DATA STRUCTURES FOR ROBUST MULTIFREQUENCY IMAGING

MIGUEL MOSCOSO*, ALEXEI NOVIKOV†, GEORGE PAPANICOLAOU‡, AND CHRYSOULA TSOGKA§

Abstract. In this paper we consider imaging problems that can be cast in the form of an underdetermined linear system of equations. When a single measurement vector is available, a sparsity promoting ℓ_1-minimization based algorithm may be used to solve the imaging problem efficiently. A suitable algorithm in the case of multiple measurement vectors would be the MUltiple SIgnal Classification (MUSIC) which is a subspace projection method. We provide in this work a theoretical framework in an abstract linear algebra setting that allows us to examine under what conditions the ℓ_1-minimization problem and the MUSIC method admit an exact solution. We also examine the performance of these two approaches when the data are noisy. Several imaging configurations that fall under the assumptions of the theory are discussed such as active imaging with single or multiple frequency data. We also show that the phase retrieval problem can be re-cast under the same linear system formalism using the polarization identity and relying on diversity of illuminations. The relevance of our theoretical analysis in imaging is illustrated with numerical simulations and robustness to noise is examined by allowing the background medium to be weakly inhomogeneous.

Key words. array imaging, phase retrieval, ℓ_1-minimization, MUSIC

1. Introduction.

Imaging is an inverse problem in which we seek to reconstruct a medium's characteristics, such as the reflectivity, by recording its response to one or more known excitations. The output is usually an image giving an estimate of an unknown characteristic in a bounded domain, the imaging window of interest. Although this problem is in all generality non-linear, it is often adequately formulated as a linear system of the form

$$\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}\,, \tag{1.1}$$

where the data vector $\boldsymbol{b} \in \mathbb{C}^N$ is a linear transformation of the unknown vector $\boldsymbol{\rho} \in \mathbb{C}^K$ [13]. $\mathcal{A} \in \mathbb{C}^{N\times K}$ is the model matrix that relates $\boldsymbol{b}$ to $\boldsymbol{\rho}$. Typically, the linear system (1.1) is underdetermined because the number of unknowns K is much larger than the number of measurements N, so $N \ll K$.

We are interested in this work in imaging problems where the unknown $\boldsymbol{\rho}$ is M-sparse with $M \ll K$. Under this assumption (1.1) falls under the compressive sensing framework [21, 16, 22]. It follows from [16] that the unique M-sparse solution of (1.1) can be obtained with ℓ_1-optimization when the mutual coherence[1] of the model matrix $\mathcal{A}$ is smaller than $1/(2M)$. The same result can be obtained assuming $\mathcal{A}$ obeys the M-restricted isometry property [7] which basically states that all sets of M-columns of $\mathcal{A}$ behave approximately as an orthonormal system.

We show that uniqueness for the minimal ℓ_1 solution of (1.1) can be obtained under less restrictive conditions on the model matrix $\mathcal{A}$ provided that the unknown $\boldsymbol{\rho}$ is such that the columns of $\mathcal{A}$ that correspond to the support T of $\boldsymbol{\rho}$ are approximately

*Department of Mathematics, Universidad Carlos III de Madrid, Leganes, Madrid 28911, Spain (moscoso@math.uc3m.es)

†Mathematics Department, Penn State University, University Park, PA 16802 (anovikov@math.psu.edu)

‡Department of Mathematics, Stanford University, Stanford, CA 94305 (papanicolaou@stanford.edu)

§Applied Math Unit, University of California, Merced, 5200 North Lake Road, Merced, CA 95343 (ctsogka@ucmerced.edu).

[1]The mutual coherence of $\mathcal{A}$ is defined as $\max_{i\neq j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle|$ with $\boldsymbol{a}_i \in \mathbb{C}^N$ the columns of $\mathcal{A}$ normalized to one, so that $\|\boldsymbol{a}_i\|_{\ell_2} = 1\ \forall\, i = 1, \ldots, K$.

orthogonal, so there exists a small value $0 < \varepsilon < 1/2$ such that

$$|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| < \frac{\varepsilon}{M}, \quad \forall\, i, j \in T, i \neq j.$$

Under this assumption, we associate to each column vector $\boldsymbol{a}_j,\ j \in T$, its vicinity

$$S_j = \left\{ k \neq j \ \text{ s.t. } \ |\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle| \geqslant \frac{1}{2M} \right\}$$

that contains all columns of $\mathcal{A}$ that are approximately parallel to $\boldsymbol{a}_j$. This result finds interesting applications in imaging since it states under what conditions the location of well separated reflectors can be determined with high precision. It can be also used to explain super-resolution, i.e., the significantly superior resolution that ℓ_1-optimization provides compared to the conventional resolution of the imaging system, i.e., the Rayleigh resolution. Moreover, we address the robustness to noise of the minimal ℓ_1 solution and show that for noisy data the solution $\boldsymbol{\rho}$ can be decomposed in two parts: the coherent part $\boldsymbol{\rho}_c$, which is supported in T or in the vicinities S_j, and the incoherent part $\boldsymbol{\rho}_i$, usually referred to as grass, that is small. Other stability results can be found in [7, 8, 17, 35, 18, 4].

The notion of vicinities and weak interaction between scatterers has been considered in [18] and [4]. In [18], several algorithms for imaging well separated sources were introduced and analyzed. These algorithms address the issue of high coherence in $\mathcal{A}$ using techniques of band exclusion and local optimization. In [4], a resolution analysis for ℓ_1-minimization and ℓ_1-penalty was carried out for array imaging in the paraxial regime. It was shown that for well separated sources or clusters of sources the minimal ℓ_1 solution is supported mainly in the vicinities of the true sources' locations.

More recently in [5], the problem of imaging sources in weakly inhomogeneous media was addressed using Coherent INTerferometry (CINT) followed by ℓ_1 convex optimization for debluring. This is a natural idea since, as it was shown in [1] (see also [3]), the CINT image is a convolution of the reflectivity with a Gaussian kernel. Hence, the resolution in CINT images can be refined by debluring as in [2], where a level set method was used. In [5], debluring was performed with ℓ_1-optimization and its performance was analyzed for well separated sources and well separated clusters of sources.

We also consider in this paper the more general form that system (1.1) takes when S multiple measurement vectors (MMV) are available, so

$$\mathcal{A}_{l_q} \boldsymbol{\rho} = \boldsymbol{b}_{l_q}, \quad q = 1, \ldots, S. \tag{1.2}$$

Here, $\boldsymbol{l}_q = [l_{1q}, l_{2q}, \ldots, l_{Kq}]^T$ denotes a parameter vector such as the excitation that we control. To simplify the notation, we will denote the different excitations by the scalar q and write $\mathcal{A}_q \boldsymbol{\rho} = \boldsymbol{b}_q$ instead, unless it is necessary to explicitly state that the model matrix depends on a vector $\boldsymbol{l}_q$. To solve (1.2) we consider the MUltiple SIgnal Classification algorithm [34] which has been used successfully in signal processing [23] and imaging [15, 25]. For a careful analysis of MUSIC for single snapshot spectral imaging we refer the reader to [26]. We show here that MUSIC gives the exact support of the solution of (1.2) in the noise free case when the matrices $\mathcal{A}_q$ admit the following factorization

$$\mathcal{A}_q = \tilde{\mathcal{A}}\, \Lambda_q, \ \text{with } \Lambda_q \text{ diagonal.} \tag{1.3}$$

In this case, (1.2) admits the following MMV formulation

$$\tilde{\mathcal{A}}\boldsymbol{\rho}_q = \boldsymbol{b}_q;\ \ \boldsymbol{\rho}_q = \Lambda_q \boldsymbol{\rho}\,,$$

where the multiple unknown vectors $\boldsymbol{\rho}_q$, $q = 1, \ldots, S$, share the same support. The main advantage of this formulation is that we can immediately infer that the data vectors $\boldsymbol{b}_q$ are linear combinations of the same M-columns of $\tilde{\mathcal{A}}$, those that belong to the support of the unknown $\boldsymbol{\rho}$. The implication is that the columns of $\tilde{\mathcal{A}}$ indexed by $T = \mathrm{supp}(\boldsymbol{\rho})$ span the column subspace of B, the 'signal' subspace of B. Hence, the support T is the zero set of the orthogonal projections of the columns of matrix $\tilde{\mathcal{A}}$ onto the null space of the data matrix B. Moreover, the support is recovered exactly under the assumption that all M-sets of columns of $\tilde{\mathcal{A}}$ are linearly independent. We discuss several imaging configurations for which the factorization (1.3) is feasible as well as instances where (1.3) holds only approximately and MUSIC is no longer exact even for noise free data.

Let us remark that for different excitations q we obtain multiple measurement vectors $\boldsymbol{b}_q$ which correspond to linear transformations of the same unknown vector $\boldsymbol{\rho}$. The data can be arranged in a matrix $B \in \mathbb{C}^{N\times S}$ whose columns are the vectors $\boldsymbol{b}_q$, and the MMV formulation may be expressed as a matrix-matrix equation

$$\tilde{\mathcal{A}}\mathrm{P} = B\,,$$

where the unknown is now the matrix $\mathrm{P} \in \mathbb{C}^{K\times S}$ whose columns are the vectors $\boldsymbol{\rho}_q = \Lambda_q \boldsymbol{\rho}$ that share the same support. The optimization can therefore be performed within the MMV formalism as described in [14, 24, 36, 37]. The main idea is to seek the solution with the minimal (2,1)-norm which consists in minimizing the ℓ_1 norm of the vector formed by the ℓ_2 norms of the rows of the unknown matrix P. This guarantees the common support of the solution's columns. We do not pursue this approach here and refer the reader to [12] for an application of this formalism to imaging strong scattering scenes as well as to [6] where an MMV formulation for synthetic aperture imaging of frequency and direction dependent reflectivity was introduced and analyzed.

We present several configurations in array imaging that can be cast under the general framework discussed here, such as single- and multiple-frequency array imaging using single- or multiple-receivers. All these problems can be formulated as (1.1) for a single measurement vector, or as (1.2) when multiple measurement vectors are available. We also consider the non-linear phase retrieval problem, which according to [31, 28, 29] can be reduced to a linear system of the form (1.2). This requires intensity data corresponding to multiple coherent illuminations which when using the polarization identity are transformed to interferometric data. We consider multiple frequency intensity data collected at a single receiver due to multiple coherent illuminations that could be generated by a spatial light modulator (SLM) [30]. The solution of (1.2) may then be computed with Single Receiver INTerferometry (SRINT) as in [29], ℓ_1-minimization or MUSIC.

The performance of these imaging methods for the non-linear phase retrieval problem is studied with numerical simulations in an optical digital microscopy imaging regime. Our simulations allow us to asses the robustness of the different methods to modeling errors resulting to perturbations in the unknown phases of the recorded data. We consider phase perturbations that are either due to grid displacements or to wave propagation in a weakly inhomogeneous medium. Our conclusions are that SRINT provides the less satisfactory image in terms of resolution but it is the more robust method when there are modeling errors, the ℓ_1 method has the best resolution but is not very robust with respect to noise, while MUSIC seems to be the more competitive method at moderate signal to noise ratio regimes because it has better resolution than SRINT and is less sensitive to noise than ℓ_1-minimization.

The paper is organized as follows. In Section 2 we present in a abstract linear algebra framework the conditions under which ℓ_1-minimization and MUSIC provide the exact solution to problems (1.1) and (1.2) respectively. We also analyze the performance of these methods for noisy data. In Section 3 we formulate the array imaging problem and consider some common configurations used in active array imaging. Moreover, we discuss how the imaging problem can be cast under the abstract framework of Section 2 and what are adequate data-structures to be used in imaging with ℓ_1-minimization and MUSIC. In Section 4, we explore with numerical simulations the robustness of the imaging methods for the phase retrieval problem in an optical (digital) microscopy regime. In Section 5 we illustrate with numerical simulations how our abstract theoretical results are relevant in assessing image resolution. Section 6 contains our conclusions.

2. Linear algebra aspects of imaging algorithms.

In this section we discuss under what conditions ℓ_1-minimization and MUSIC algorithms provide the exact solution when there is no noise in the data. We also discuss the performance of these algorithms for noisy data. We assume that imaging can be formulated as a linear inverse problem of the form

$$\mathcal{A}_l \boldsymbol{\rho} = \boldsymbol{b}_l \,, \tag{2.1}$$

that is underdetermined. In (2.1), the model matrix

$$\mathcal{A}_l = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \boldsymbol{a}_1^{(l_1)} & \boldsymbol{a}_2^{(l_2)} & \dots & \boldsymbol{a}_K^{(l_K)} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{C}^{N\times K} \tag{2.2}$$

relates the unknown vector $\boldsymbol{\rho} \in \mathbb{C}^K$, which is the “image” to be constructed, to the transformed vector $\boldsymbol{b}_l \in \mathbb{C}^N$, which contains the data. This matrix is fixed by the physical setup of the imaging system and, therefore, it is given to us. However, the important observation here is that $\mathcal{A}_l$ also depends on a parameter vector $\boldsymbol{l} = [l_1, l_2, \dots, l_K]^T$ which may be varied so as several transformed vectors $\boldsymbol{b}_l$ of the same unknown $\boldsymbol{\rho}$ can be obtained.

If only one snapshot of array measurements is available for imaging, we solve (2.1) for a single measurement vector (SMV) $\boldsymbol{l}$ using ℓ_1 minimization that promotes the assumed sparsity of the vector $\boldsymbol{\rho}$. In that case, we will write (2.1) simply as $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$. When several snapshots of array measurements corresponding to different parameter vectors $\boldsymbol{l}_q$ are available, we solve the corresponding MMV problem using MUSIC. In that case, we will write (2.1) as $\mathcal{A}_q\boldsymbol{\rho} = \boldsymbol{b}_q$.

2.1. ℓ_1 minimization-based methods.

In the imaging problems considered here we assume that the scatterers occupy only a small fraction of a region of interest called the image window IW. This means that the true reflectivity vector $\boldsymbol{\rho}_0$ is sparse, so the number of its entries that are different than zero, denoted by M, is much smaller than its length K. Thus, $M = |\operatorname{supp}(\boldsymbol{\rho}_0)| \ll K$. This prior knowledge changes the imaging problem substantially because we can exploit the sparsity of $\boldsymbol{\rho}_0$ by formulating (2.1) as an optimization problem which seeks the sparsest vector in

$\mathbb{C}^K$ that equates model and data. Thus, for a single measurement vector $\boldsymbol{b}$ we solve

$$\min \|\boldsymbol{\rho}\|_{\ell_1} \quad \text{subject to} \quad \mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}. \tag{2.3}$$

In this form, we may be able to pick the true solution $\boldsymbol{\rho}_0$ if the matrix $\mathcal{A}$ and the sparsity of $\boldsymbol{\rho}_0$ fulfill certain conditions. In particular, we have the following four theorems whose proofs are given in Appendix A. We denote by $\|\cdot\|_{\ell_2}$ and $\|\cdot\|_{\ell_1}$ the ℓ_2 and ℓ_1 norms of a vector, respectively.

THEOREM 2.1. *M-sparse solutions of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$ are unique, if*

$$|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| < \frac{1}{2M} \quad \forall i \neq j, \tag{2.4}$$

where we assume that the columns of matrix $\mathcal{A}$ are normalized so that $\|\boldsymbol{a}_i\|_{\ell_2} = 1$ $\forall i$.

THEOREM 2.2. *The M-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$ can be found as the solution of*

$$\min \|\boldsymbol{\eta}\|_{\ell_1}, \text{ subject to } \mathcal{A}\boldsymbol{\eta} = \boldsymbol{b}, \tag{2.5}$$

if

$$|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| < \frac{1}{2M}, \quad \forall i \neq j, \tag{2.6}$$

where we assume that the columns of matrix $\mathcal{A}$ are normalized so that $\|\boldsymbol{a}_i\|_{\ell_2} = 1$ $\forall i$.

THEOREM 2.3. *Let $\boldsymbol{\rho}$ be a solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$, and let T be the index set of the support of $\boldsymbol{\rho}$, so*

$$T = \operatorname{supp}(\boldsymbol{\rho}), \quad \text{and} \quad M = |T|.$$

Fix a positive $\varepsilon < 1/2$, and suppose that the matrix $\mathcal{A}$ satisfies:

(i) The column vectors are normalized so that $\|\boldsymbol{a}_i\|_{\ell_2} = 1$ $\forall i$.

(ii) The column vectors in the set T are approximately orthogonal, so

$$|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| < \frac{\varepsilon}{M}, \quad \forall\, i, j \in T, i \neq j. \tag{2.7}$$

(iii) For any $j \in T$ the vicinity

$$S_j = \left\{ k \neq j \ \ s.t. \ \ |\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle| \geqslant \frac{1}{2M} \right\} \tag{2.8}$$

has the properties

$$|\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle| \leqslant 1 - 2\varepsilon \quad \forall k \in S_j\,, \tag{2.9}$$

and

$$|\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle| < \frac{\varepsilon}{M} \quad \forall k \in S_i, \ \forall i \neq j. \tag{2.10}$$

Then $\boldsymbol{\rho}$, the M-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$, can be found as the solution of

$$\min \|\boldsymbol{\eta}\|_{\ell_1}, \text{ subject to } \mathcal{A}\boldsymbol{\eta} = \boldsymbol{b}.$$

THEOREM 2.4. **Noisy case.** *Let $\boldsymbol{\rho}$ be an M-sparse solution of*

$$\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b},$$

and let $T = \mathrm{supp}(\boldsymbol{\rho})$, so $M = |T|$. Fix a positive $\varepsilon < 1/2$, and suppose that $\mathcal{A}$ satisfies conditions (i), (ii), and (iii) of Theorem 2.3.

Furthermore, let $\boldsymbol{\rho}_\delta$ be the minimal ℓ_1-norm solution of the noisy problem

$$\min \|\boldsymbol{\eta}\|_{\ell_1}, \ \textit{subject to } \mathcal{A}\boldsymbol{\eta} = \boldsymbol{b}^\delta, \tag{2.11}$$

with $\boldsymbol{b}^\delta$ defined by

$$\boldsymbol{b}^\delta = \boldsymbol{b} + \boldsymbol{\delta b}, \tag{2.12}$$

such that the noise $\boldsymbol{\delta b}$ is bounded for some small positive δ, so that

$$\|\boldsymbol{\delta b}\|_{\ell_2} \leqslant \delta\,. \tag{2.13}$$

Assume that $\mathcal{A}$ has the property that the solution $\boldsymbol{\delta\rho}$ to

$$\min \|\boldsymbol{\eta}\|_{\ell_1}, \ \textit{subject to } \mathcal{A}\boldsymbol{\eta} = \boldsymbol{\delta b}, \tag{2.14}$$

satisfies

$$\|\boldsymbol{\delta\rho}\|_{\ell_1} \leqslant C\|\boldsymbol{\delta b}\|_{\ell_2}. \tag{2.15}$$

Then, we can show that the solution $\boldsymbol{\rho}_\delta$ of (2.11) can be decomposed as

$$\boldsymbol{\rho}_\delta = \boldsymbol{\rho}_c + \boldsymbol{\rho}_i, \tag{2.16}$$

with $\boldsymbol{\rho}_c$ the coherent part of the solution supported on T or in the vicinities S_j with $j \in T$, and $\boldsymbol{\rho}_i$ the incoherent part of the solution which is supported away from the vicinities and it is small. Specifically, for $\boldsymbol{\rho}_c$ we have that for any $j \in T$

$$|\ |(\boldsymbol{\rho})_j| - |(\boldsymbol{\rho}_c)_j + \sum_{k \in S_j} \langle a_j, a_k \rangle (\boldsymbol{\rho}_c)_k|\ | \leqslant \delta_0 + C\delta,$$

with

$$\delta_0 = \frac{2C\delta(1-\varepsilon)}{M(1-2\varepsilon)} + \frac{2\varepsilon(\|\boldsymbol{\rho}\|_{\ell_1} + C\delta)}{M}.$$

While for $\boldsymbol{\rho}_i$ we can show that:

$$\|\boldsymbol{\rho}_i\|_{\ell_1} \leqslant \delta_1,$$

with δ_1 given by

$$\delta_1 = C\delta + \frac{4C\delta(1-\varepsilon)}{(1-2\varepsilon)}\,.$$

Theorems 2.1 and 2.2 are well known results in the literature of compressive sensing [21, 16, 22]. The first theorem tells us that the M-sparse solution of the linear system $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$ is unique when the columns of the matrix satisfy the orthonormality condition (2.4). This condition is satisfied when the mutual coherence of the matrix $\mathcal{A}$, defined as $\max_{i \neq j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|$, is smaller than $1/(2M)$. This first theorem is an ℓ_0 uniqueness result. The second result, Theorem 2.2, tells us that the unique M-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$ can be found by solving the ℓ_1 minimization problem (2.5).

Algorithm 1 GelMa for solving (2.5)

Require: Set $\boldsymbol{y} = \boldsymbol{0}$, $\boldsymbol{z} = \boldsymbol{0}$. Pick the step size β, and a regularization parameter τ.
  repeat
    Compute the residual $\boldsymbol{r} = \boldsymbol{b} - \mathcal{A}\boldsymbol{y}$
    $\boldsymbol{y} \Leftarrow \eta_{\tau\beta}(\boldsymbol{y} + \beta\mathcal{A}^*(\boldsymbol{z} + \boldsymbol{r}))$
    $\boldsymbol{z} \Leftarrow \boldsymbol{z} + \beta\boldsymbol{r}$
  until Convergence

This is a very useful result because it is the ℓ_1 minimization problem that can be solved efficiently in practice, for example, by using the algorithm GelMa described in Algorithm 1, which involves only simple matrix-vector multiplications followed by a shrinkage-thresholding step defined by the operator $\eta_\tau(y_i) = \operatorname{sign}(y_i)\max\{0, |y_i| - \tau\}$. In the noiseless case, this algorithm converges to the exact solution independently of the value of the regularization parameter τ. For more details we refer to [27].

Theorem 2.3 is to the best of our knowledge new. Its proof is given in Appendix A. This theorem tells us that the M-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$ can be recovered by solving the ℓ_1 minimization problem under a less stringent condition than (2.6) provided that the column vectors of the matrix $\mathcal{A}$ that are in the support of the true solution $\boldsymbol{\rho}_0$ are approximately orthogonal, that is, they satisfy (2.7). Note that we allow for the columns of $\mathcal{A}$ to be close to collinear. Moreover, we define the vicinities S_j for the column vectors $\boldsymbol{a}_j$ in the support of the true solution, and we assume that all the column vectors that are in the vicinity of a *support column vector* are close enough to it, so (2.9) holds. We also assume that the vicinities S_i and S_j, for $i \neq j$, are far enough, so (2.10) holds.

The last result, Theorem 2.4, is the noisy version of Theorem 2.3. It shows that when the data $\boldsymbol{b}$ is not exact but is known up to some bounded vector $\boldsymbol{\delta b}$, the solution $\boldsymbol{\rho}_\delta$ of the minimization problem (2.11)-(2.12) is close to the solution of the original (noiseless) problem in the following sense. The solution $\boldsymbol{\rho}_\delta$ can be decomposed in two parts: the coherent part $\boldsymbol{\rho}_c$ supported in T or in the vicinities S_j, $j \in T$, of the true solution, and the incoherent part $\boldsymbol{\rho}_i$ usually referred to as grass in imaging. The grass is supported away from the vicinities S_j and it is shown to be small assuming that (2.15) holds for the solution to (2.14) and assuming that the norm of the noise is small so (2.13) holds. Other stability results can be found in [7, 8, 17, 35, 18, 4].

We will see in Section 5 how Theorems 2.3 and 2.4 can be applied in imaging.

2.2. MUSIC. MUSIC is a subspace imaging algorithm based on the decomposition of the measurements into two orthogonal domains: the signal and noise subspaces [34]. The key is to be able to form a data matrix

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1S} \\ b_{21} & b_{22} & \dots & b_{2S} \\ \dots & \dots & \dots & \dots \\ b_{N1} & b_{N2} & \dots & b_{NS} \end{pmatrix} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \boldsymbol{b}_1 & \boldsymbol{b}_2 & \dots & \boldsymbol{b}_S \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{C}^{N\times S}, \tag{2.17}$$

whose column vectors $\boldsymbol{b}_q$ are obtained from a family of linear systems $\mathcal{A}_q\boldsymbol{\rho} = \boldsymbol{b}_q$ that can be rewritten in the form

$$\tilde{\mathcal{A}}\Lambda_q\boldsymbol{\rho} = \boldsymbol{b}_q\,, \quad q = 1, \dots, S, \tag{2.18}$$

where Λ_q is a diagonal matrix whose entries can be controlled to form the images. The assumption here is that the model matrices $\mathcal{A}_q$ relating the unknown vector $\boldsymbol{\rho}$

with the data vectors $\boldsymbol{b}_q$ can be factorized into two matrices

(2.19)
$$\tilde{\mathcal{A}} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \tilde{\boldsymbol{a}}_1 & \tilde{\boldsymbol{a}}_2 & \dots & \tilde{\boldsymbol{a}}_K \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{C}^{N\times K} \quad \text{and} \quad \Lambda_q = \begin{pmatrix} l_{1q} & 0 & & \\ 0 & l_{2q} & & \\ & & \ddots & \\ & & 0 & l_{Kq} \end{pmatrix} \in \mathbb{C}^{K\times K},$$

with $\tilde{\mathcal{A}}$ independent of the parameter vector $\boldsymbol{l}_q = [l_{1q}, l_{2q}, \dots, l_{Kq}]^T$, and Λ_q diagonal. Under this assumption, the imaging problem (2.18) can be reinterpreted in the form of an MMV problem

$$\tilde{\mathcal{A}}\boldsymbol{\rho}_q = \boldsymbol{b}_q, \tag{2.20}$$

with $\boldsymbol{\rho}_q = \Lambda_q\boldsymbol{\rho}$. Physically, each $\boldsymbol{\rho}_q$ is a transformed version of the same unknown vector $\boldsymbol{\rho}$. The data can be arranged into the data matrix (2.17), and (2.20) may be expressed as a matrix-matrix equation

$$\tilde{\mathcal{A}}\mathrm{P} = B\,, \tag{2.21}$$

where the columns of $\mathrm{P} \in \mathbb{C}^{K\times S}$, $\boldsymbol{\rho}_q = \Lambda_q\boldsymbol{\rho}$, share the same support.

The important element of the new formulation (2.20) (or (2.21)) is that now all the data vectors $\boldsymbol{b}_q$ are linear combinations of the same M columns of $\tilde{\mathcal{A}}$ (or $\mathcal{A}$), those columns that correspond to $T = \mathrm{supp}(\boldsymbol{\rho})$, with $M = |T|$. Thus, every column of $\tilde{\mathcal{A}}$ indexed by T is contained in the column space of B, the signal subspace, which is orthogonal to the noise subspace. Hence, one can simply find the unknown support T by projecting the columns of $\tilde{\mathcal{A}}$ onto the noise subspace. Both, the signal and the noise subspaces can be obtained via the singular value decomposition (SVD) of B.

More precisely, the objective of a MUSIC algorithm is to find the support T of an unknown sparse vector $\boldsymbol{\rho} = [\rho_1, \rho_2, \dots, \rho_K]^T$ with a number of nonzero entries M much smaller than its length K. With a sufficiently diverse number of experiments $S \geq M$ we create a data matrix B, and we compute its SVD

$$B = U\Sigma V^* = \sum_{j=1}^{K} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^*\,. \tag{2.22}$$

If the data is noiseless there are M nonzero singular values $\sigma_1 > \sigma_2 > \cdots > \sigma_M > 0$ with corresponding (left) singular vectors $\boldsymbol{u}_j$, $j = 1, \dots, M$ that span the signal subspace of $\mathbb{C}^N$. The remaining singular values σ_j, $j = M+1, \dots, K$, are zero, and the corresponding (left) singular vectors span the noise subspace of $\mathbb{C}^N$. Because the set of columns of $\tilde{\mathcal{A}}$ indexed by $T = \mathrm{supp}(\boldsymbol{\rho})$ also spans the signal subspace, the sought support T corresponds to the zero set of the orthogonal projections of the columns vectors $\tilde{\boldsymbol{a}}_k$ onto the noise subspace. Thus, it follows that the support of $\boldsymbol{\rho}$ can be found among the zeros of the imaging functional

$$\mathcal{I}_k^{\mathrm{SIGNAL}} = \sum_{j=1}^{M} |\tilde{\boldsymbol{a}}_k^* \boldsymbol{u}_j|^2, \; k = 1, \dots, K, \tag{2.23}$$

or, equivalently, among the peaks of the imaging functional

$$\mathcal{I}_k^{\mathrm{MUSIC}} = \frac{\|\tilde{\boldsymbol{a}}_k\|_{\ell_2}}{\sum_{j=M+1}^{N} |\tilde{\boldsymbol{a}}_k^* \boldsymbol{u}_j|^2}\,, \; k = 1, \dots, K. \tag{2.24}$$

Furthermore, if all sets of M columns of $\tilde{\mathcal{A}}$ are linearly independent, then the peaks exactly coincide with the support of $\boldsymbol{\rho}$ in the noiseless case. In (2.24), the numerator is a normalization factor.

Once the support is recovered, the problem typically becomes overdetermined ($N > |\operatorname{supp}(\boldsymbol{\rho})|$) and the nonzero values of $\boldsymbol{\rho}$ can be easily found by solving the linear system restricted to the given support with an ℓ_2 method [13].

Regarding imaging with noisy data, it follows from Weyl's theorem [39] that when noise is added to the data so $B \to B^\delta = B + E$ with $\|E\|_{\ell_2} < \delta$, then no singular value σ^δ moves more than the norm of the perturbation, i.e., $\|\sigma^\delta - \sigma\|_{\ell_2} < \delta$. Hence, (i) perturbed and unperturbed singular values are paired, and (ii) the spectral gap between the zero and the nonzero singular values remains large if the smallest nonzero unperturbed singular value $\sigma_M \gg \delta$. If the noise is not too large, then the rank of the data matrix B^δ can be determined, and so is $M = |T|$.

The signal and noise subspaces are also perturbed in the presence of noise. It can be shown, however, that the perturbed subspaces remain close to the unperturbed ones, with changes that are proportional to the reciprocal of the spectral gap $\beta = \sigma^\delta_M - \sigma_{M+1}$. This follows from Wedin's Theorem [38].

THEOREM 2.5. (Wedin) *Let B have the SVD $B = Q + Q_0$ with $Q = U\Sigma V^T$ and $Q_0 = U_0\Sigma_0 V_0^T$, and let the perturbed matrix $B^\delta = B + E$ have the SVD $B^\delta = Q^\delta + Q_0^\delta$ with $Q^\delta = U^\delta \Sigma^\delta {V^\delta}^T$ and $Q_0^\delta = U_0^\delta \Sigma_0^\delta {V_0^\delta}^T$. If there exist two constants $\alpha \geq 0$ and $\beta > 0$ such that $\sigma_{max}(Q_0) \leq \alpha$ and $\sigma_{min}(Q^\delta) \geq \alpha + \beta$, then the distance between the orthogonal projections onto the subspaces $R(Q)$ and $R(Q^\delta)$ is bounded by*

$$\|P_{R(Q^\delta)} - P_{R(Q)}\|_{\ell_2} \leqslant \frac{\delta}{\beta}\,, \tag{2.25}$$

where $\delta = \max(\|EV\|_{\ell_2}, \|E^* U\|_{\ell_2})$.

There is much work done on the robustness of MUSIC with respect to noise. We refer to [26], and references therein, for a recent discussion about how much noise the MUSIC algorithm can tolerate. When we apply the Theorem 2.5 to our imaging problem, where $Q_0 = 0$, we obtain the following result whose proof is in Appendix B.

THEOREM 2.6. *Let $X =$Diag$(\boldsymbol{\rho})$ be a diagonal matrix that solves*

$$\tilde{\mathcal{A}} X L = B, \tag{2.26}$$

where $\tilde{\mathcal{A}}$ satisfies conditions (i), (ii), and (iii) of Theorem 2.3 for a fixed $\varepsilon < 1/3$,

$$L = \begin{pmatrix} l_{11} & l_{12} & & l_{1S} \\ l_{21} & l_{22} & & l_{2S} \\ \vdots & \vdots & & \vdots \\ l_{K1} & l_{K2} & & l_{KS} \end{pmatrix} \in \mathbb{C}^{K\times S}\,,$$

and B is the noiseless data matrix (2.17) with SVD $B = Q = U\Sigma V^T$. Let the perturbed matrix $B^\delta = Q^\delta + Q_0$ be such that $\sigma_{\max}(B^\delta - B) \leqslant \delta$. Suppose $\boldsymbol{\rho}$, the vector diagonal entries of X, is sparse with $T = \operatorname{supp}(\boldsymbol{\rho})$, $M = |T|$, $M \ll$ size$(\boldsymbol{\rho})$, and

$$\rho_m = \min_{\rho_i \neq 0}\{|\rho_i|\}.$$

Let L_T be the submatrix of L, formed by the rows corresponding to T, has

$$\sigma_m^T = \sigma_{\min}(L_T). \tag{2.27}$$

If

$$2\delta < \rho_m \sigma_m^T (1-3\varepsilon), \tag{2.28}$$

the orthogonal projections onto the subspaces $R(Q^\delta)$ and $R(B)$ are close:

$$\|P_{R(Q^\delta)} - P_{R(B)}\|_{\ell_2} \leqslant \frac{\delta}{\rho_m \sigma_m^T (1-3\varepsilon)}. \tag{2.29}$$

To conclude, the main step in setting up MUSIC is to be able to find a suitable factorization of the model matrix as $\mathcal{A}_q = \tilde{\mathcal{A}}\Lambda_q$, where Λ_q is diagonal. In that case, the imaging vectors are just the columns of $\tilde{\mathcal{A}}$ that are given. We discuss next imaging situations in which this factorization is possible and MUSIC can form form images with high precision. We also discuss applications in which the factorization is only approximate and, hence, images obtained with MUSIC lose resolution.

3. Array imaging: data models.

The goal of array imaging is to form images inside a region of interest called the image window IW. In active array imaging the array probes the medium by sending signals and recording the echoes. Probing of the medium can be done with many different types of arrays that differ in their number of transmitters and receivers, their geometric layouts, or the type of signals they use for illumination. They may use single frequency signals sent from different positions, or multifrequency signals sent from one or more positions. Of course, the problem of active array imaging also depends on the receivers. They can record the intensities and phases of the signals that arrive to the array or only their intensities.

In this section, we describe some common configurations used in active array imaging. The array, with N transducers separated by a distance h, has a characteristic length a (see Figure 1). The transducers emit signals from positions $\vec{\boldsymbol{x}}_s$ and record the echoes at positions $\vec{\boldsymbol{x}}_r$, $s, r = 1, 2, \ldots, N$. They can use single or multifrequency signals, with frequencies ω_l, $l = 1, \ldots, S$. Our goal is to reconstruct a sparse scene consisting of M point-scatterers at a distance L from the array, whose positions $\vec{\boldsymbol{y}}_{n_j}$ and reflectivities $\alpha_{n_j} \in \mathbb{C}$, $j = 1, \ldots, M$, we seek to determine. The ambient medium between the array and the scatterers can be homogeneous or inhomogeneous.

In order to form the images we discretize the IW using a uniform grid of points $\vec{\boldsymbol{y}}_k$, $k = 1, \ldots, K$, and we introduce the *true reflectivity vector*

$$\boldsymbol{\rho}_0 = [\rho_{01}, \ldots, \rho_{0K}]^T \in \mathbb{C}^K,$$

such that $\rho_{0k} = \sum_{j=1}^M \alpha_{n_j} \delta_{\vec{y}_{n_j}\vec{y}_k}$, $k = 1, \ldots, K$, where $\delta_{\cdot\cdot}$ is the classical Kronecker delta. We will not assume that the scatterers lie on the grid, i.e., $\{\vec{\boldsymbol{y}}_{n_1}, \ldots, \vec{\boldsymbol{y}}_{n_M}\} \not\subset \{\vec{\boldsymbol{y}}_1, \ldots, \vec{\boldsymbol{y}}_K\}$ in general. To write the data received on the array in a compact form, we define the Green's function vector

$$\widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}};\omega) = [\widehat{G}(\vec{\boldsymbol{x}}_1, \vec{\boldsymbol{y}};\omega), \widehat{G}(\vec{\boldsymbol{x}}_2, \vec{\boldsymbol{y}};\omega), \ldots, \widehat{G}(\vec{\boldsymbol{x}}_N, \vec{\boldsymbol{y}};\omega)]^T \tag{3.1}$$

at location $\vec{\boldsymbol{y}}$ in the IW, where $\widehat{G}(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}};\omega)$ denotes the free-space Green's function of the homogeneous or inhomogeneous medium. This function characterizes the propagation of a signal of angular frequency ω from point $\vec{\boldsymbol{y}}$ to point $\vec{\boldsymbol{x}}$, so (3.1) represents the signal received at the array due to a point source of frequency ω at $\vec{\boldsymbol{y}}$. When the medium is homogeneous,

$$\widehat{G}(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}};\omega) = \widehat{G}_0(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}};\omega) = \frac{\exp(\mathrm{i}\kappa|\vec{\boldsymbol{x}} - \vec{\boldsymbol{y}}|)}{4\pi|\vec{\boldsymbol{x}} - \vec{\boldsymbol{y}}|}, \quad \kappa = \frac{\omega}{c_0}. \tag{3.2}$$

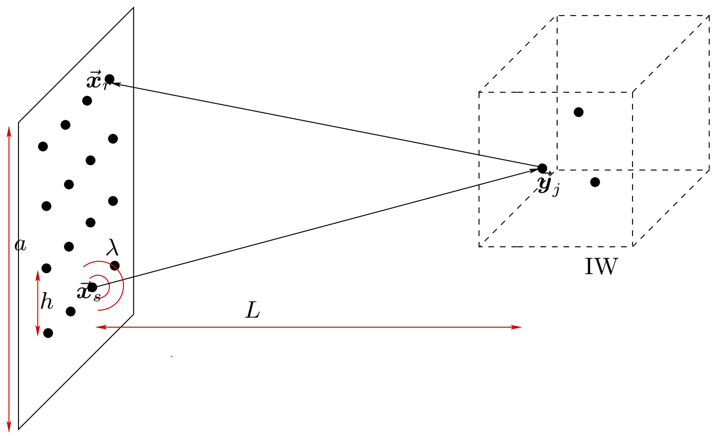

FIG. 1. *General setup of an array imaging problem. The transducer at $\vec{\boldsymbol{x}}_s$ emits a probing signal and the reflected intensities are recorded at $\vec{\boldsymbol{x}}_r$. The scatterers located at $\vec{\boldsymbol{y}}_j$, $j = 1, \ldots, M$ are at distance L from the array and inside the image window IW.*

In this case, the Green's function vector is

$$\widehat{\boldsymbol{g}}_0(\vec{\boldsymbol{y}};\omega) = [\widehat{G}_0(\vec{\boldsymbol{x}}_1,\vec{\boldsymbol{y}};\omega), \widehat{G}_0(\vec{\boldsymbol{x}}_2,\vec{\boldsymbol{y}};\omega), \ldots, \widehat{G}_0(\vec{\boldsymbol{x}}_N,\vec{\boldsymbol{y}};\omega)]^T.$$

We assume that the scatterers are far apart or that the reflectivities are small, so multiple scattering between them is negligible. In this case, the Born approximation holds and, thus, the response at $\vec{\boldsymbol{x}}_r$ (including phases) due to a pulse of angular frequency ω_l sent from $\vec{\boldsymbol{x}}_s$, and reflected by the M scatterers, is given by

$$P(\vec{\boldsymbol{x}}_r,\vec{\boldsymbol{x}}_s;\omega_l) = \sum_{j=1}^{M} \alpha_j G(\vec{\boldsymbol{x}}_r,\vec{\boldsymbol{y}}_{n_j};\omega_l)\, G(\vec{\boldsymbol{y}}_{n_j},\vec{\boldsymbol{x}}_s;\omega_l)\,, \tag{3.3}$$

and the the full response matrix that contains all posible information for imaging by

$$P(\omega_l) = [P(\vec{\boldsymbol{x}}_r,\vec{\boldsymbol{x}}_s;\omega_l)] = \sum_{j=1}^{M} \alpha_j \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_{n_j};\omega_l)\, \widehat{\boldsymbol{g}}^T(\vec{\boldsymbol{y}}_{n_j};\omega_l)\,. \tag{3.4}$$

Next, we describe different situations of interest in active array imaging.

3.1. Single frequency signals and multiple receivers. Let us first consider the case in which only one illumination of frequency ω is sent using the N sources in the array located at positions $\vec{\boldsymbol{x}}_s$, $s = 1, \ldots, N$. The echoes are also recorded at the N receivers located at $\vec{\boldsymbol{x}}_r$, $r = 1, \ldots, N$. If $\widehat{\boldsymbol{f}}(\omega) = [\widehat{f}_1(\omega), \ldots, \widehat{f}_N(\omega)]^T$ represents the illumination vector whose entries are the signals sent from the sources in the array, then $\widehat{g}^{(k)}_{\widehat{f}(\omega)} = \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_k;\omega)^T \widehat{\boldsymbol{f}}(\omega)$ is the field at the grid position $\vec{\boldsymbol{y}}_k$ in the IW. Thus,

$$\mathcal{A}_{\widehat{f}(\omega)} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \widehat{g}^{(1)}_{\widehat{f}(\omega)}\widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_1;\omega) & \widehat{g}^{(2)}_{\widehat{f}(\omega)}\widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_2;\omega) & \ldots & \widehat{g}^{(K)}_{\widehat{f}(\omega)}\widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_K;\omega) \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{C}^{N\times K} \tag{3.5}$$

is the model matrix that connects the unknown reflectivity vector $\boldsymbol{\rho} \in \mathbb{C}^K$ to the data vector $\boldsymbol{b}_{\widehat{f}(\omega)} \in \mathbb{C}^N$ that depends on the illumination $\widehat{\boldsymbol{f}}(\omega)$.

If a single illumination is used to form an image, then active array imaging amounts to finding $\boldsymbol{\rho}$ from the system of linear equations

$$\mathcal{A}_{\widehat{f}(\omega)}\boldsymbol{\rho} = \boldsymbol{b}_{\widehat{f}(\omega)}\,. \tag{3.6}$$

Abusing a little bit the notation used in Section 2, we have indicated in (3.6) that the control parameter vector is the illumination $\widehat{\boldsymbol{f}}(\omega)$. According to (2.1)-(2.2), the parameter vector is $\boldsymbol{l} = [\widehat{g}^{(1)}_{\widehat{f}(\omega)}, \widehat{g}^{(2)}_{\widehat{f}(\omega)}, \ldots, \widehat{g}^{(K)}_{\widehat{f}(\omega)}]^T$ which depends on the Green's function vectors $\widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}};\omega)$ fixed by the physical layout, and on the illumination vector $\widehat{\boldsymbol{f}}(\omega)$ that we control. The system of linear equations (3.6) can be solved using appropriate ℓ_2 or ℓ_1 methods. If an ℓ_1-norm minimization method is chosen, we would seek the sparsest vector $\boldsymbol{\rho}$ among all possible vectors satisfying (3.6).

If, instead, multiple illuminations are used to form the images, then we can use an MMV approach to find the solution with MUSIC. Indeed, note that the model matrix (3.5) can be factorized into two matrices

$$\tilde{\mathcal{A}} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_1;\omega) & \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_2;\omega) & \ldots & \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_K;\omega) \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{C}^{N\times K} \tag{3.7}$$

and

$$\Lambda_{\widehat{f}(\omega)} = \begin{pmatrix} \widehat{g}^{(1)}_{\widehat{f}(\omega)} & 0 & & \\ 0 & \widehat{g}^{(2)}_{\widehat{f}(\omega)} & & \\ & & \ddots & \\ & & 0 & \widehat{g}^{(k)}_{\widehat{f}(\omega)} \end{pmatrix} \in \mathbb{C}^{K\times K}\,, \tag{3.8}$$

so that $\mathcal{A}_{\widehat{f}(\omega)} = \tilde{\mathcal{A}}\,\Lambda_{\widehat{f}(\omega)}$. Hence, it follows from the discussion in Section 2 that (3.6) can be written in the MMV form

$$\tilde{\mathcal{A}}\tilde{\boldsymbol{\rho}}_q = \boldsymbol{b}_q\,, \quad q = 1,\ldots,S, \tag{3.9}$$

and the support of $\boldsymbol{\rho}$ can be found exactly with MUSIC if enough data vectors $\boldsymbol{b}_{\widehat{f}_q(\omega)}$ are available. In (3.9), $\boldsymbol{b}_q = \boldsymbol{b}_{\widehat{f}_q(\omega)}$, and $\tilde{\boldsymbol{\rho}}_q = \Lambda_{\widehat{f}_q(\omega)}\boldsymbol{\rho}$ represents an *effective source weighted reflectivity vector* with the same support as $\boldsymbol{\rho}$, and whose nonzero entries vary with $\widehat{\boldsymbol{f}}_q(\omega)$. We remark that the equivalent source problem (3.9) can be used to account for multiple scattering between the scatterers (see [12] for details).

To show that Theorem 2.6 is relevant for imaging we write (3.9) as (2.26) with the unknown matrix $X =$Diag$(\boldsymbol{\rho})$, the data matrix B formed by the S vectors $\boldsymbol{b}_q$, and the illumination matrix

$$L = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \tilde{\mathcal{A}}^T\widehat{\boldsymbol{f}}_1(\omega) & \tilde{\mathcal{A}}^T\widehat{\boldsymbol{f}}_2(\omega) & \ldots & \tilde{\mathcal{A}}^T\widehat{\boldsymbol{f}}_S(\omega) \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{C}^{K\times S}$$

whose ith column $\tilde{\mathcal{A}}^T\widehat{\boldsymbol{f}}_i(\omega) = [\widehat{g}^{(1)}_{\widehat{f}_i(\omega)}, \widehat{g}^{(2)}_{\widehat{f}_i(\omega)}, \ldots, \widehat{g}^{(K)}_{\widehat{f}_i(\omega)}]^T$ contains the fields at all grid positions $\vec{\boldsymbol{y}}_k$, $k = 1,\ldots,K$ due to the illumination $\widehat{\boldsymbol{f}}_i(\omega)$. Then, condition (2.27) can be interpreted as an orthogonality condition on the illuminations. Furthermore, if we suppose that $S = N$ and use the illuminations $\widehat{\boldsymbol{f}}_q(\omega) = \widehat{f}(\omega)\widehat{\boldsymbol{e}}_q$ ($\widehat{\boldsymbol{e}}_q$ is the vector with a 1 in the qth coordinate and 0's elsewhere) for all $q = 1, \dots, S$, then $L = \widehat{f}(\omega)\tilde{\mathcal{A}}^T$. In this case, $\sigma_m^T = \sigma_{\min}(L_T) \geq (1 - 3\varepsilon)|\widehat{f}(\omega)|$, assuming $\tilde{\mathcal{A}}$ satisfies conditions (i), (ii) and (iii) of Theorem 2.3 (see proof of Theorem 2.6 in Appendix B).

3.2. Multifrequency signals and one receiver: the one-dimensional problem. Consider now a one-dimensional problem with scatterers located at different ranges. To determine their positions we only use one transducer that emits and receives multiple frequency signals. We assume that the scatterers are far from the transducer, but not far from each other so the denominator of the Green's function in (3.2) can be approximated by a constant. In that case, the collected data are approximately the Fourier transform of the reflectivity vector to be imaged.

To fix ideas, denote by $z_n = L + (n-1)\Delta z$ the distance between the single transducer and the scatterer of reflectivity ρ_n, $n = 1, \dots, K$. Then,

$$\sum_{n=1}^{K} e^{i2\kappa_m z_n} \rho_n = b_m\,, \quad m = 1, \dots, 2S, \tag{3.10}$$

relates the positions and reflectivities of the scatterers to the measurements b_m at frequencies $\omega_m = \kappa_m\, c_0$, where c_0 is the wave speed in a homogeneous medium. In this problem, we seek to recover the unknown vector $\boldsymbol{\rho} = [\rho_1, \rho_2, \dots, \rho_K]$ from the multifrequency data vector $\boldsymbol{b} = [b_1, b_2, \dots, b_{2S}]$ recorded at a single receiver.

The next assumption allows to succinctly formulate one-dimensional multifrequency MUSIC in the form of an MMV problem using the Prony-type argument (see, for example, [25]). Namely, suppose that the measurements are obtained at equally spaced (spatial) frequencies $\kappa_m = \kappa_1 + (m-1)\Delta\kappa$, $m = 1, 2, \dots, 2S$. Then, we write (3.10) in matrix form as

$$\mathcal{A}_{2S}\,\boldsymbol{\rho} = \boldsymbol{b}\,, \tag{3.11}$$

where

$$\mathcal{A}_{2S} = \begin{pmatrix} e^{i2\kappa_1 z_1} & e^{i2\kappa_1 z_2} & \dots & e^{i2\kappa_1 z_K} \\ e^{i2\kappa_2 z_1} & e^{i2\kappa_2 z_2} & \dots & e^{i2\kappa_2 z_K} \\ \dots & \dots & \dots & \dots \\ e^{i2\kappa_{2S} z_1} & e^{i2\kappa_{2S} z_2} & \dots & e^{i2\kappa_{2S} z_K} \end{pmatrix} \tag{3.12}$$

is a Vandermonde matrix of dimensions $2S \times K$. Since we only have one data vector $\boldsymbol{b} \in \mathbb{C}^{2S}$ we cannot determine from it a signal space of dimension $M = |\operatorname{supp}(\boldsymbol{\rho})|$. However, following the general idea of Prony-type [32] methods we form the $S \times S$ data matrix

$$B = \begin{pmatrix} b_1 & b_2 & \dots & b_S \\ b_2 & b_3 & \dots & b_{S+1} \\ \dots & \dots & \dots & \dots \\ b_S & b_{S+1} & \dots & b_{2S} \end{pmatrix}, \tag{3.13}$$

whose rank is M if $S > M$. If we now set the $S \times K$ matrix

$$\tilde{\mathcal{A}} = \mathcal{A}_S = \begin{pmatrix} e^{i2\kappa_1 z_1} & e^{i2\kappa_1 z_2} & \dots & e^{i2\kappa_1 z_K} \\ e^{i2\kappa_2 z_1} & e^{i2\kappa_2 z_2} & \dots & e^{i2\kappa_2 z_K} \\ \dots & \dots & \dots & \dots \\ e^{i2\kappa_S z_1} & e^{i2\kappa_S z_2} & \dots & e^{i2\kappa_S z_K} \end{pmatrix} \tag{3.14}$$

and the $K \times K$ diagonal matrices

$$\Lambda_q = \begin{pmatrix} e^{i2\Delta\kappa z_1} & 0 & \dots & 0 & 0 \\ 0 & e^{i2\Delta\kappa z_2} & \dots & 0 & 0 \\ \dots & \dots & \dots & e^{i2\Delta\kappa z_{K-1}} & 0 \\ 0 & 0 & \dots & 0 & e^{i2\Delta\kappa z_K} \end{pmatrix}^q , \tag{3.15}$$

with $q = 1, \dots, S$, then it is straightforward to verify that $\tilde{\mathcal{A}}\,\Lambda_q \boldsymbol{\rho} = \boldsymbol{b}_q$, where $\boldsymbol{b}_q$ is the qth column of the matrix B in (3.13). Thus, we obtain the desired structure

$$\tilde{\mathcal{A}}\boldsymbol{\rho}_q = \boldsymbol{b}_q,$$

and MUSIC can be applied directly to find the support of $\boldsymbol{\rho}$. Subsequently, as noted above $\boldsymbol{\rho}$ itself can be determined by solving the linear system restricted on the support $\boldsymbol{\rho}$.

If $M \ll K$, so the vector $\boldsymbol{\rho}$ is M-sparse, then the solution can also be found directly from (3.11) by using an ℓ_1-norm minimization approach. Note that (3.11) always has a unique M-sparse solution if $M < S$. Indeed, we argue by contradiction that it is not possible to have more than one M-sparse solution if $M < S$. Suppose there are two M-sparse solutions $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$. Then, $\mathcal{A}_{2S}\boldsymbol{y} = 0$ for $\boldsymbol{y} = \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2$. Since the support of $\boldsymbol{y}$ is less or equal than $2M$, we have $2M$ linearly dependent columns of $\mathcal{A}_{2S}$, which is impossible for Vandermonde matrices since they are full rank.

3.3. The single frequency phase retrieval problem. In its classical form, the phase retrieval problem consists in finding a function h from the amplitude of its Fourier transform $\widehat{h}$. In imaging, it consists in finding a vector $\boldsymbol{\rho}$ that is compatible with a set of quadratic equations for measured amplitudes. This occurs in imaging regimes where only intensity data is recorded, which means that most of the information encoded in the phases is lost. Phase retrieval algorithms have been developed over a long time to deal with this problem [20, 19]. They are flexible and effective but depend on prior information about the image and can give uneven results. An alternative convex approach that guarantees exact recovery has been considered in [10, 9] but its computational cost is extremely high when the problem is large. When, however, multiple measurements of the object to be imaged are available, we may recover the missing phase information and image holographically much more efficiently [31, 28, 29]. By holographic imaging we mean the use of interference patterns between two or more coherent sources in order to form the images [40].

Indeed, let us consider single frequency imaging with multiple sources and receivers as in problem (3.9), where the data vectors $\boldsymbol{b}_q = \tilde{\mathcal{A}}\tilde{\boldsymbol{\rho}}_q$, that depend on the illumination $\widehat{\boldsymbol{f}}_q(\omega)$, contained the amplitudes and phases of the recorded signals We now, however, assume that only the amplitudes squared of the components of these data vectors can be measured. Then, the phase retrieval problem is to find the unknown vector $\boldsymbol{\rho}$ from a family of quadratic equations

$$|\mathcal{A}_q\boldsymbol{\rho}|^2 = |\boldsymbol{b}_q|^2\,, \quad q = 1, \dots, Q,$$

understood component wise. This problem is nonlinear and nonconvex and, hence, difficult to solve. In fact, it is in general NP hard [33]. However, if an appropriate set of illuminations is used, we can take advantage of the polarization identity

$$\begin{aligned} 2\,\mathrm{Re} < u, v > &= |u+v|^2 - |u|^2 - |v|^2 \\ 2\,\mathrm{Im} < u, v > &= |u-iv|^2 - |u|^2 - |v|^2 \end{aligned} \tag{3.16}$$

to solve a simple linear system of the form

$$\mathcal{A}_q \boldsymbol{\rho} = \boldsymbol{m}_q^{(r)} . \tag{3.17}$$

The polarization identity allows us to find the inner product between two complex numbers and, therefore, its phase differences. In (3.17), $\boldsymbol{m}_q^{(r)}$ is the vector whose ith component is the correlation $\overline{b_q^{(r)}} b_{\widehat{e}_i}^{(r)}$ between two signals measured at $\vec{\boldsymbol{x}}_r$, one corresponding to a general illumination $\widehat{\boldsymbol{f}}_q(\omega)$ and the other to an illumination $\widehat{\boldsymbol{e}}_i = [0, 0, \ldots, 0, 1, 0, \ldots, 0]^T$ whose entries are all zero except the ith entry which is 1. Using the polarization identity (3.16) we can obtain $\overline{b_q^{(r)}} b_{\widehat{e}_i}^{(r)}$ from linear combinations of the magnitudes (squared) $|b_q^{(r)}|^2$, $|b_{\widehat{e}_i}^{(r)}|^2$, $|b_q^{(r)} + b_{\widehat{e}_i}^{(r)}|^2$, and $|b_q^{(r)} + i b_{\widehat{e}_i}^{(r)}|^2$. A physical interpretation of (3.17) is as follows. Send an illumination $\widehat{\boldsymbol{f}}_q(\omega)$, collect the response at $\vec{\boldsymbol{x}}_r$, time reverse the received signal at $\vec{\boldsymbol{x}}_r$, and send it back to probe the medium again. Then, $\boldsymbol{m}_q^{(r)}$ represents the signals recorded at all receivers $\vec{\boldsymbol{x}}_i$, $i = 1, \ldots, N$.

To wrap up, if the phases are not measured at the array but we control the illuminations, the images can be formed by solving (3.17). We can use ℓ_1-norm minimization if only one vector $\boldsymbol{m}_q^{(r)}$ is obtained in the data acquisition process, or we can use MUSIC if enough vectors of this form are available [31, 28]. Note that in this approach, where only one frequency ω is used, the receiver $\vec{\boldsymbol{x}}_r$ is fixed.

3.4. Multiple frequency signals and multiple receivers. Finally, we consider the most general case in which multiple frequency signals are used to probe the medium from several source positions, and the echoes are measured at several receiver positions. This case considers all the possible diversity of information that can be obtained from the illuminations. We discuss first the situation in which the receivers measure amplitudes and phases and, then, the situation in which they can only measure amplitudes squared.

3.4.1. Imaging with phases. Assume that the data (including phases)

$$d(\vec{\boldsymbol{x}}_r, \vec{\boldsymbol{x}}_s, \omega_l) = P(\vec{\boldsymbol{x}}_r, \vec{\boldsymbol{x}}_s; \omega_l) , \tag{3.18}$$

for all receiver locations $\vec{\boldsymbol{x}}_r$, source locations $\vec{\boldsymbol{x}}_s$, and frequencies ω_l are available for imaging. For an array with N colocated sources and receivers that emit S different frequencies the number of measurements is then equal to $N^2 S$. To make use of the coherence of these data over all the frequencies we could stack them in a column vector $\boldsymbol{b}$, but then we would have to deal with a huge linear system $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$ of size $N^2 S \times K$. To reduce the number of data used in an ℓ_1 approach, we consider that the illumination is of separable form, i.e., $\widehat{\boldsymbol{f}}(\omega_l) = f(\omega_l)\widehat{\boldsymbol{f}}$ and the same vector $\widehat{\boldsymbol{f}}$ is used for all the frequencies ω_l, $l = 1, \ldots, S$. Thus, for an illumination $\widehat{\boldsymbol{f}} = [\widehat{\boldsymbol{f}}(\omega_1)^T, \widehat{\boldsymbol{f}}(\omega_2)^T, \ldots, \widehat{\boldsymbol{f}}(\omega_S)^T]^T$ we stack the data (including phases) in a column vector

$$\boldsymbol{b}_{\widehat{f}} = [\boldsymbol{b}^T_{\widehat{f}(\omega_1)}, \boldsymbol{b}^T_{\widehat{f}(\omega_2)}, \ldots, \boldsymbol{b}^T_{\widehat{f}(\omega_S)}]^T , \tag{3.19}$$

and we solve the system of equations

$$\mathcal{A}_{\widehat{f}} \boldsymbol{\rho} = \boldsymbol{b}_{\widehat{f}} , \tag{3.20}$$

with the $(N \cdot S) \times K$ matrix

$$
\mathcal{A}_{\widehat{f}} = \begin{pmatrix}
\uparrow & \uparrow & & \uparrow \\
\widehat{g}^{(1)}_{\widehat{f}(\omega_1)} \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_1;\omega_1) & \widehat{g}^{(2)}_{\widehat{f}(\omega_1)} \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_2;\omega_1) & \dots & \widehat{g}^{(K)}_{\widehat{f}(\omega_1)} \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_K;\omega_1) \\
\downarrow & \downarrow & & \downarrow \\
\uparrow & \uparrow & & \uparrow \\
\widehat{g}^{(1)}_{\widehat{f}(\omega_2)} \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_1;\omega_2) & \widehat{g}^{(2)}_{\widehat{f}(\omega_2)} \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_2;\omega_2) & \dots & \widehat{g}^{(K)}_{\widehat{f}(\omega_2)} \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_K;\omega_2) \\
\downarrow & \downarrow & & \downarrow \\
\vdots & \vdots & & \vdots \\
\uparrow & \uparrow & & \uparrow \\
\widehat{g}^{(1)}_{\widehat{f}(\omega_S)} \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_1;\omega_S) & \widehat{g}^{(2)}_{\widehat{f}(\omega_S)} \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_2;\omega_S) & \dots & \widehat{g}^{(K)}_{\widehat{f}(\omega_S)} \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_K;\omega_S) \\
\downarrow & \downarrow & & \downarrow
\end{pmatrix}. \tag{3.21}
$$

Here, $\widehat{g}^{(j)}_{\widehat{f}(\omega_l)} = \widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_j;\omega_l)^T \widehat{\boldsymbol{f}}(\omega_l)$ denotes the field with frequency ω_l at position $\vec{\boldsymbol{y}}_j$. The system (3.20) relates the unknown vector $\boldsymbol{\rho} \in \mathbb{C}^K$ to the data vector $\boldsymbol{b}_{\widehat{f}} \in \mathbb{C}^{(N \cdot S)}$ in a coherent way. The system of linear equations (3.20) can, of course, be solved by appropriate ℓ_2 and ℓ_1 methods.

However, because (3.20) cannot be written in the form of an MMV problem, MUSIC cannot be used to identify the support of $\boldsymbol{\rho}$ as in the previous imaging problems. The issue here is that matrix (3.21) cannot be factorized in the form $\mathcal{A}_{\widehat{f}} = \tilde{\mathcal{A}} \Lambda_{\widehat{f}}$ because the scalars $\widehat{g}^{(j)}_{\widehat{f}(\omega_l)}$ depend on frequency. However, in the paraxial regime, where the scatterers are far from the array, and the array and the IW are small so the wavefronts that illuminate the scatterers are planar, we can take into account these changes over frequencies explicitly to image coherently with MUSIC.

Indeed, assume for simplicity that only one source at $\vec{\boldsymbol{x}}_s = (\boldsymbol{x}_s, 0)$ with cross-range vector $\boldsymbol{x}_s = (x_{sx}, x_{sy})$ emits the signals, i.e., for all the frequencies ω_l we use the N-vector $\widehat{\boldsymbol{f}}(\omega_l) \equiv \widehat{\boldsymbol{f}}_{l,s} = [0, 0 \dots, 0, 1, 0, \dots, 0]^T$ with all the entries equal to zero except the sth entry which is one. In the paraxial regime, where $\lambda \ll a \ll L$ and the IW is small compared to L, the illumination at position $\vec{\boldsymbol{y}}_j = (\boldsymbol{y}_j, L + \eta_j)$ can be approximated by $\widehat{g}^{(j)}_{\widehat{f}_{l,s}} \approx e^{i\kappa_l(\eta_j + (\boldsymbol{x}_s - \boldsymbol{y}_j)^2/2L)} \approx e^{i\kappa_l \eta_j} e^{i\kappa_c (\boldsymbol{x}_s - \boldsymbol{y}_j)^2/2L}$ and, thus, $\mathcal{A}_{\widehat{f}_{l,s}} \approx \tilde{\mathcal{A}} \Lambda_{\widehat{f}_{c,s}}$ where

$$
\tilde{\mathcal{A}} = \begin{pmatrix}
\uparrow & \uparrow & & \uparrow \\
\widehat{\boldsymbol{h}}(\vec{\boldsymbol{y}}_1;\omega_1) & \widehat{\boldsymbol{h}}(\vec{\boldsymbol{y}}_2;\omega_1) & \dots & \widehat{\boldsymbol{h}}(\vec{\boldsymbol{y}}_K;\omega_1) \\
\downarrow & \downarrow & & \downarrow \\
\uparrow & \uparrow & & \uparrow \\
\widehat{\boldsymbol{h}}(\vec{\boldsymbol{y}}_1;\omega_2) & \widehat{\boldsymbol{h}}(\vec{\boldsymbol{y}}_2;\omega_2) & \dots & \widehat{\boldsymbol{h}}(\vec{\boldsymbol{y}}_K;\omega_2) \\
\downarrow & \downarrow & & \downarrow \\
\vdots & \vdots & & \vdots \\
\uparrow & \uparrow & & \uparrow \\
\widehat{\boldsymbol{h}}(\vec{\boldsymbol{y}}_1;\omega_S) & \widehat{\boldsymbol{h}}(\vec{\boldsymbol{y}}_2;\omega_S) & \dots & \widehat{\boldsymbol{h}}(\vec{\boldsymbol{y}}_K;\omega_S) \\
\downarrow & \downarrow & & \downarrow
\end{pmatrix} \tag{3.22}
$$

with $\widehat{\boldsymbol{h}}(\vec{\boldsymbol{y}}_j;\omega_l) = e^{i\kappa_l \eta_j}\widehat{\boldsymbol{g}}(\vec{\boldsymbol{y}}_j;\omega_l)$, and

$$\Lambda_{\widehat{f}_{c,s}} = \begin{pmatrix} e^{i\kappa_c(\boldsymbol{x}_s-\boldsymbol{y}_1)^2/2L} & 0 & & \\ 0 & e^{i\kappa_c(\boldsymbol{x}_s-\boldsymbol{y}_2)^2/2L} & & \\ & & \ddots & \\ & & 0 & e^{i\kappa_c(\boldsymbol{x}_s-\boldsymbol{y}_K)^2/2L} \end{pmatrix}. \tag{3.23}$$

In this approximation, the nonzero entries of the diagonal matrix (3.23) are given by the illumination relative to the central frequency κ_c. Then, the multiple-frequency MUSIC formulation is of the MMV form

$$\tilde{\mathcal{A}}\Lambda_{\widehat{f}_{c,s}}\boldsymbol{\rho} = B\,, \tag{3.24}$$

with $\tilde{\mathcal{A}}$ as in (3.22), $\Lambda_{\widehat{f}_{c,s}}$ as in (3.23), and the $(N\cdot S)\times N$ matrix

$$B = P^c = [P(\omega_1)^T, P(\omega_2)^T, \ldots, P(\omega_S)^T]^T \tag{3.25}$$

corresponding to stacking the array response data matrices (3.4) for multiple frequencies in a column. With this data structure, multiple-frequency imaging can be carried out coherently using MUSIC with the column vectors of (3.22) as the imaging vectors.

We could have used instead the alternative data structure

$$B = P^d = \begin{pmatrix} P(\omega_1) & \ldots & 0 & 0 \\ 0 & P(\omega_2) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & P(\omega_S) \end{pmatrix} \tag{3.26}$$

to image with MUSIC. However, that would be as if imaging with each frequency separately and summing up the resulting images incoherently, so there would be no significant improvement over single frequency imaging.

To summarize, multiple frequency imaging with phases can be done in all regimes by solving (3.20) with suitable ℓ_2-norm or ℓ_1-norm methods. The matrix-matrix formulation (3.24) can be used to form the images with MUSIC or using (2,1)-matrix minimization as in [12]. Recall that (3.24) is an approximate formulation, which is valid for the paraxial regime.

3.4.2. Imaging without phases. Assume now that only the intensities can be recorded at the array. In subsection 3.3 we showed that with multiple sources and multiple receivers, but a single frequency, we could recover cross correlated data from intensity-only measurements if we control the illuminations and, then, we could image holographically. In general, if several frequencies are used for imaging, we can fix one of the three possible variables $(\vec{\boldsymbol{x}}_r, \vec{\boldsymbol{x}}_s, \omega)$ and proceed similarly. For example, we can fix the receiver position $\vec{\boldsymbol{x}}_r$, and recover the multifrequency interferometric data

$$d((\vec{\boldsymbol{x}}_r, \vec{\boldsymbol{x}}_r), (\vec{\boldsymbol{x}}_s, \vec{\boldsymbol{x}}_{s'}), (\omega, \omega')) = \overline{P(\vec{\boldsymbol{x}}_r, \vec{\boldsymbol{x}}_s;\omega)}P(\vec{\boldsymbol{x}}_r, \vec{\boldsymbol{x}}_{s'};\omega') \tag{3.27}$$

for all pairs of frequencies (ω, ω') and source locations $(\vec{\boldsymbol{x}}_s, \vec{\boldsymbol{x}}_{s'})$.

To understand the type of data that we can use in this situation, let us consider one row of the $N\times(N\cdot S)$ full response matrix for multiple frequencies

$$P^r = [P(\omega_1), P(\omega_2), \ldots, P(\omega_S)]\,, \tag{3.28}$$

and denote the r-th row of this matrix by

$$\boldsymbol{p}_r = [p_{r1}, p_{r2}, \ldots, p_{rN\cdot S}]\,. \tag{3.29}$$

Here, p_{rj} with $j \equiv j(s,l) = s + (l-1)\cdot N$, denotes the received signal at $\vec{\boldsymbol{x}}_r$ when the source at $\vec{\boldsymbol{x}}_s$ sends a signal of frequency ω_l. With this notation, and denoting by the superscript $\cdot^*$ the conjugate transpose of a vector,

$$M^r = \boldsymbol{p}_r^* \boldsymbol{p}_r \tag{3.30}$$

is the rank-one matrix whose jth column corresponds to the vector $\boldsymbol{m}^r_{\widehat{e}_j}$ in the right hand side of the linear system (3.17), introduced in subsection 3.3 for single frequency imaging, but generalized here so as to account for multiple frequencies, i.e., for $l = 1, \ldots, S$. That is, the jth column of (3.30) contains the correlations of the response received at $\vec{\boldsymbol{x}}_r$ when a signal of unit amplitude and frequency ω_l is sent from $\vec{\boldsymbol{x}}_s$ to probe the medium ($j = s + (l-1)N$), with all the other responses received also at $\vec{\boldsymbol{x}}_r$ when unit amplitude signals are sent from all the sources with all the different frequencies. In short,

$$[M^r]_{ij} = \overline{p}_{ri} p_{rj} = (\boldsymbol{p}_r \widehat{\boldsymbol{e}}_i)^* \, \boldsymbol{p}_r \widehat{\boldsymbol{e}}_j\,. \tag{3.31}$$

Since M^r is rank one, all the columns are linearly dependent, so we can only use one of its columns to solve the imaging problem

$$\mathcal{A}_{\widehat{e}_j} \boldsymbol{\rho} = \boldsymbol{m}^r_{\widehat{e}_j} \tag{3.32}$$

for one $\widehat{\boldsymbol{e}}_j$, and form the images with an ℓ_2-norm or ℓ_1-norm method. The matrix $\mathcal{A}_{\widehat{e}_j}$ is given by (3.21) and, hence, the model (3.32) is exact.

Alternatively, once the matrix M^r has been obtained from intensity-only measurements, imaging can be done using the Kirchhoff migration functional

$$\mathcal{I}^{\mathrm{KM}} = diag(\mathcal{A}^*_{\widehat{e}_j} M^r \mathcal{A}_{\widehat{e}_j}). \tag{3.33}$$

The ℓ_2 images (3.33) are very robust with respect to additive measurement noise, but they are statistically unstable when imaging is done in a randomly inhomogeneous medium or when there are modeling errors due to off-grid scatterers. Both situations lead to perturbations in the (unknown) phases that may make the $\mathcal{I}^{\mathrm{KM}}$ images dependent on the particular realization of the medium and/or the positions of the scatterers. In [29], we showed that statistical stability can be enhanced by masks that limit the frequency and source offsets of the measurements used in (3.33). Hence, if the perturbations of the phases are important, we can use the Single Receiver INTerferometric (SRINT) imaging functional given by

$$\mathcal{I}^{SRINT} = diag(\mathcal{A}^*_{\widehat{e}_j} \boldsymbol{\mathcal{Z}} \odot M^r \mathcal{A}_{\widehat{e}_j})\,. \tag{3.34}$$

In (3.34), the mask $\boldsymbol{\mathcal{Z}}$ is a matrix composed by zeros and ones restricting the data to coherent nearby source locations and frequencies, and $\odot$ denotes component-wise multiplication. The same idea can be used for stabilizing the ℓ_1-norm minimization method if the perturbation of the phases are important. We can just replace the jth column of the matrix M^r by the jth column of the masked data $\boldsymbol{\mathcal{Z}} \odot M^r$, and remove the corresponding rows from the model matrix $\mathcal{A}_{\widehat{e}_j}$.

On the other hand, as noted in [31, 28], the support of the reflectivity $\boldsymbol{\rho}$ can be recovered exactly by using the MUSIC algorithm on the single frequency interferometric matrix $M(\omega) = P^*(\omega)P(\omega)$. Once the support of $\boldsymbol{\rho}$ is found, we can estimate the reflectivities by solving a trace minimization problem restricted to the support of $\boldsymbol{\rho}$ (see [10, 31] for details).

For multiple frequencies, multiple sources and multiple receivers one can use the data structure

$$M^c = \begin{pmatrix} P(\omega_1)^* P(\omega_1) \\ P(\omega_2)^* P(\omega_1) \\ \vdots \\ P(\omega_S)^* P(\omega_1) \end{pmatrix} \tag{3.35}$$

for pairs of frequencies (ω_l, ω_1), $l = 1, \ldots, S$, to image coherently using MUSIC. Indeed, the matrices M^c as in (3.35) and P^c defined in (3.25) have the same column space and, therefore, MUSIC can form the images using the SVD of M^c and the column vectors of (3.22) as imaging vectors. We denote these data structures with the superscript c to point out that we have stacked the one frequency matrices $P(\omega_l)$ and the two frequencies matrices $P(\omega_l)^* P(\omega_1)$ in a column.

As noted in the previous section we could have used instead the alternative data structure

$$M^d = \begin{pmatrix} P(\omega_1)^* P(\omega_1) & \ldots & 0 & 0 \\ 0 & P(\omega_2)^* P(\omega_2) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & P(\omega_S)^* P(\omega_S) \end{pmatrix} \tag{3.36}$$

to image using MUSIC. However, as we have already explained, if we used the SVD of M^d to obtain the signal and noise subspaces, then the frequencies are not used coherently and there is no improvement over single frequency imaging.

In summary, multiple frequency imaging with intensity-only can be done in all regimes by solving (3.32) with appropriate ℓ_2-norm or ℓ_1-norm methods or, in the paraxial regime, by forming the images using MUSIC on the data structure (3.35) with imaging vectors given by the column vectors of the matrix (3.22). MUSIC on the data structure (3.36) should not be used since multiple frequencies are not processed coherently. The performance of these methods will be assessed in Section 4, where we show numerical experiments in homogeneous and weakly inhomogeneous media.

4. Numerical Simulations.

We present here numerical simulations that illustrate the performance of the different imaging methods discussed in the previous sections. Specifically, we consider multifrequency interferometric imaging without phases discussed in subsection 3.4.2, and we present the images obtained with ℓ_1-norm minimization, SRINT, and MUSIC using the data structures M^c and M^d. Our objective is to study the robustness of these imaging methods in the presence of noise, that is perturbations in the unknown phases of the collected signals. Two types of phase perturbations are considered, systematic due to off-grid placement of the scatterers and random resulting from wave propagation in an inhomogeneous ambient medium.

4.1. Imaging setup.

We consider a typical imaging regime in optics, with a central frequency $f_0 = 600$ THz corresponding to a central wavelength $\lambda_0 = 500nm$. We use $S = 12$ equally spaced frequencies covering a total bandwidth of 30THz. In this regime, the decoherence frequency of the data Ω_d is equal to the total bandwidth. All considered wavelengths are in the visible spectrum of green light.

The size of the array is $a = 500\lambda_0$, and the distance between the array and the IW is $L = 10000\lambda_0$. The IW, whose size is $120\lambda_0 \times 60\lambda_0$, is discretized using a uniform lattice with mesh size $4\lambda_0 \times 2\lambda_0$. The medium between the array and the IW is inhomogeneous, with weak fluctuations and long correlation lengths with respect to the central wavelength. The propagation distance L is large so cumulative scattering effects are important, but not too large so the phases of the signals received at the array still maintain certain degree of coherence. In all the figures, the true locations of the scatterers are indicated with white crosses, and the length scales are measured in units of λ_0.

Again, we assume that the phases of the signals received at the array cannot be measured. Hence, only their intensities are available for imaging. These measurements are collected at only one receiver, so we can use the methods explained in subsection 3.4.2 to image interferometrically. We consider imaging in homogeneous and inhomogeneous media.

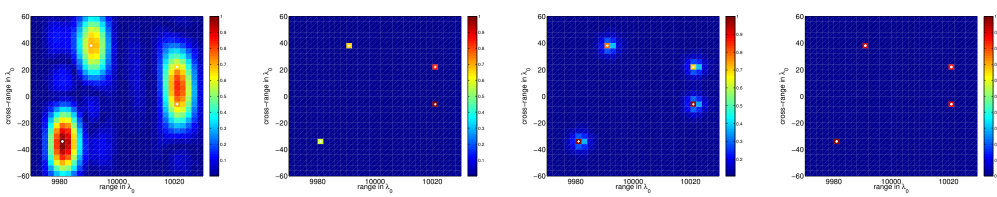

FIG. 2. *Imaging in a homogeneous medium. There is no noise added to the data and the scatterers are on the grid. From left to right: SRINT image, MUSIC with M^d, MUSIC with M^c coupling over frequencies, and ℓ_1-norm minimization applied on one column of the masked matrix $\boldsymbol{\mathcal{Z}} \odot M^r$.*

4.2. Imaging in homogeneous media. Let us first consider imaging in homogeneous media. For the imaging system described above, we expect cross-range and range resolutions of $\lambda_0 L/a = 20\lambda_0$ and $C_0/B = \lambda_0 f_0/B = 20\lambda_0$, respectively. In order to keep the resolution fixed with respect to imaging in inhomogeneous media that we consider afterwards, we also apply masks to the data used to image in the homogeneous medium. This reduces the cross-range resolution to $\lambda_0 L/X_d = 32\lambda_0$ corresponding to $X_d = 5a/8$. The range resolution does not change because the decoherence frequency Ω_d is equal to the total bandwidth.

In Figure 2, the scatterers lie on the grid and there is no noise in the data. We observe that SRINT (left image) provides a quite limited resolution and it cannot resolve two of the four scatterers. On the other hand, imaging with MUSIC (two middle images) or imaging using ℓ_1-norm minimization (right image) give much better results. MUSIC using the block-diagonal matrix M^d (second image from the left) gives exact recovery, while MUSIC using the M^c matrix (third image from the left), that couples all the frequencies, is less accurate. This is so because, as we explained in Section 3.4, MUSIC with M^c is not exact as it provides approximate locations of the scatterers only in the paraxial regime. Finally, the ℓ_1-norm approach recovers exactly the four scatterers as can be seen in the right image of this figure.

Figure 3 shows the same experiment as Figure 2 but with the scatterers displaced by half the grid size with respect to the grid points in range and cross-range directions. This produces perturbations in the unknown phases of the collected signals due to modeling errors. Because the point spread function is, in this case, much wider (of the order of $20\lambda_0$) than the off-grid displacements, the image formed with SRINT (left plot) is very robust with respect to these perturbations in the phases. However, the image obtained with MUSIC using the data structure M^d (second plot from the left) deteriorates dramatically because the multiple-frequency information contained in the data is not processed in a coherent way. On the other hand, both MUSIC with the M^c data structure (third plot from the left) and ℓ_1-norm minimization (right plot) are very robust with respect to the off-grid displacements.

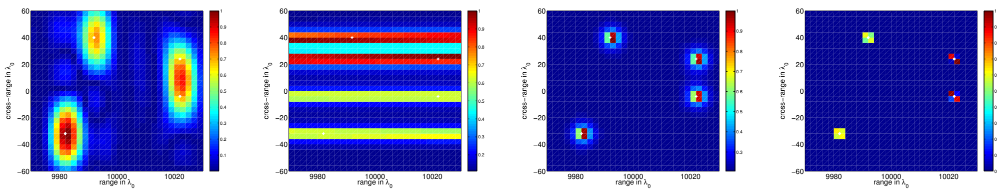

FIG. 3. *Same as Figure 2 but with the scatterers off the grid. The scatterers are displaced by half the grid size in both directions from a grid point.*

We study next the performance of the proposed methods for imaging in inhomogeneous media with weak fluctuations and long correlation lengths with respect to λ_0. The challenge is to obtain similar results in this case.

4.3. Imaging in random inhomogeneous media. Consider the setup displayed in Figure 4 with four scatterers in the right (black circles) at a distance $L = 10000\lambda_0$ from the array (black stars). The data used in the numerical experiments are generated using the random phase model which is frequently used to account for weak phase distortions [3, 13, 5, 29]. In this model, the standard deviation of the perturbations of the phases is given by $\sigma\sqrt{lL}/\lambda_0$, where σ and l denote the strength and the correlation length of the fluctuations of the medium, respectively. If we introduce the characteristic strength $\sigma_0 = \lambda_0/\sqrt{lL}$, for which the standard deviation of the random phases is $O(1)$, we can quantify the perturbations of the unknown phases by the dimensionless parameter $\varepsilon = \sigma/\sigma_0$.

In order to study the effect of phase distortions due to a random medium on imaging, we consider that the scatterers lie on the grid. Imaging in random media with ℓ_1-norm minimization has also been considered in [13, 5].

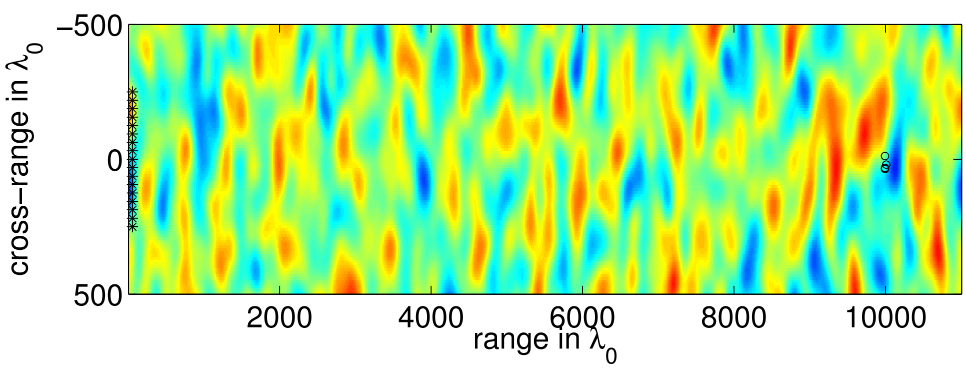

FIG. 4. *One realization of the random medium used in the simulations. The correlation length of the fluctuations is $l = 100\lambda_0$.*

Figure 5 displays the images obtained in a very weak fluctuating random medium with $\varepsilon = 0.05$. Comparing these images with the ones obtained in a homogeneous

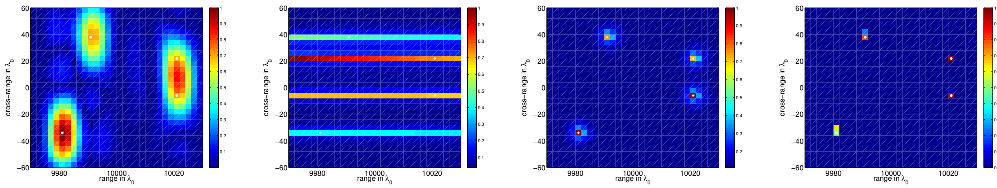

FIG. 5. *Same as Figure 2 but the medium is inhomogeneous. The strength of the fluctuations is* $\sigma = 0.5\,10^{-4}$ *which corresponds to* $\varepsilon = 0.05$. *The scatterers are on-grid.*

medium with scatterers on and off the grid (Figures 2 and 3, respectively) we observe that (i) SRINT (left plot), MUSIC using M^c (third plot from the left) and ℓ_1-norm minimiation (right plot) are stable, and (ii) MUSIC using M^d (second plot from the left) is not. Note that off-grid scatterers and a random medium both induce similar noise in the data, as both occur in the phases. In the off-grid case, the noise is systematic and similar for all array elements, while the noise induced by the random phase model depends on the path that connects the scatterer to each array element. Hence, depending on the correlation length of the random medium the noise produced in the phases is more or less correlated over the array elements.

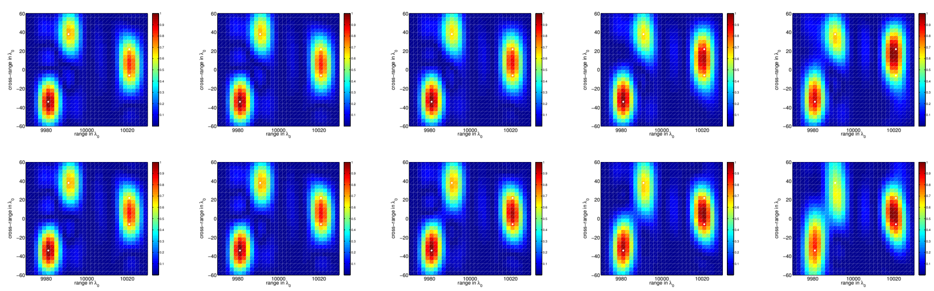

FIG. 6. *Imaging with SRINT in inhomogeneous media illustrating its stability with respect to the random fluctuations of the media. The strength of the fluctuations increases from left to right so* $\varepsilon = 0.1, 0.2, 0.4, 0.6$ *and* 0.8. *The top and bottom rows are two realizations of the random medium.*

Since MUSIC using M^d is not robust with respect to perturbations in the phases (see Figures 3 and 5) because the data are not processed coherently over frequencies, we do not present more results using this method.

To further examine the robustness of the other imaging methods with respect to random medium fluctuations, we consider in the next figures five noise levels corresponding to $\varepsilon = 0.1, 0.2, 0.4, 0.6$ and 0.8. Each figure presents results for two realizations of the random medium. In Figure 6 we see that, as expected, SRINT is highly robust, although its resolution is not very good. Even for $\varepsilon = 0.8$ (right column) the images do not change much respect to the ones obtained in a homogeneous medium. Figure 7 shows the images obtained with ℓ_1-norm minimization. The resolution is much better than that provided by SRINT, but it is much more sensitive to noise. Only for fluctuation strengths below or equal $\varepsilon = 0.2$ the images are good. Above this strength the images are useless. However, the use of masks on the data effectively removes the distortion imposed by the medium up to $\varepsilon = 0.4$, as it can be seen in Figure 8. This is so because by using masks we discard the incoherent data and, thus, we improve the robustness of the ℓ_1-norm method (even though we reduce the number of equations in the linear system by about 40%).

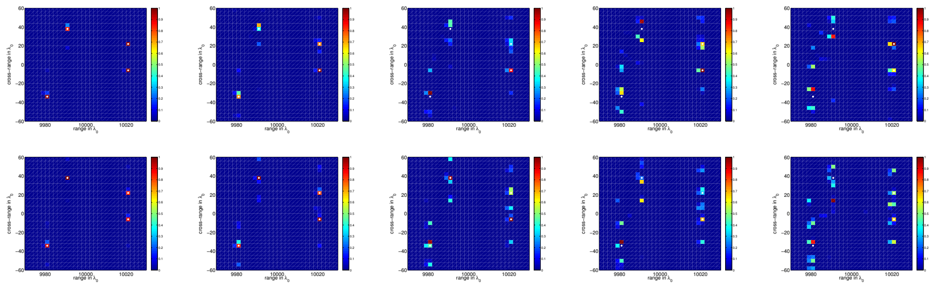

FIG. 7. *Images obtained with ℓ_1-norm minimization without masks in the same media and the same scatterer's configuration as in Figure 6. Imaging with ℓ_1-norm minimization without masks is stable only for $\varepsilon \leq 0.2$.*

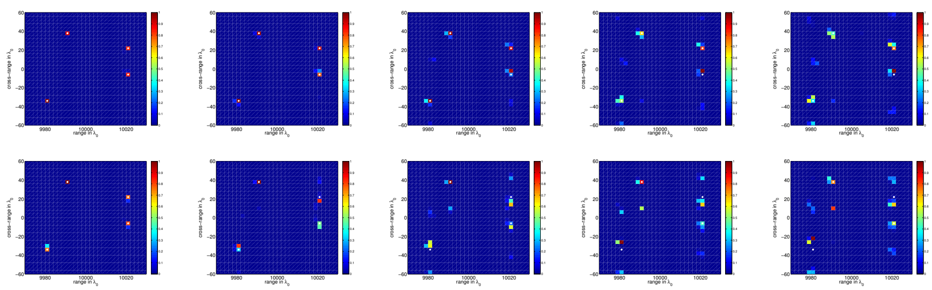

FIG. 8. *Same as Figure 7 but using masked data. The results are now stable for $\varepsilon \leq 0.4$.*

Finally, the images shown in Figure 9 formed using MUSIC with M^c are also very good. They have significantly better resolution than the SRINT images but not as good as the ones obtained with ℓ_1-norm minimization. We stress that MUSIC with M^c is not exact even for perfect data and, therefore, ℓ_1-norm minimization should be preferred if the fluctuations of the medium are weak. However, as the strength of the fluctuations increases, MUSIC with M^c becomes competitive. Observe that at lower SNR, when the ℓ_1-norm images are not usefull, MUSIC with M^c is robust and the resolution is better than the one provided by SRINT. Therefore, it should be the preferred method among the three for imaging in moderate SNR regimes.

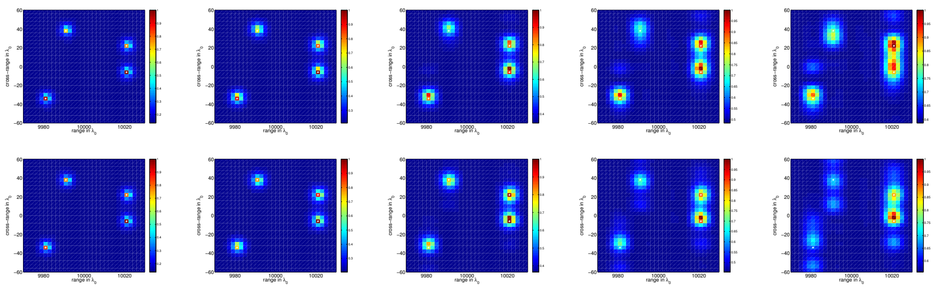

FIG. 9. *Images obtained with MUSIC using M^c in the same media and the same scatterer's configuration as in Figures 6-8. MUSIC using M^c is stable for $\varepsilon \leq 0.6$.*

5. Imaging results in the framework of Theorems 2.3 and 2.4. To illustrate the relevance of Theorems 2.3 and 2.4 for imaging, we consider in this section the equivalent source problem of active array imaging with multiple frequencies and multiple receivers described in subsection 3.4.1. In this setting we have to solve the linear system

$$\tilde{\mathcal{A}}\boldsymbol{\rho} = \boldsymbol{b}_{\widehat{f}}$$

with $\tilde{\mathcal{A}}$ the model matrix (3.22). We compare the corresponding ℓ_2 and ℓ_1 solutions of this problem for different imaging configurations. Our results illustrate the well know super-resolution for ℓ_1, meaning that $\boldsymbol{\rho}_{\ell_1}$ determines the support of the unknown $\boldsymbol{\rho}$ with higher accuracy than the conventional resolution limits, provided the assumptions of Theorem 2.3 for the noiseless case or Theorem 2.4 for the noisy case are satisfied. We also show how the bandwidth, the array size and the number of scatterers affect the vicinities defined in (2.8). The numerical results are not specialized to a paticular physical regime. They illustrate only the role of the theorems in solving the associated linear systems.

Imaging methods. We compare the solution $\boldsymbol{\rho}_{\ell_1}$ obtained with the ℓ_1-norm minimization algorithm GelMa described in section 2, and the ℓ_2-norm solution

$$\boldsymbol{\rho}_{\ell_2} = \tilde{\mathcal{A}}^* \boldsymbol{b}_{\widehat{f}}. \tag{5.1}$$

where $\tilde{\mathcal{A}}^*$ is the conjugate transpose of $\tilde{\mathcal{A}}$.

Imaging setup. The images are obtained in a homogeneous medium with an active array of $N = 37$ transducers. The ratio between the array aperture a and the distance L to IW, as well as the ratio between the bandwidth $2B$ and the central frequency f_0, vary in the numerical experiments. The IW is discretized using a uniform grid of $K = 3721$ points of size $\lambda_0/2$ in range and cross-range directions. The classical resolution theory suggests that the range and cross-range resolutions are $c_0/(2B)$ and $\lambda_0 L/a$, respectively. There is no additive noise in the data, but we consider on-grid and off-grid scatterers which produces perturbations in the recorded phases.

Imaging results. In Figure 10 we show the results obtained for a large array and a large bandwidth corresponding to $a/L = 1$ and $(2B)/f_0 = 1$. From left to right we show the $\boldsymbol{\rho}_{\ell_2}$ solution, the $\boldsymbol{\rho}_{\ell_1}$ solution, and the vicinities S_j defined in (2.8) plotted with different colors. In the top and bottom rows there are $M = 4$ and $M = 8$ scatterers, respectively. All the scatterers are on the grid and their exact locations are indicated with white crosses. The four scatterers in the top row are far apart and, therefore, their vicinities do not overlap as it can be seen in the top right image of this figure. In this case, all the conditions of Theorem 2.3 are satisfied and we find the exact locations of scatterers with the ℓ_1-norm minimization algorithm. The eight scatterers in the bottom row are closer and their vicinities are larger (according to (2.8) the size of the vicinities increases with M). We observe in the bottom right image of this figure that the vicinities overlap, so condition (2.10) is not satisfied in this case. We still, however, find the exact locations of scatterers with the ℓ_1-norm minimization algorithm which means that the conditions of Theorem 2.3 have pessimistic bounds. Because the array and the bandwidth are large, the ℓ_2-norm solutions also give very good estimates of the scatterer's locations (see the left column images).

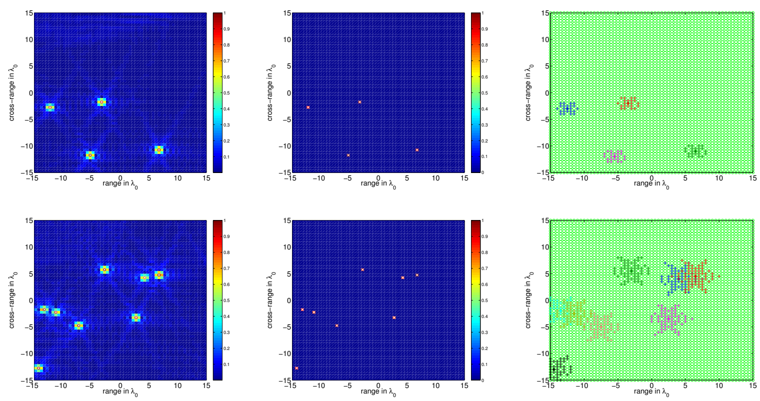

Fig. 10. *Imaging in a homogeneous medium and scatterers on grid. From left to right:* $\boldsymbol{\rho}_{\ell_2}$*,* $\boldsymbol{\rho}_{\ell_1}$*, and the vicinities* S_j*,* $j = 1, \ldots, M$*, plotted with different colours. Top row* $M = 4$*, bottom row* $M = 8$*. Large array aperture and large bandwidth so* $a/L = 1$ *and* $(2B)/f_0 = 1$*.*

In Figure 11 we show the results for the same configurations of scatterers as in Figure 10, but using a smaller array aperture and a smaller bandwidth so $a/L = 1/2$ and $(2B)/f_0 = 1/2$. Thus, the classical resolution limits become $c_0/(2B) = 2\lambda_0$ in range and $\lambda_0 L/a = 2\lambda_0$ in cross-range. Hence, the resolution of the ℓ_2-norm solutions deteriorate, as can be observed in the left column images of this figure. In fact, we only recover seven scatterers instead of eight for $M = 8$ (there are two scatterers that are quite close). The ℓ_1-norm minimization approach, however, still gives exact recovery for both $M = 4$ and $M = 8$ scatterers. This is referred to as super-resolution, which means that we can determine the location of the scatterers with a better accuracy than the classical resolution limits.

To illustrate the effect of the array and bandwidth sizes on the size of the vicinities we plot them in Figure 12 for the case $M = 4$. From left to right we plot the vicinities for $a/L = 1/2$ and $(2B)/f_0 = 1/2$, $a/L = 1/2$ and $(2B)/f_0 = 1/4$, and $a/L = 1/4$ and $(2B)/f_0 = 1/2$. As expected, cross-range and range resolutions deteriorate and consequently vicinity sizes increase as the ratios a/L and $(2B)/f_0$ decrease.

In Figure 13 we use a relatively small array and bandwidth so $a/L = 1/4$ and $(2B)/f_0 = 1/4$. In this case, the conditions of Theorem 2.3 are not satisfied for neither $M = 4$ nor $M = 8$, but the images obtained with ℓ_1-norm minimization are still very good. They are exact for $M = 4$ and very close to the true image for $M = 8$.

By further decreasing the array aperture and the bandwidth so that $a/L = 0.1$ and $(2B)/f_0 = 0.1$, we consider in Figure 14 a very challenging situation even for well separated scatterers. The ℓ_2-norm solutions shown in the left column of this figure

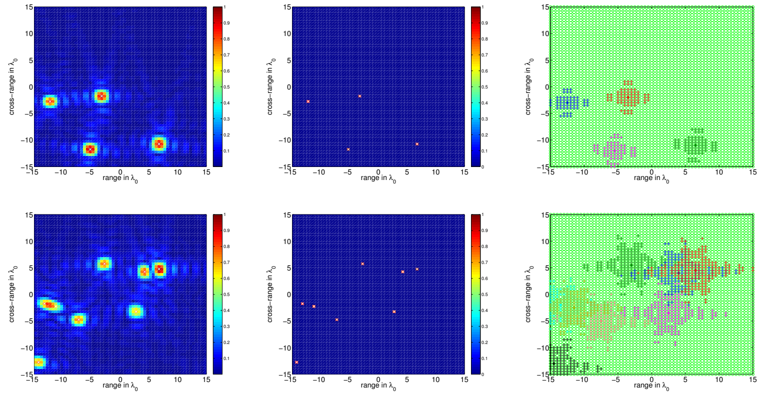

FIG. 11. *Same as Figure 10 but using a smaller array aperture and a smaller bandwidth so* $a/L = 1/2$ *and* $(2B)/f_0 = 1/2$*.*

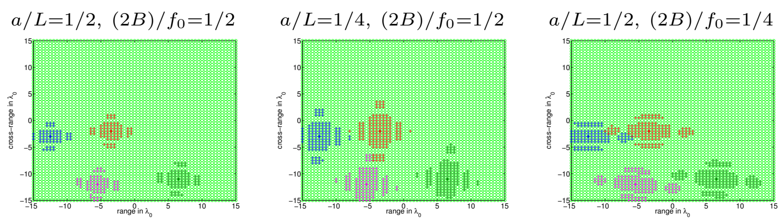

FIG. 12. *Vicinities* S_j*,* $j = 1, \dots, 4$*, for different array and bandwidth sizes. From left to right:* $a/L = 1/2$ *and* $(2B)/f_0 = 1/2$*,* $a/L = 1/2$ *and* $(2B)/f_0 = 1/4$ *and* $a/L = 1/4$ *and* $(2B)/f_0 = 1/2$*.*

are not able to locate the positions of the scatterers because of the low resolution of the imaging system. However, when the number of the scatterers is very small (see the top row corresponding to $M = 4$) the ℓ_1-norm approach provides a precise image even though the discretization of the IW is 20 times finer than the classical resolution limits of the imaging system. On the other hand, when we increase the number of scatterers to $M = 8$ (bottom row) the interaction between the vicinities is very strong and the ℓ_1-norm image in not good neither.

We now consider the same situation as in Figure 10, so the array aperture and the bandwidth are large, but with scatterers off the grid. This means that there are modeling errors and, therefore, there is not a vector $\boldsymbol{\rho}$ for which $\tilde{\mathcal{A}}\boldsymbol{\rho} = \boldsymbol{b}_{\widehat{f}}$. In the case considered next, the scatterers are displaced by $\lambda_0/4$ from a grid point in range and cross-range directions. The left column of Figure 15 shows, as expected, that the ℓ_2-norm solutions (5.1) are not affected by off-grid displacements. This is so because the resolution is larger than the displacements of the scatterers with respect to the grid points. The right column shows, however, that the ℓ_1-norm solutions are sensitive to these displacements. They are no longer exact, although they remain very close to the true solutions. By carefully examining the results of this figure we observe that the ℓ_1-norm solutions behave as it is predicted by Theorem 2.4. The coherent part of the solution is supported in the vicinities of the exact solution while the incoherent part remains very small.

Figure 16 shows similar results but for a smaller array and a smaller bandwidth. We use $a/L = 1/4$ and $(2B)/f_0 = 1/4$, so the classical resolution limits increase as

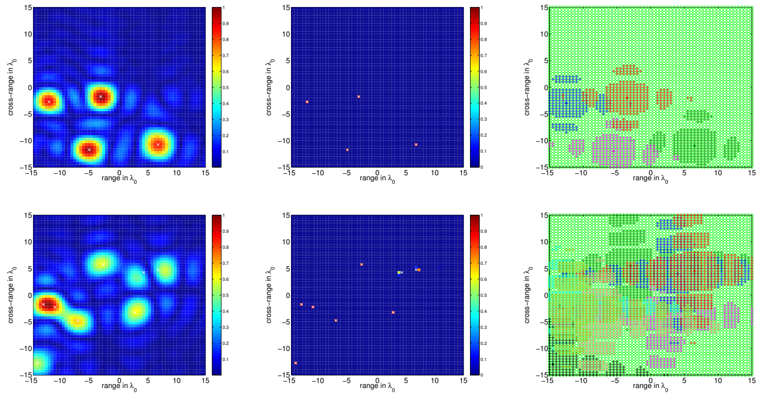

FIG. 13. *Same as Figures 10 and 11 but using a smaller array aperture and a smaller bandwidth so $a/L = 1/4$ and $(2B)/f_0 = 1/4$.*

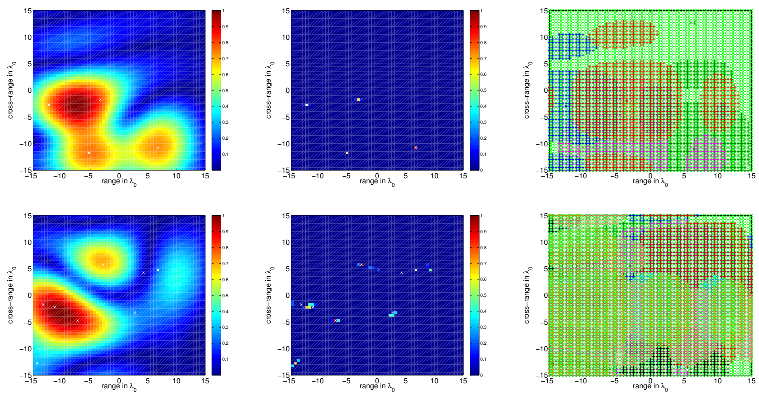

FIG. 14. *Imaging in a homogeneous medium with $a/L = 0.1$ and $(2B)/f_0 = 0.1$. Top and bottom rows: $M = 4$ and $M = 8$ scatterers, respectively. From left to right: $\boldsymbol{\rho}_{\ell_2}$ as in (5.1), $\boldsymbol{\rho}_{\ell_1}$ obtained with GelMa, and the vicinities S_j, $j = 1, \ldots, M$ plotted with different colors.*

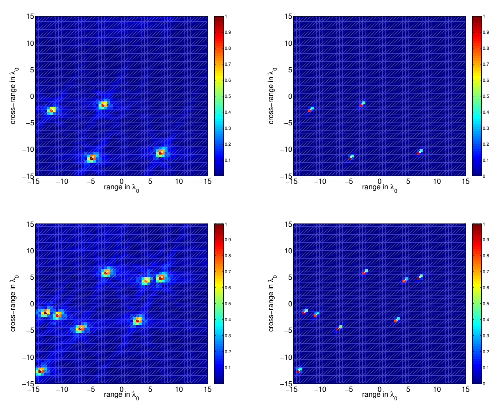

FIG. 15. *Imaging in a homogeneous medium with scatterers off the grid. As in Figure 10, we use a large array aperture and a large bandwidth so $a/L = 1$ and $(2B)/f_0 = 1$. Top and bottom rows show the images for $M = 4$ and $M = 8$ sactterers, respectively. Left and right columns show the ℓ_2-norm and ℓ_1-norm solutions, respectively.*

can be observed in the ℓ_2-norm solutions shown in the left column. As in the previous figure, the ℓ_1-norm solutions shown in the right column have a coherent part whose support is contained in the vicinities of the true solutions and an incoherent part that is very small. We also refer to [18, 4] for nice discussions about what to expect from ℓ_1-norm minimization when the scatterers do not lie on the grid.

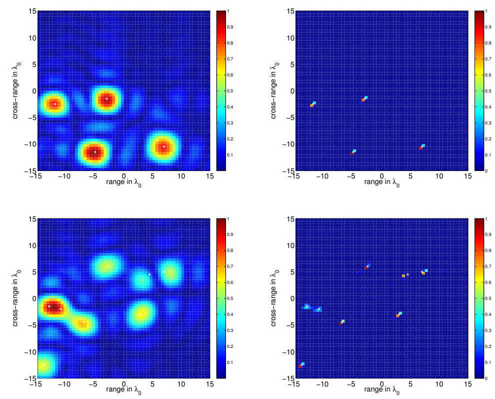

FIG. 16. *Same as Figure 15 but with $a/L = 1/4$ and $(2B)/f_0 = 1/4$.*

6. Conclusions. In this paper we addressed the question of what are appropriate data structures so as to obtain robust images with two widely used methods: ℓ_1-norm minimization and MUSIC. Both methods are well adapted to finding sparse solutions of linear underdetermined systems of equations of the form $\mathcal{A}_l \boldsymbol{\rho} = \boldsymbol{b}_l$ where $\boldsymbol{l}$ is a parameter vector that can be varied, such as the illumination profile in space and/or frequency. ℓ_1-norm minimization is well suited for solving problems with a single measurement vector corresponding to one parameter vector $\boldsymbol{l}$. On the other hand, MUSIC requires multiple measurement vectors that are obtained for several parameter vectors $\boldsymbol{l}_i$, $i = 1, \dots, S$. Given the data $\boldsymbol{b}_l$, our first main result concerns the uniqueness and robustness to noise of the minimal ℓ_1-norm solution of $\mathcal{A}_l \boldsymbol{\rho} = \boldsymbol{b}_l$. This is the subject of Theorems 2.3 and 2.4. The second important result is the key observation that MUSIC provides the exact support of the unknown $\boldsymbol{\rho}$ when the matrix $\mathcal{A}_l$ admits a factorization of the form $\mathcal{A}_l = \tilde{\mathcal{A}} \Lambda_l$ with Λ_l diagonal. Furthermore, we show in Theorem 2.6 that MUSIC is robust with respect to noise. Our third main contribution is the formulation of several common imaging configurations, including multifrequency imaging and imaging without phases, under a common linear algebra framework. For imaging without phases (the phase retrieval problem) the robustness of ℓ_1-norm minimization and MUSIC is studied with numerical simulations in weakly inhomogeneous media. Our results suggest that ℓ_1-norm minimization may be used for low noise levels while MUSIC should be the method of choice for higher noise levels.

Acknowledgments. Part of this material is based upon work supported by the National Science Foundation under Grant No. DMS-1439786 while the authors were in residence at the Institute for Computational and Experimental Research in Mathematics (ICERM) in Providence, RI, during the Fall 2017 semester. The work of M. Moscoso was partially supported by Spanish grant MICINN FIS2013-41802-R. The work of A.Novikov was partially supported by NSF grant DMS-0908011. The work of C. Tsogka was partially supported by AFOSR FA9550-17-1-0238.

REFERENCES

[1] L. Borcea, G. Papanicolaou, and C. Tsogka, *Interferometric array imaging in clutter,* Inverse Problems, 21 (2011), pp 1419–1460, 2005.

[2] L. Borcea, G. Papanicolaou, and C. Tsogka, *Adaptive interferometric imaging in clutter and optimal illumination,* Inverse Problems, 22 (2006), pp. 1405–1436.

[3] L. Borcea, J. Garnier, G. Papanicolaou, and C. Tsogka, *Enhanced statistical stability in coherent interferometric imaging,* Inverse Problems 27 (2011), 085003.

[4] L. Borcea and I. Kocyigit, *Resolution analysis of imaging with ℓ_1 optimization*, SIAM J. Imaging Sci. 8 (2015), pp. 3015–3050.

[5] L. Borcea and I. Kocyigit, *Imaging in random media with convex optimization,* SIAM J. Imaging Sci. 10 (2017), pp. 147–190.

[6] L. Borcea, M. Moscoso, G. Papanicolaou and C. Tsogka, *Synthetic aperture imaging of directional and frequency dependent reflectivity*, SIAM J. Imaging Sci., 9 (2016), pp. 52–81.

[7] E.J Candès. J.K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate information,* Communications on Pure and Applied Mathematics 59 (2006), pp. 1207–33.

[8] E.J Candès and T. Tao, *Near optimal signal recovery from random projections: universal encoding strategies?,* IEEE Trans. Inf. Theory 52 (2006), pp. 5406–25.

[9] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, *Phase Retrieval via Matrix Completion,* SIAM J. Imaging Sci. 6 (2013), pp. 199–225.

[10] A. Chai, M. Moscoso and G. Papanicolaou, *Array imaging using intensity-only measurements*, Inverse Problems 27 (2011), 015005.

[11] A. Chai, M. Moscoso and G. Papanicolaou, *Robust imaging of localized scatterers using the singular value decomposition and ℓ_1 optimization*, Inverse Problems 29 (2013), 025016.

[12] A. Chai, M. Moscoso and G. Papanicolaou, *Imaging strong localized scatterers with sparsity promoting optimization*, SIAM J. Imaging Sci., 10 (2014), pp. 1358–1387.

[13] A. Chai, M. Moscoso and G. Papanicolaou, *Array imaging of localized objects in homogeneous and heterogeneous media*, Inverse Problems 32 (2016), 104003.

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

[15] A. Devaney, E. Marengo and F. Gruber, *Time-reversal-based imaging and inverse scattering of multiply scattering point targets*, J. Acoust. Soc. Am. 118 (2005), pp. 3129–3138.

[16] D. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization*, Proceedings of the National Academy of Sciences 100 (2003), pp. 2197–2202.

[17] D. Donoho, M. Elad and V. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Information Theory 52 (2006), pp. 6–18.

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

[19] J.R. Fienup, *Phase retrieval algorithms: a comparison,* Applied Optics 21 (1982), pp. 2758–2768.

[20] R. W. Gerchberg and W. O. Saxton, *A practical algorithm for the determination of the phase from image and diffraction plane pictures,* Optik 35 (1972), pp. 237–246.

[21] I.F. Gorodnitsky, and B.D. Rao, *Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Re-weighted Minimum Norm Algorithm,* Trans. Sig. Proc. 45 (1997), pp. 600–616.

[22] R. Gribonval and M. Nielsen, *Sparse representations in unions of bases,* IEEE Transactions on Information Theory 49 (2003), pp. 3320–3325.

[23] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, John Wiley & Sons, Inc., New York, NY, USA, 1996.

[24] D. Malioutov, M. Cetin and A. Willsky, *A sparse signal reconstruction perspective for source localization with sensor arrays*, IEEE Trans. on Signal Processing 53 (2005), pp. 3010–3022.

[25] R.Griesmaier and C.Schmiedecke, *A multifrequency MUSIC algorithm for locating small inhomogeneities in inverse scattering* Inverse Problems 33 (2017) 035015 (17pp)

[26] Liao W and Fannjiang, *A MUSIC for single-snapshot spectral estimation: stability and superresolution* Appl. Comput. Harmon. Anal. 40 (2016), pp 33–67.

[27] M. Moscoso, A. Novikov, G. Papanicolaou and L. Ryzhik, A differential equations approach to l1-minimization with applications to array imaging, Inverse Problems 28 (2012).

[28] M. Moscoso, A. Novikov and G. Papanicolaou, *Coherent imaging without phases,* SIAM

J. Imaging Sci. 9 (2016), pp. 1689–1707.

[29] M. Moscoso, A. Novikov, G. Papanicolaou and C. Tsogka, *Multifrequency interferometric imaging with intensity-only measurements,* SIAM J. Imaging Sci. 10 (2017), pp. 1005–1032.

[30] V. Nikolenko, B. O. Watson, R. Araya, A. Woodruff, D. S. Peterka and R. Yuste, *SLM microscopy: scanless two-photon imaging and photostimulation with spatial light modulators,* Frontiers in Neural Circuits. 2:5 (2008), pp 1-14.

[31] A. Novikov, M. Moscoso and G. Papanicolaou, *Illumination strategies for intensity-only imaging,* SIAM J. Imaging Sci., 8 (2015), pp. 1547–1573.

[32] G. R. B. de Prony *Essai Experimentale et Analytique*, J. de L'Ecole Polytechnique 2 (1795), pp. 24–76.

[33] J. L. C. Sanz, *Mathematical Considerations for the Problem of Fourier Transform Phase Retrieval from Magnitude,* SIAM J. Appl. Math. 45 (1985), pp. 651–664

[34] R. O. Schmidt, *Multiple emitter location and signal parameter estimation*, IEEE Trans. Antennas Propag., 34 (1986), pp. 276–280.

[35] J. Tropp, *Just relax: Convex programming methods for identifying sparse signals in noise*, IEEE Trans. Information Theory 52 (2006), pp. 1030–1051.

[36] J. Tropp, A Gilbert, and M Strauss, *Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit*, Signal Processing 86 (2006), pp. 572–588.

[37] J. Tropp, *Algorithms for simultaneous sparse approximation. Part II: Convex relaxation*, Signal Processing 86 (2006), pp. 589–602.

[38] P. Wedin, *Perturbation bounds in connection with singular value decomposition.* BIT Numerical Mathematics 12 (1972), pp. 99–111.

[39] H. Weyl, *Das asymptotische Verteilungsgesetz der Eigenwert linearer partieller Differentialgleichungen (mit einer Anwendung auf der Theorie der Hohlraumstrahlung)*, Mathematische Annalen 71 (1912), pp. 441–479.

[40] E. Wolf, *Determination of the Amplitude and the Phase of Scattered Fields by Holography,* J. Opt. Soc. Am. 60 (1969), pp. 18–20.

Appendix A. Proofs of theorems 2.1 to 2.4.

THEOREM 2.1. *M-sparse solutions of $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$ are unique, if*

$$|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| < \frac{1}{2M}, \quad \forall i \neq j, \tag{A.1}$$

where we assume that the columns of matrix $\mathcal{A}$ are normalized so that $\forall i$, $\|\boldsymbol{a}_i\|_{\ell_2} = 1$.

Proof. Assume that there exist two M-sparse solutions $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ of $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$. Then their difference $\boldsymbol{z} = \boldsymbol{x}_1 - \boldsymbol{x}_2$ is at most $2M$-sparse, and $\boldsymbol{z}$ is in the kernel: $\mathcal{A}\boldsymbol{z} = 0$. This implies that there exist a 1-sparse vector $\boldsymbol{z}_1$ and a $(2M-1)$-sparse vector $\boldsymbol{z}_2$ with disjoint support such that $\boldsymbol{z}_1 - \boldsymbol{z}_2 = \boldsymbol{z}$, and

$$\|\boldsymbol{z}_1\|_{\ell^\infty} \geqslant \|\boldsymbol{z}_2\|_{\ell^\infty}. \tag{A.2}$$

This means that the vector $\boldsymbol{z}_1$ was constructed so as to contain only the largest in magnitude component of $\boldsymbol{z}$ (one of them if there are several) while $\boldsymbol{z}_2$ contains all the other components of $\boldsymbol{z}$. Suppose that the unique non-zero coordinate of $\boldsymbol{z}_1$ is i. Multiplying the identity $\mathcal{A}\boldsymbol{z}_1 = \mathcal{A}\boldsymbol{z}_2$ by $\boldsymbol{a}_i$, we get

$$\langle \boldsymbol{a}_i, \mathcal{A}\boldsymbol{z}_1 \rangle = \langle \boldsymbol{a}_i, \mathcal{A}\boldsymbol{z}_2 \rangle,$$

which reduces to

$$(\boldsymbol{z}_1)_i = \langle \boldsymbol{a}_i, \mathcal{A}\boldsymbol{z}_2 \rangle = \sum_{j=1, j\neq i}^{2M} \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle (\boldsymbol{z}_2)_j$$

Using now (A.1) we obtain

$$\|\boldsymbol{z}_1\|_{l^\infty} < \frac{1}{2M}(2M-1)\|\boldsymbol{z}_2\|_{\ell^\infty} < \|\boldsymbol{z}_2\|_{\ell^\infty},$$

which is in contradiction with (A.2). ∎

THEOREM 2.2. *M-sparse solutions of $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$ can be found as solutions of*

$$\min \|\boldsymbol{y}\|_{\ell_1}, \text{ subject to } \mathcal{A}\boldsymbol{y} = \boldsymbol{b},$$

if

$$|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| < \frac{1}{2M}, \quad \forall i \neq j,$$

where we assume that the columns of matrix $\mathcal{A}$ are normalized so that $\forall i$, $\|\boldsymbol{a}_i\|_{\ell_2} = 1$.

Proof. Assume that there exist two solutions $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ of $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$. Suppose x_1 is M-sparse, and x_2 is arbitrary. Their difference $\boldsymbol{z} = \boldsymbol{x}_1 - \boldsymbol{x}_2$ is in the kernel: $\mathcal{A}\boldsymbol{z} = 0$. We will show that $\|\boldsymbol{x}_1\|_{\ell_1} < \|\boldsymbol{x}_2\|_{\ell_1}$. Without loss of generality, we may assume that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ have disjoint support. Otherwise we decompose $\boldsymbol{z}$ in $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ such that $\boldsymbol{z} = \boldsymbol{z}_1 - \boldsymbol{z}_2$ and

$$\begin{aligned} &\operatorname{supp}(\boldsymbol{z}_1) \subset \operatorname{supp}(\boldsymbol{x}_1), \\ &\operatorname{supp}(\boldsymbol{z}_2) \cap \operatorname{supp}(\boldsymbol{x}_1) = \emptyset. \end{aligned}$$

If we assume

$$\|\boldsymbol{x}_2\|_{\ell_1} < \|\boldsymbol{x}_1\|_{\ell_1} \tag{A.3}$$

then necessarily

$$\|\boldsymbol{z}_2\|_{\ell_1} < \|\boldsymbol{z}_1\|_{\ell_1}. \tag{A.4}$$

Indeed, if $\|\boldsymbol{z}_1\|_{\ell_1} \geqslant \|\boldsymbol{x}_1\|_{\ell_1}$, it is obvious that (A.3) implies (A.4). Otherwise, if $\|\boldsymbol{z}_1\|_{\ell_1} < \|\boldsymbol{x}_1\|_{\ell_1}$ we have

$$\|\boldsymbol{z}_1 - \boldsymbol{x}_1\|_{\ell_1} \geqslant \|\boldsymbol{x}_1\|_{\ell_1} - \|\boldsymbol{z}_1\|_{\ell_1} > 0.$$

Since $\boldsymbol{x}_2 = \boldsymbol{x}_1 - \boldsymbol{z} = \boldsymbol{x}_1 - \boldsymbol{z}_1 + \boldsymbol{z}_2$ we obtain $\|\boldsymbol{x}_2\|_{\ell_1} = \|\boldsymbol{x}_1 - \boldsymbol{z}_1\|_{\ell_1} + \|\boldsymbol{z}_2\|_{\ell_1}$ and from (A.3) we get

$$\|\boldsymbol{x}_1\|_{\ell_1} > \|\boldsymbol{x}_2\|_{\ell_1} = \|\boldsymbol{x}_1 - \boldsymbol{z}_1\|_{\ell_1} + \|\boldsymbol{z}_2\|_{\ell_1},$$

which implies

$$\|\boldsymbol{z}_2\|_{\ell_1} < \|\boldsymbol{x}_1\|_{\ell_1} - \|\boldsymbol{z}_1 - \boldsymbol{x}_1\|_{\ell_1} \leqslant \|\boldsymbol{z}_1\|_{\ell_1}.$$

This finishes the proof of the statement that (A.3) implies (A.4).

We return now in the proof of the theorem and let i be the coordinate of the component of $\boldsymbol{z} = \boldsymbol{z}_1 - \boldsymbol{z}_2$ with the largest absolute value. Without loss of generality, we may suppose this component is real and positive. Then by multiplying the identity $\mathcal{A}\boldsymbol{z} = 0$ by $\boldsymbol{a}_i$ we conclude

$$\|\boldsymbol{z}\|_{l^\infty} \leqslant \frac{1}{2M} \sum_{j \neq i} |z_j| < \frac{1}{2M} \|\boldsymbol{z}\|_{\ell_1} = \frac{1}{2M} \left(\|\boldsymbol{z}_1\|_{\ell_1} + \|\boldsymbol{z}_2\|_{\ell_1} \right).$$

Since $\|\boldsymbol{z}_1\|_{\ell_1} \leqslant M \|\boldsymbol{z}_1\|_{\ell^\infty} \leqslant M \|\boldsymbol{z}\|_{\ell^\infty}$, we obtain

$$\|\boldsymbol{z}\|_{l^\infty} < \frac{1}{2} \|\boldsymbol{z}\|_{\ell^\infty} + \frac{1}{2M} \|\boldsymbol{z}_2\|_{\ell_1}. \qquad \square$$

It implies $M\|\boldsymbol{z}\|_{\ell^\infty} < \|\boldsymbol{z}_2\|_{\ell_1}$. Again using $\|\boldsymbol{z}_1\|_{\ell_1} \leqslant M\|\boldsymbol{z}\|_{\ell^\infty}$, we obtain $\|\boldsymbol{z}_1\|_{\ell_1} < \|\boldsymbol{z}_2\|_{\ell_1}$ which is in contradiction with (A.4).

THEOREM 2.3. *Let $\boldsymbol{x}$ be a solution of $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$. Let T be the index set of the support of $\boldsymbol{x}$:*

$$T = \operatorname{supp}(\boldsymbol{x}), \quad M = |T|.$$

Fix a positive $\varepsilon < 1/2$ and suppose that $\mathcal{A}$ satisfies

i. *The columns of matrix $\mathcal{A}$ are normalized so that $\forall i$, $\|\boldsymbol{a}_i\|_{\ell_2} = 1$.*

ii. *The vectors $\boldsymbol{a}_i$ in the set T are approximately orthogonal, that is they satisfy*

$$|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| < \frac{\varepsilon}{M}, \quad \forall i, j \in T, i \neq j.$$

iii. *For any $j \in T$ the vicinity S_j defined as*

$$S_j = \left\{ k \neq j | \ |\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle| \geqslant \frac{1}{2M} \right\},$$

has the properties

$$|\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle| \leqslant 1 - 2\varepsilon, \quad \forall k \in S_j$$

and

$$|\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle| < \frac{\varepsilon}{M}, \quad \forall k \in S_i, \ \forall i \neq j.$$

Then $\boldsymbol{x}$, the M-sparse solution of $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$, can be found as the solution of

$$\min \|\boldsymbol{y}\|_{\ell_1}, \ \textit{subject to } \mathcal{A}\boldsymbol{y} = \boldsymbol{b}.$$

Proof. Assume $\boldsymbol{y}$ is another solution of $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$. Then $\mathcal{A}\boldsymbol{x} = \mathcal{A}\boldsymbol{y}$. As in the proof of Theorem 2.2 we may suppose that $\boldsymbol{x}$ and $\boldsymbol{y}$ have disjoint support. For any $p \in T$ multiplying the identity $\mathcal{A}\boldsymbol{x} = \mathcal{A}\boldsymbol{y}$ by $\boldsymbol{a}_p$ we get

$$\begin{aligned} x_p + \sum_{i \in T, i \neq p} \langle \boldsymbol{a}_i, \boldsymbol{a}_p \rangle x_i &= \sum_{i \in S_p} \langle \boldsymbol{a}_i, \boldsymbol{a}_p \rangle y_i + \sum_{i \notin \cup_j S_j} \langle \boldsymbol{a}_i, \boldsymbol{a}_p \rangle y_i + \sum_{i \in S_j, j \neq p} \langle \boldsymbol{a}_i, \boldsymbol{a}_p \rangle y_i \\ &\leqslant (1 - 2\varepsilon) \sum_{i \in S_p} |y_i| + \frac{1}{2M} \sum_{i \notin \cup_j S_j} |y_i| + \frac{\varepsilon}{M} \sum_{i \in S_j, j \neq p} |y_i|. \end{aligned}$$

This implies

$$|x_p| < (1 - 2\varepsilon) \sum_{i \in S_p} |y_i| + \frac{1}{2M} \sum_{i \notin \cup_j S_j} |y_i| + \frac{\varepsilon}{M} \sum_{i \in S_j, j \neq p} |y_i| + \frac{\varepsilon}{M} \|\boldsymbol{x}\|_{\ell_1}.$$

Adding up the inequalities for all $p \in T$ we obtain

$$\|\boldsymbol{x}\|_{\ell_1} < (1 - \varepsilon) \sum_{i \in \cup_j S_j} |y_i| + \varepsilon \|\boldsymbol{x}\|_{\ell_1} + \frac{1}{2} \sum_{i \notin \cup_j S_j} |y_i|.$$

Thus

$$\|\boldsymbol{x}\|_{\ell_1} < \sum_{i \in \cup_j S_j} |y_i| + \frac{1}{2(1 - \varepsilon)} \| \sum_{i \notin \cup_j S_j} |y_i| \leqslant \|\boldsymbol{y}\|_{\ell_1}. \tag{A.5}$$

Contradiction. □

THEOREM 2.4. **Noisy case** *Let $\boldsymbol{x}$ be an M-sparse solution of*

$$\mathcal{A}\boldsymbol{x} = \boldsymbol{b},$$

and let as before T denote the index set of the support of $\boldsymbol{x}$, that is $T = \operatorname{supp}(\boldsymbol{x})$ and $M = |T|$. Fix a positive $\varepsilon < 1/2$ and suppose that $\mathcal{A}$ satisfies conditions i, ii, and iii of Theorem 2.3.

Furthermore, let $\boldsymbol{x}_\delta$ be the ℓ_1-norm minimal solution of the noisy problem

$$\min \|\boldsymbol{y}\|_{\ell_1}, \text{ subject to } \mathcal{A}\boldsymbol{y} = \boldsymbol{b}^\delta, \tag{A.6}$$

with $\boldsymbol{b}^\delta$ defined by

$$\boldsymbol{b}^\delta = \boldsymbol{b} + \boldsymbol{\delta b}.$$

We assume that the noise $\boldsymbol{\delta b}$ is bounded, that is we have

$$\|\boldsymbol{\delta b}\|_{\ell_2} \leqslant \delta,$$

for some small positive δ. We further assume that $\mathcal{A}$ has the property that the solution $\boldsymbol{\delta x}$ of

$$\min \|\boldsymbol{y}\|_{\ell_1}, \text{ subject to } \mathcal{A}\boldsymbol{y} = \boldsymbol{\delta b}, \tag{A.7}$$

satisfies

$$\|\boldsymbol{\delta x}\|_{\ell_1} \leqslant C\|\boldsymbol{\delta b}\|_{\ell_2}. \tag{A.8}$$

Then we can show that the solution $\boldsymbol{x}_\delta$ of (A.6) can be decomposed as

$$\boldsymbol{x}_\delta = \boldsymbol{x}_c + \boldsymbol{x}_i, \tag{A.9}$$

with $\boldsymbol{x}_c$ the coherent part of the solution that is supported on T or in the vicinities S_j with $j \in T$, and $\boldsymbol{x}_i$ the incoherent part of the solution which is supported away from the vicinities and is small. Specifically, for $\boldsymbol{x}_c$ we have: for any $j \in T$

$$\left|\, |(\boldsymbol{x})_j| - |(\boldsymbol{x}_c)_j + \sum_{k\in S_j} \langle a_j, a_k\rangle (\boldsymbol{x}_c)_k| \,\right| \leqslant \delta_0 + C\delta,$$

with

$$\delta_0 = \frac{2C\delta(1-\varepsilon)}{M(1-2\varepsilon)} + \frac{2\varepsilon(\|\boldsymbol{x}\|_{\ell_1} + C\delta)}{M}.$$

While for $\boldsymbol{x}_i$ we can show that:

$$\|\boldsymbol{x}_i\|_{\ell_1} \leqslant \delta_1,$$

with δ_1 given by

$$\delta_1 = C\delta + \frac{4C\delta(1-\varepsilon)}{(1-2\varepsilon)}$$

Proof. By assumption (A.7)-(A.8) there exist $\boldsymbol{\delta x}$ such that $\mathcal{A}\boldsymbol{\delta x} = \boldsymbol{\delta b}$, and $\|\boldsymbol{\delta x}\|_{\ell_1} \leqslant C\delta$. Suppose $\boldsymbol{x}$ is the M-sparse solution of $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$. Note that

$$\mathcal{A}\left(\boldsymbol{x}_\delta - \boldsymbol{\delta x}\right) = \boldsymbol{b}, \quad \mathcal{A}\left(\boldsymbol{x} + \boldsymbol{\delta x}\right) = \boldsymbol{b}^{\boldsymbol{\delta}}.$$

Since both $\boldsymbol{x}$ and $\boldsymbol{x}_\delta$ are respective minimizers, we obtain

$$\|\boldsymbol{x}\|_{\ell_1} \leqslant \|\boldsymbol{x}_\delta - \boldsymbol{\delta x}\|_{\ell_1}, \tag{A.10}$$

and

$$\|\boldsymbol{x}_\delta\|_{\ell_1} \leqslant \|\boldsymbol{x} + \boldsymbol{\delta x}\|_{\ell_1}.$$

Using the triangle inequalities

$$\|\boldsymbol{x}_\delta - \boldsymbol{\delta x}\|_{\ell_1} \leqslant \|\boldsymbol{x}_\delta\|_{\ell_1} + \|\boldsymbol{\delta x}\|_{\ell_1}, \quad \|\boldsymbol{x} + \boldsymbol{\delta x}\|_{\ell_1} \leqslant \|\boldsymbol{x}\|_{\ell_1} + \|\boldsymbol{\delta x}\|_{\ell_1}$$

we obtain

$$\|\boldsymbol{x}_\delta - \boldsymbol{\delta x}\|_{\ell_1} \leqslant \|\boldsymbol{x}_\delta\|_{\ell_1} + \|\boldsymbol{\delta x}\|_{\ell_1} \leqslant \|\boldsymbol{x} + \boldsymbol{\delta x}\|_{\ell_1} + \|\boldsymbol{\delta x}\|_{\ell_1} \leqslant \|\boldsymbol{x}\|_{\ell_1} + 2\|\boldsymbol{\delta x}\|_{\ell_1}$$

which implies

$$\|\boldsymbol{x}_\delta - \boldsymbol{\delta x}\|_{\ell_1} \leqslant \|\boldsymbol{x}\|_{\ell_1} + 2C\delta. \tag{A.11}$$

Combining (A.10) and (A.11) we conclude that

$$\|\boldsymbol{x}\|_{\ell_1} \leqslant \|\boldsymbol{x}_\delta - \boldsymbol{\delta x}\|_{\ell_1} \leqslant \|\boldsymbol{x}\|_{\ell_1} + 2C\delta. \tag{A.12}$$

For any $p \in T$, taking the inner product of

$$\mathcal{A}(\boldsymbol{x} - \boldsymbol{x}_\delta + \boldsymbol{\delta x}) = 0$$

with $\boldsymbol{a}_p$ we get

$$\begin{aligned} (\boldsymbol{x} - \boldsymbol{x}_\delta + \boldsymbol{\delta x})_p &+ \sum_{k\in T, k\neq p} \langle \boldsymbol{a}_k, \boldsymbol{a}_p\rangle (\boldsymbol{x} - \boldsymbol{x}_\delta + \boldsymbol{\delta x})_k + \sum_{k\in S_p} \langle \boldsymbol{a}_k, \boldsymbol{a}_p\rangle (\boldsymbol{\delta x} - \boldsymbol{x}_\delta)_k \\ &+ \sum_{k\in S_j, j\neq p} \langle \boldsymbol{a}_k, \boldsymbol{a}_p\rangle (\boldsymbol{\delta x} - \boldsymbol{x}_\delta)_k - \sum_{k\notin \cup S_j, k\notin T} \langle \boldsymbol{a}_k, \boldsymbol{a}_p\rangle (\boldsymbol{\delta x} - \boldsymbol{x}_\delta)_k = 0. \end{aligned} \tag{A.13}$$

Using properties (ii)-(iii) we obtain

$$\begin{aligned} |(\boldsymbol{x} - \boldsymbol{x}_\delta + \boldsymbol{\delta x})_p| \quad &< \frac{\varepsilon}{M} \sum_{k\in T, k\neq p} |(\boldsymbol{x} - \boldsymbol{x}_\delta + \boldsymbol{\delta x})_k| \\ &+ (1-2\varepsilon) \sum_{k\in S_p} |(\boldsymbol{x}_\delta - \boldsymbol{\delta x})_k| + \frac{\varepsilon}{M} \sum_{k\in S_j, j\neq p} |(\boldsymbol{x}_\delta - \boldsymbol{\delta x})_k| \\ &+ \frac{1}{2M} \sum_{k\notin \cup S_j, k\notin T} |(\boldsymbol{x}_i - \boldsymbol{\delta x})_k|. \end{aligned} \tag{A.14}$$

Summing over all $p \in T$ we get

$$\begin{aligned} \sum_{p\in T} |(\boldsymbol{x} - \boldsymbol{x}_\delta + \boldsymbol{\delta x})_p| \quad &< \varepsilon \sum_{p\in T} |(\boldsymbol{x} - \boldsymbol{x}_\delta + \boldsymbol{\delta x})_p| + (1-2\varepsilon) \sum_{k\in \cup_{p=1}^M S_p} |(\boldsymbol{x}_\delta - \boldsymbol{\delta x})_k| \\ &+ \varepsilon \sum_{k\in \cup_{p=1}^M S_p} |(\boldsymbol{x}_\delta - \boldsymbol{\delta x})_k| + \frac{1}{2} \sum_{k\notin \cup S_j, k\notin T} |(\boldsymbol{x}_i - \boldsymbol{\delta x})_k|. \end{aligned}$$

Thus

$$\begin{aligned} \sum_{k\in T} |(\boldsymbol{x} - \boldsymbol{x}_\delta + \boldsymbol{\delta x})_k| \quad &< \sum_{k\in \cup_{p=1}^M S_p} |(\boldsymbol{x}_\delta - \boldsymbol{\delta x})_k| + \frac{1}{2(1-\varepsilon)} \sum_{k\notin \cup S_j, k\notin T} |(\boldsymbol{x}_i - \boldsymbol{\delta x})_k| \\ &= \sum_{k\notin T} |(\boldsymbol{x}_\delta - \boldsymbol{\delta x})_k| - \frac{1-2\varepsilon}{2(1-\varepsilon)} \sum_{k\notin \cup S_j, k\notin T} |(\boldsymbol{x}_i - \boldsymbol{\delta x})_k| \end{aligned} \quad .$$

We therefore obtain

$$\|\boldsymbol{x}\|_{\ell_1} < \|\boldsymbol{x}_\delta - \boldsymbol{\delta x}\|_{\ell_1} - \frac{1-2\varepsilon}{2(1-\varepsilon)} \sum_{k \notin \cup S_j, k \notin T} |(\boldsymbol{x}_i - \boldsymbol{\delta x})_k|$$

By (A.12) we conclude

$$\sum_{k \notin \cup S_j, k \notin T} |(\boldsymbol{x}_i - \boldsymbol{\delta x})_k| \leqslant \frac{4C\delta(1-\varepsilon)}{1-2\varepsilon}.$$

By the triangle inequality

$$\|\boldsymbol{x}_i\|_{\ell_1} \leqslant \|\boldsymbol{\delta x}\|_{\ell_1} + \frac{4C\delta(1-\varepsilon)}{1-2\varepsilon} \leqslant C\delta + \frac{4C\delta(1-\varepsilon)}{1-2\varepsilon} = \delta_1. \tag{A.15}$$

It remains to investigate $\boldsymbol{x}_c$, the coherent part of the solution. From (A.13) we have

$$\begin{aligned}
\left| (\boldsymbol{x})_p + \sum_{k \in S_p \cup \{p\}} \langle \boldsymbol{a}_k, \boldsymbol{a}_p \rangle (\boldsymbol{\delta x} - \boldsymbol{x}_\delta)_k \right| &< \frac{\varepsilon}{M} \sum_{k \in T, k \neq p} |(\boldsymbol{x} - \boldsymbol{x}_\delta + \boldsymbol{\delta x})_k| + \frac{\varepsilon}{M} \sum_{k \in S_j, j \neq p} |(\boldsymbol{x}_\delta - \boldsymbol{\delta x})_k| \\
&\quad + \frac{1}{2M} \sum_{k \notin \cup S_j, k \notin T} |(\boldsymbol{x}_i - \boldsymbol{\delta x})_k| \\
&\leqslant \frac{\varepsilon}{M} \|\boldsymbol{x} - \boldsymbol{x}_\delta + \boldsymbol{\delta x}\|_{\ell_1} + \frac{1}{2M} \frac{4C\delta(1-\varepsilon)}{1-2\varepsilon} \\
&\leqslant \frac{\varepsilon}{M} \left(\|\boldsymbol{x}\|_{\ell_1} + \|\boldsymbol{x}_\delta\|_{\ell_1} + \|\boldsymbol{\delta x}\|_{\ell_1} \right) + \frac{2C\delta(1-\varepsilon)}{M(1-2\varepsilon)} \\
&\leqslant \frac{\varepsilon}{M} \left(2\|\boldsymbol{x}\|_{\ell_1} + 2C\delta \right) + \frac{2C\delta(1-\varepsilon)}{M(1-2\varepsilon)} = \delta_0.
\end{aligned}$$

Applying the triangle inequality:

$$\begin{aligned}
\left| (\boldsymbol{x})_p - \sum_{k \in S_p \cup \{p\}} \langle \boldsymbol{a}_k, \boldsymbol{a}_p \rangle (\boldsymbol{x}_\delta)_k \right| &\leqslant \left| (\boldsymbol{x})_p + \sum_{k \in S_p \cup \{p\}} \langle \boldsymbol{a}_k, \boldsymbol{a}_p \rangle (\boldsymbol{\delta x} - \boldsymbol{x}_\delta)_k \right| + \left| \sum_{k \in S_p} \langle \boldsymbol{a}_k, \boldsymbol{a}_p \rangle (\boldsymbol{\delta x})_k \right| \\
&\leqslant \delta_0 + C\delta,
\end{aligned}$$

we obtain the result. ∎

Appendix B. Proof of theorem 2.6.

THEOREM 2.6. Let $X = \text{Diag}(\boldsymbol{x})$ be a diagonal matrix that solves

$$\tilde{\mathcal{A}} X L = B,$$

where $\tilde{\mathcal{A}}$ satisfies conditions (i), (ii), and (iii) of Theorem 2.3 for a fixed $\varepsilon < 1/3$,

$$L = \begin{pmatrix} l_{11} & l_{12} & & l_{1S} \\ l_{21} & l_{22} & & l_{2S} \\ \vdots & \vdots & & \vdots \\ l_{K1} & l_{K2} & & l_{KS} \end{pmatrix} \in \mathbb{C}^{K \times S},$$

and B is the noiseless data matrix (2.17) with SVD $B = Q = U\Sigma V^T$. Let the perturbed matrix $B^\delta = Q^\delta + Q_0$ be such that $\sigma_{\max}(B^\delta - B) \leqslant \delta$. Suppose $\boldsymbol{x}$, the vector diagonal entries of X, is sparse with $T = \text{supp}(\boldsymbol{x})$, $M = |T|$, $M \ll \text{size}(\boldsymbol{x})$, and

$$x_m = \min_{x_i \neq 0} \{|x_i|\}.$$

Let L_T be the submatrix of L, formed by the rows corresponding to T, has

(B.1) $$\sigma_m^T = \sigma_{\min}(L_T).$$

If

(B.2) $$2\delta < x_m \sigma_m^T (1 - 3\varepsilon),$$

the orthogonal projections onto the subspaces $R(Q^\delta)$ and $R(B)$ are close:

(B.3) $$\|P_{R(Q^\delta)} - P_{R(B)}\|_{\ell_2} \leqslant \frac{\delta}{x_m \sigma_m^T (1 - 3\varepsilon)}.$$

Proof. Denote by X_T be the submatrix of X where we keep the rows that correspond to the support of $\boldsymbol{x}$. Similarly, denote by $\boldsymbol{y}_T$ be the subvector of $\boldsymbol{y}$ where we keep the entries that correspond to the support of $\boldsymbol{x}$. We claim that

(B.4) $$(1 - 3\,\varepsilon)^2 \|\boldsymbol{z}\|_{\ell_2}^2 \leqslant \|(\tilde{\mathcal{A}}^* \boldsymbol{z})_T\|_{\ell_2}^2 \leqslant (1 + 3\,\varepsilon)^2 \|\boldsymbol{z}\|_{\ell_2}^2$$

if $\boldsymbol{z} \in R(B)$. Indeed, suppose that

$$\boldsymbol{z} = \sum_{i \in T} \alpha_i \boldsymbol{a}_i.$$

Then, defining $\boldsymbol{\alpha}$ as the vector in $\mathbb{C}^K$ whose components are zero except the ith components with $i \in T$ that are equal to α_i, we get

$$\left| \|\boldsymbol{z}\|_{\ell_2}^2 - \|\boldsymbol{\alpha}\|_{\ell_2}^2 \right| = \left| \sum_{i,j \in T, i \neq j} \bar{\alpha}_i \alpha_j \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle \right| \leqslant \varepsilon \|\boldsymbol{\alpha}\|_{\ell_2}^2 \,,$$

and

$$(1 - \varepsilon) \|\boldsymbol{\alpha}\|_{\ell_2}^2 \leqslant \|z\|_{\ell_2}^2 \leqslant (1 + \varepsilon) \|\boldsymbol{\alpha}\|_{\ell_2}^2.$$

For any $j \in T$ we have

$$(\tilde{\mathcal{A}}^* \boldsymbol{z})_j = \sum_{i \in T} \alpha_i \langle \boldsymbol{a}_j, \boldsymbol{a}_i \rangle \,,$$

and, therefore,

$$\|(\tilde{\mathcal{A}}^* \boldsymbol{z})_T\|_{\ell_2}^2 = \sum_{i,j,k \in T} \bar{\alpha}_j \alpha_i \langle \boldsymbol{a}_k, \boldsymbol{a}_i \rangle \overline{\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle} \,.$$

Hence,

$$\left| \|(\tilde{\mathcal{A}}^* \boldsymbol{z})_T\|_{\ell_2}^2 - \|\boldsymbol{\alpha}\|_{\ell_2}^2 \right| \leqslant \left| \sum_{i,j,k \in T, i \neq j} \bar{\alpha}_j \alpha_i \langle \boldsymbol{a}_k, \boldsymbol{a}_i \rangle \overline{\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle} \right|$$

$$\leqslant \sum_{i,j \in T, i \neq j} \frac{|\alpha_j|^2 + |\alpha_i|^2}{2} \varepsilon \left(\frac{2}{M} + \frac{\varepsilon}{M} \right) \leqslant 3\,\varepsilon \|\boldsymbol{\alpha}\|_{\ell_2}^2.$$

Therefore,

$$(1 - 3\,\varepsilon) \|\boldsymbol{\alpha}\|_{\ell_2}^2 \leqslant \|(\tilde{\mathcal{A}}^* \boldsymbol{z})_T\|_{\ell_2}^2 \leqslant (1 + 3\,\varepsilon) \|\boldsymbol{\alpha}\|_{\ell_2}^2,$$

and we obtain

$$\frac{1 - 3\,\varepsilon}{1 + \varepsilon} \|\boldsymbol{z}\|_{\ell_2}^2 \leqslant \|(\tilde{\mathcal{A}}^* \boldsymbol{z})_T\|_{\ell_2}^2 \leqslant \frac{1 + 3\,\varepsilon}{1 - \varepsilon} \|\boldsymbol{z}\|_{\ell_2}^2,$$

which implies (B.4).

In order to compute the smallest nonzero singular value of B we observe that

$$\min_{\boldsymbol{z}\in R(B),||\boldsymbol{z}||_{\ell_2}=1} \boldsymbol{z}^* BB^* \boldsymbol{z} = \min_{\boldsymbol{z}\in R(B),||\boldsymbol{z}||_{\ell_2}=1} (\tilde{\mathcal{A}}^* \boldsymbol{z})_T^* X_T L_T L_T^* \bar{X}_T (\tilde{\mathcal{A}}^* \boldsymbol{z})_T$$

$$\geqslant (1-3\varepsilon)^2 \min_{\boldsymbol{y}\in\mathbb{C}^M ||\boldsymbol{y}||_{\ell_2}=1} \boldsymbol{y}^* X_T L_T L_T^* \bar{X}_T \boldsymbol{y} \geqslant (1-3\varepsilon)^2 x_m^2 (\sigma_m^T)^2 \,,$$

where we have used the condition (B.1). Since $\sigma_{\max}(B^\delta - B) \leqslant \delta$, we conclude that $B^\delta = Q^\delta + Q_0^\delta$, where Q^δ has M nonzero singular values, with smallest nonzero singular value

$$\sigma_{\min}(Q^\delta) \geqslant x_m \sigma_m^T (1-3\varepsilon) - \delta \,,$$

and Q_0^δ has largest singular value

$$\sigma_{\max}(Q_0^\delta) \leqslant \delta.$$

If (B.2) holds, then we can discard Q_0^δ by truncation of the singular values smaller than the noise level. We now apply Theorem 2.5 to obtain (B.3). ∎