# Imaging with highly incomplete and corrupted data

**Miguel Moscoso**[1], **Alexei Novikov**[2], **George Papanicolaou**[3], **Chrysoula Tsogka**[4]

[1] Department of Mathematics, Universidad Carlos III de Madrid, Leganes, Madrid 28911, Spain
E-mail: moscoso@math.uc3m.es

[2] Mathematics Department, Penn State University, University Park, PA 16802
E-mail: anovikov@math.psu.edu

[3] Department of Mathematics, Stanford University, Stanford, CA 94305
E-mail: papanicolaou@stanford.edu

[4] Applied Math Unit, University of California, Merced, 5200 North Lake Road, Merced, CA 95343
E-mail: ctsogka@ucmerced.edu

January 2019

**Abstract.** We consider the problem of imaging sparse scenes from a few noisy data using an $\ell_1$-minimization approach. This problem can be cast as a linear system of the form $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$, where $\mathcal{A}$ is an $N \times K$ measurement matrix. We assume that the dimension of the unknown sparse vector $\boldsymbol{\rho} \in \mathbb{C}^K$ is much larger than the dimension of the data vector $\boldsymbol{b} \in \mathbb{C}^N$, i.e, $K \gg N$. We provide a theoretical framework that allows us to examine under what conditions the $\ell_1$-minimization problem admits a solution that is close to the exact one in the presence of noise. Our analysis shows that $\ell_1$-minimization is not robust for imaging with noisy data when high resolution is required. To improve the performance of $\ell_1$-minimization we propose to solve instead the augmented linear system $[\mathcal{A}\,|\,\mathcal{C}]\boldsymbol{\rho} = \boldsymbol{b}$, where the $N \times \Sigma$ matrix $\mathcal{C}$ is a noise collector. It is constructed so as its column vectors provide a frame on which the noise of the data, a vector of dimension $N$, can be well approximated. Theoretically, the dimension $\Sigma$ of the noise collector should be $e^N$ which would make its use not practical. However, our numerical results illustrate that robust results in the presence of noise can be obtained with a large enough number of columns $\Sigma \precsim 10K$.

*Keywords*: array imaging, $\ell_1$-minimization, noise

Submitted to: *Inverse Problems*

## 1. Introduction

In this paper, we are interested in imaging problems formulated as

$$\mathcal{A}\,\boldsymbol{\rho} = \boldsymbol{b}\,, \tag{1}$$

so the data vector $\boldsymbol{b} \in \mathbb{C}^N$ is a linear transformation of the unknown vector $\boldsymbol{\rho} \in \mathbb{C}^K$ that represents the image. The model matrix $\mathcal{A} \in \mathbb{C}^{N\times K}$, which is given to us, depends on the geometry of the imaging system and the sought resolution. Typically, the linear system (1) is underdetermined because only a few linear measurements are gathered, so $N \ll K$. Hence, there exist infinitely many solutions to (1) and, thus, it is a priori not possible to find the correct one without some additional information.

We are interested, however, in imaging problems with sparse scenes. We seek to locate the positions and amplitudes of a small number $M$ of point sources that illuminate a linear array of detectors. This means that the unknown vector $\boldsymbol{\rho}$ is M-sparse, with only a few $M \ll K$ non-zero entries. Under this assumption, (1) falls under the compressive sensing framework [24, 18, 25, 10]. It follows from [18] that the unique M-sparse solution of (1) can be obtained with $\ell_1$-norm minimization when the model matrix $\mathcal{A}$ is incoherent, i.e., when its mutual coherence† is smaller than $1/(2M)$. The same result can be obtained assuming $\mathcal{A}$ obeys the M-restricted isometry property [10], which states that all sets of M-columns of $\mathcal{A}$ behave approximately as an orthonormal system.

In our imaging problems, these incoherence conditions can be satisfied only for coarse image discretizations that imply poor resolution. To retain the resolution and recover the position of the sources with higher precision we propose to extend the theory to allow for some coherence in $\mathcal{A}$. To this end, we show that uniqueness for the minimal $\ell_1$-norm solution of (1) can be obtained under less restrictive conditions on the model matrix $\mathcal{A}$. More specifically, given the columns of $\mathcal{A}$ that correspond to the support of $\boldsymbol{\rho}$, we define their vicinities as the sets of columns that are almost parallel ‡ to them. With this definition, our first result set out in Proposition 1 states that if the sources are located far enough from each other, so that their vicinities do not overlap, we can recover their positions exactly with noise-free data. Furthermore, in the presence of small noise, their position is still approximately recoverable, in the sense that most of the solution vector is supported in the vicinities while some small noise (grass) is present away from them.

This result finds interesting applications in imaging. As we explain in Section 2, in array imaging we seek to find the position of point sources that are represented as the non-zero entries of $\boldsymbol{\rho}$. Our result states under what conditions the location of these objects can be determined with high precision. It can be also used to explain super-resolution, i.e., the significantly superior resolution that $\ell_1$-norm minimization provides compared to the conventional resolution of the imaging system, i.e., the Rayleigh

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

‡ The vicinity of a column $\boldsymbol{a}_i$ is defined as the set of all columns $\boldsymbol{a}_j$ such that $|\langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle| \geq 1/(3M)$.

resolution. For instance, super-resolution has been studied using sparsity promotion for sparse spike trains recovery from band-limited measurements. Donoho [17] showed that spike locations and their weights can be exactly recovered for a cutoff frequency $f_c$ if the minimum spacing $\Delta$ between spikes is large enough, so $\Delta > 1/f_c$. Candès and Fernandez-Granda [13] showed that $\ell_1$-norm minimization guarantees the exact recovery if $\Delta > 1/2f_c$. Super-resolution has also been studied for highly coherent model matrices $\mathcal{A}$ that arise in imaging under the assumption of well-separated objects when the resolution is below the Rayleigh threshold [26, 21, 22, 5, 6]. These works include results regarding the robustness of super-resolution in the presence of noise.

Our theory also addresses the robustness to noise of the minimal $\ell_1$-norm solution. Specifically, we show that for noisy data the solution $\boldsymbol{\rho}$ can be separated into two parts: (1) the coherent part which is supported inside the vicinities, and (2) the incoherent part, usually referred to as grass, that is small and it is present everywhere. A key observation of our work is that the $\ell_1$-images get worse as $\sqrt{N}$ when there is noise in the data and, thus, $\ell_1$-norm minimization fails when the number of measurements $N$ is large. This basically follows from (12) in Proposition 1 which relates the $\ell_1$ norm of the solution to the $\ell_2$ norm of the data, so

$$\|\boldsymbol{\rho}\|_{\ell_1} \leqslant \gamma \, \|\boldsymbol{b}\|_{\ell_2}.$$

The key quantity here is the constant $\gamma$, which for usual imaging matrices $\mathcal{A}$ is proportional to $\sqrt{N}$.

To overcome this problem we introduce in Proposition 2 the noise collector matrix $\mathcal{C} \in \mathbb{C}^{N\times\Sigma}$ and propose to solve instead of the augmented linear system $[\mathcal{A}\,|\,\mathcal{C}]\boldsymbol{\rho} = \boldsymbol{b}$. The dimension of the unknown vector $\boldsymbol{\rho}$ is, thus, augmented by $\Sigma$ components which do not have any physical meaning. They correspond to fictitious sources that allow us to better approximate the noisy data. The natural question is how to build the noise collector matrix. Theoretically, the answer is given in the proof of Proposition 2 in Section 3, which is constructive. The key is that the column vectors of $[\mathcal{A}\,|\,\mathcal{C}]$ form now a frame in which the noisy vector $\boldsymbol{b}$ can be well approximated. As a consequence, we obtain a bound on the constant $\gamma$ ($\gamma < 18M^2$) which is now independent of $N$. The drawback of this construction is that we need exponentially many vectors, that is $\Sigma \lesssim e^N$. This would suggest that the noise collector may not be practical. However, the numerical experiments show that with a large enough number of columns in $\mathcal{C}$ selected at random (as i.i.d. Gaussian random variables with mean zero and variance $1/N$) the $\ell_1$-norm minimization problem is regularized and the minimal $\ell_1$-norm solution is found.

The paper is organized as follows. In Section 2, we formulate the array imaging problem. In Section 3, we present in a abstract linear algebra framework the conditions under which $\ell_1$-minimization provides the exact solution to problem (1) with and without noise. This section contains our main results. In Section 4, we illustrate with numerical simulations how our abstract theoretical results are relevant in imaging sparse sources with noisy data. Section 5 contains our conclusions.

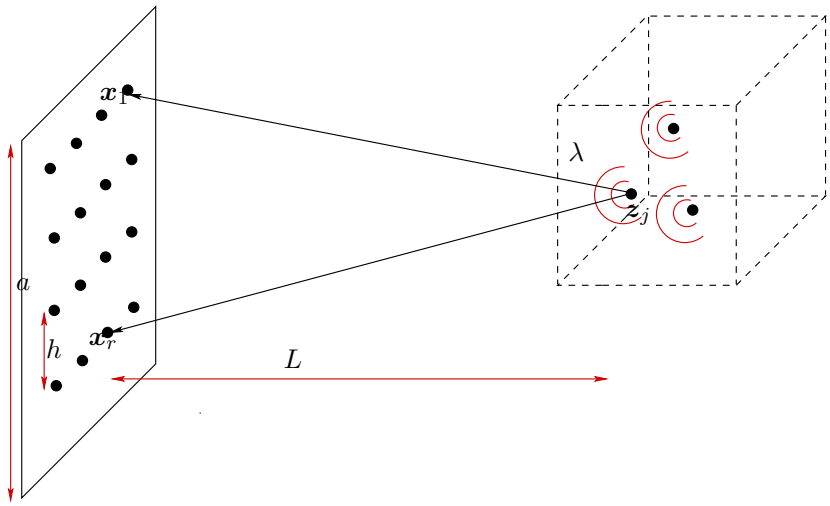


**Figure 1.** General setup of a passive array imaging problem. The sources located at $\vec{\boldsymbol{z}}_j$, $j = 1, \ldots, M$ are at distance $L$ from the array and inside the image window IW. They emit a signal and the response is recorded at all array elements $\vec{\boldsymbol{x}}_r$, $r = 1, \ldots, N$.

## 2. Passive array imaging

We consider point sources located inside a region of interest called the image window IW. The goal of array imaging is to determine their positions and amplitudes using measurements obtained on an array of receivers. The array of size $a$ has $N$ receivers separated by a distance $h$ located at positions $\vec{\boldsymbol{x}}_r$, $r = 1, \ldots, N$ (see Figure 1). They can measure single or multifrequency signals with frequencies $\omega_l$, $l = 1, \ldots, S$. The $M$ point sources, whose positions $\vec{\boldsymbol{z}}_j$ and complex-valued amplitudes $\alpha_j \in \mathbb{C}$, $j = 1, \ldots, M$, we seek to determine, are at a distance $L$ from the array. The ambient medium between the array and the sources 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 source vector*

$$\boldsymbol{\rho} = [\rho_1, \ldots, \rho_K]^\intercal \in \mathbb{C}^K \,,$$

such that

$$\rho_k = \begin{cases} \alpha_j, & \text{if } \|\vec{\boldsymbol{z}}_j - \vec{\boldsymbol{y}}_k\|_\infty < \text{ grid-size, for some } j = 1, \ldots, M, \\ 0, & \text{otherwise.} \end{cases}$$

We will not assume that the sources lie on the grid, i.e., typically $\vec{\boldsymbol{z}}_j \neq \vec{\boldsymbol{y}}_k$ for all $j$ and $k$. To write the data received on the array in a compact form, we define the Green's function vector

$$\boldsymbol{g}(\vec{\boldsymbol{y}}; \omega) = [G(\vec{\boldsymbol{x}}_1, \vec{\boldsymbol{y}}; \omega), G(\vec{\boldsymbol{x}}_2, \vec{\boldsymbol{y}}; \omega), \ldots, G(\vec{\boldsymbol{x}}_N, \vec{\boldsymbol{y}}; \omega)]^\intercal \tag{2}$$

at location $\vec{\boldsymbol{y}}$ in the IW, where $G(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}; \omega)$ denotes the free-space Green's function of the homogeneous medium. This function characterizes the propagation of a signal of

angular frequency $\omega$ from point $\vec{\boldsymbol{y}}$ to point $\vec{\boldsymbol{x}}$, so (2) represents the signal received at the array due to a point source of amplitude one, phase zero, and frequency $\omega$ at $\vec{\boldsymbol{y}}$. If the medium is homogeneous

$$G(\vec{\boldsymbol{x}},\vec{\boldsymbol{y}};\omega)=\frac{\exp\left(\mathrm{i}\frac{\omega|\vec{\boldsymbol{x}}-\vec{\boldsymbol{y}}|}{c_0}\right)}{4\pi|\vec{\boldsymbol{x}}-\vec{\boldsymbol{y}}|}\,. \tag{3}$$

The signal received at $\vec{\boldsymbol{x}}_r$ at frequency $\omega_l$ is given by

$$b(\vec{\boldsymbol{x}}_r,\omega_l)=\sum_{j=1}^{M}\alpha_j G(\vec{\boldsymbol{x}}_r,\vec{\boldsymbol{z}}_j;\omega_l). \tag{4}$$

If we normalize the columns of $\mathcal{A}$ to one and stack the data in a column vector

$$\boldsymbol{b}=\frac{1}{\sqrt{NS}}[b(\vec{\boldsymbol{x}}_1,\omega_1),b(\vec{\boldsymbol{x}}_2,\omega_1),\ldots,b(\vec{\boldsymbol{x}}_N,\omega_S)]^{\intercal}\,, \tag{5}$$

then the source vector $\boldsymbol{\rho}$ solves the system $\mathcal{A}\,\boldsymbol{\rho}=\boldsymbol{b}$, with the $(N\cdot S)\times K$ matrix

$$\mathcal{A}=\frac{1}{\sqrt{NS}}\begin{pmatrix}
\uparrow & \uparrow & & \uparrow\\
\boldsymbol{g}(\vec{\boldsymbol{y}}_1;\omega_1) & \boldsymbol{g}(\vec{\boldsymbol{y}}_2;\omega_1) & \ldots & \boldsymbol{g}(\vec{\boldsymbol{y}}_K;\omega_1)\\
\downarrow & \downarrow & & \downarrow\\
\uparrow & \uparrow & & \uparrow\\
\boldsymbol{g}(\vec{\boldsymbol{y}}_1;\omega_2) & \boldsymbol{g}(\vec{\boldsymbol{y}}_2;\omega_2) & \ldots & \boldsymbol{g}(\vec{\boldsymbol{y}}_K;\omega_2)\\
\downarrow & \downarrow & & \downarrow\\
\vdots & \vdots & & \vdots\\
\uparrow & \uparrow & & \uparrow\\
\boldsymbol{g}(\vec{\boldsymbol{y}}_1;\omega_S) & \boldsymbol{g}(\vec{\boldsymbol{y}}_2;\omega_S) & \ldots & \boldsymbol{g}(\vec{\boldsymbol{y}}_K;\omega_S)\\
\downarrow & \downarrow & & \downarrow
\end{pmatrix}:=\begin{pmatrix}
\uparrow & \uparrow & & \uparrow\\
\boldsymbol{a}_1 & \boldsymbol{a}_2 & .. & \boldsymbol{a}_K\\
\downarrow & \downarrow & & \downarrow
\end{pmatrix}. \tag{6}$$

The system $\mathcal{A}\,\boldsymbol{\rho}=\boldsymbol{b}$ relates the unknown vector $\boldsymbol{\rho}\in\mathbb{C}^K$ to the data vector $\boldsymbol{b}\in\mathbb{C}^{(N\cdot S)}$. This system of linear equations can be solved by appropriate $\ell_2$ and $\ell_1$ methods.

**Remark 1** *For simplicity of the presentation, we restricted ourselves to the passive array imaging problem where we seek to determine a distribution of sources. The active array imaging problem can be cast under the same linear algebra framework assuming the linearized Born approximation for scattering [14]. In that case, we still obtain a system of the form $\mathcal{A}_s\,\boldsymbol{\rho}=\boldsymbol{b}$, where $\boldsymbol{\rho}$ is the reflectivity of the scatterers, $\boldsymbol{b}$ is the data, and $\mathcal{A}_s$ is a model matrix for the scattering problem defined similarly to (6). Even more, when multiple scattering is not negligible the problem can also be cast as in (1); see [15] for details. Therefore, the theory presented in the next sections can be applied to the scattering problems provided that the matrix $\mathcal{A}_s$ satisfies the assumptions of Propositions 1 and 2.*

## 3. $\ell_1$ minimization-based methods

In the imaging problems considered here we assume that the sources occupy only a small fraction of the image window IW. This means that the true source vector $\boldsymbol{\rho}$ is

sparse, so the number of its entries that are different than zero, denoted by $M$, is much smaller than its length $K$. Thus, we assume $M = |\mathrm{supp}(\boldsymbol{\rho})| \ll K$. This prior knowledge changes the imaging problem substantially because we can exploit the sparsity of $\boldsymbol{\rho}$ by formulating it as an optimization problem which seeks the sparsest vector in $\mathbb{C}^K$ that equates model and data. Thus, for a measurement vector $\boldsymbol{b}$ we solve

$$\boldsymbol{\rho}_{\ell_1} = \mathrm{argmin}\, \|\boldsymbol{\rho}\|_{\ell_1}, \text{ subject to } \mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}. \tag{7}$$

Above, and in the sequel, we denote by $\|\cdot\|_{\ell_2}$, $\|\cdot\|_{\ell_1}$ and $\|\cdot\|_{\ell_\infty}$ the $\ell_2$, $\ell_1$ and $\ell_\infty$ norms of a vector, respectively.

In the literature of compressive sensing, we find the following theoretical justification of the $\ell_1$-norm minimization approach. If we assume decoherence of the columns of $\mathcal{A}$, so

$$|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| < \frac{1}{2M}, \quad \forall i \neq j, \tag{8}$$

then the $M$-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$ is unique, and it can be found as the solution of (7) [24, 18, 25]. Numerically, the $\ell_1$-norm minimization approach works under less restrictive conditions than the decoherence condition (8) suggests. In fact, our imaging matrices almost never satisfy (8).

Consider a typical imaging regime with the central wavelength $\lambda_0$. Assume we use $S = 36$ equally spaced frequencies covering a bandwidth that is 10% of the central frequency. The size of the array is $a$ and the distance between the array and the IW is $L = a$. An IW of size is $30\lambda_0 \times 30\lambda_0$ is discretized using a uniform grid with mesh size $\lambda_0/2 \times \lambda_0/2$. For such parameters, *every* column vector $\boldsymbol{a}_i$ has at least *sixty two* other column vectors $\boldsymbol{a}_j$ so that $|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| \geqslant 1/16$. Thus, our matrices are fairly far from satisfying the decoherence condition (8) if we want to recover, say, 8 sources. Numerically, however, the $\ell_1$ minimization works flawlessly.

Physically, a pair of columns $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$ are *coherent*, so $|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| \approx 1$, if the corresponding grid-points in the image are close to each other. In other words, when $\boldsymbol{a}_i$ lies in a vicinity of $\boldsymbol{a}_j$ (and vice versa). We assume, though, that the sources are far apart and, thus, the set of columns indexed by the support of the true source vector $\boldsymbol{\rho}$ does satisfy the decoherence condition (8). The above observation motivates the following natural conjecture. Perhaps, the $\ell_1$ minimization works well because it suffices to satisfy (8) only on the support of $\boldsymbol{\rho}$. Our theoretical results support this conjecture.

### *3.1. Main results*

When data is perturbed by small noise, the following qualitative description of the $\ell_1$ image (7) could be observed. Firstly, some pixels close to the points where the sources are located become visible. Secondly, a few pixels away from the sources are also visible. The latter is usually referred to as grass. In order to quantify the observed results we need to modify the decoherence condition (8) and introduce the vicinities.

**Definition 1** *Let $\boldsymbol{\rho} \in \mathbb{C}^K$ be an $M$-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$, with support $T = \{i : \rho_i \neq 0\}$ †. For any $j \in T$ define the corresponding vicinity of $\boldsymbol{a}_j$ as*

$$S_j = \left\{ k \ : \ |\langle \boldsymbol{a}_k, \boldsymbol{a}_j \rangle| \geqslant \frac{1}{3M} \right\}. \tag{9}$$

*For any vector $\boldsymbol{\eta} \in \mathbb{C}^K$ its coherent misfit to $\boldsymbol{\rho}$ is*

$$\mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\eta}) = \sum_{j \in T} \left| \rho_j - \sum_{k \in S_j} \langle \boldsymbol{a}_j, \boldsymbol{a}_k \rangle \eta_k \right|, \tag{10}$$

*whereas its incoherent remainder with respect to $\boldsymbol{\rho}$ is*

$$\mathbf{In}(\boldsymbol{\rho}, \boldsymbol{\eta}) = \sum_{k \notin \Upsilon} |\eta_k|, \quad \Upsilon = \cup_{j \in T} S_j. \tag{11}$$

**Proposition 1** *Let $\boldsymbol{\rho}$ be an $M$-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$, and let $T$ be its support. Suppose the vicinities $S_j$ from Definition 1 do not overlap, and let $\gamma > 0$ be defined as*

$$\gamma = \sup_{\boldsymbol{c}} \frac{\|\boldsymbol{\xi}\|_{\ell_1}}{\|\boldsymbol{c}\|_{\ell_2}}, \ \textit{where } \boldsymbol{\xi} \textit{ is the minimal } \ell_1\textit{−norm solution of } \mathcal{A}\boldsymbol{\xi} = \boldsymbol{c}. \tag{12}$$

*Let $\boldsymbol{\rho}_\delta$ be the minimal $\ell_1$-norm solution of the noisy problem*

$$\min \|\boldsymbol{\rho}_\delta\|_{\ell_1}, \ \textit{subject to } \mathcal{A}\boldsymbol{\rho}_\delta = \boldsymbol{b}_\delta, \tag{13}$$

*with $\|\boldsymbol{b} - \boldsymbol{b}_\delta\|_{\ell_2} \leqslant \delta$. Then,*

$$\mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\rho}_\delta) \leqslant 3\gamma\delta, \tag{14}$$

*and*

$$\mathbf{In}(\boldsymbol{\rho}, \boldsymbol{\rho}_\delta) \leqslant 5\gamma\delta. \tag{15}$$

*If $\delta = 0$, and $\Upsilon$ does not contain collinear vectors, we have exact recovery: $\boldsymbol{\rho}_\delta = \boldsymbol{\rho}$.*

Proposition 1 is proved in Appendix A. As it follows from this proof, our pessimistic bound $1/(3M)$ could be sharpened to the usual bound (8) found in the literature. We did not strive to obtain sharper results because it will make the proofs more technical and, more importantly, because the concept of vicinities describes well the observed phenomena in imaging with this bound.

When there is no noise so $\delta = 0$, Proposition 1 tells us that the M-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$ can be recovered exactly by solving the $\ell_1$ minimization problem under a less stringent condition than (8). Note that we allow for the columns of $\mathcal{A}$ to be close to collinear. When there is noise so $\delta \neq 0$, this Proposition shows that if the data $\boldsymbol{b}$ is not exact but it is known up to some bounded vector, the solution $\boldsymbol{\rho}_\delta$ of the minimization problem (13) is close to the solution of the original (noiseless) problem in the following sense. The solution $\boldsymbol{\rho}_\delta$ can be separated into two parts: the coherent part supported in the vicinities $S_j$ of the true solution, $j \in T$, and the incoherent part, which is small for

† Below and in the rest of the paper the notation $\rho_i$ means the $i$th entry of the vector $\boldsymbol{\rho}$. In contrast, we use the notation $\boldsymbol{\rho}_i$ to represent the $i$th vector of a set of vectors.

low noise, and that is supported away from these vicinities. Other stability results can be found in [10, 11, 19, 34, 26, 21, 22, 5].

Let us now make some comments regarding the relevance of this result in imaging. Vicinities, as defined in (9), are related to the classical $\ell_2$-norm resolution theory. Indeed, recall Kirchhoff migration imaging given by the $\ell_2$-norm solution

$$\boldsymbol{\rho}_{\ell_2} = \mathcal{A}^* \boldsymbol{b}, \tag{16}$$

where $\mathcal{A}^*$ is the conjugate transpose of $\mathcal{A}$. Note that (16) is an $\ell_2$ solution because it is an approximation of the least-squares solution obtained via the normal equations [3, 4]. The approximation consists in assuming that $\mathcal{A}^*\mathcal{A}$ is close to the identity matrix. Typically $\mathcal{A}^*\mathcal{A}$ is close to a diagonal matrix in which case (16) has to be modified accordingly to provide the correct amplitude of $\boldsymbol{\rho}_{\ell_2}$. Since $\boldsymbol{b} = \mathcal{A}\boldsymbol{\rho}$ the resolution analysis of KM relies on studying the behaviour of the inner products $|\langle \boldsymbol{a}_i, \boldsymbol{a}_k \rangle|$. We know from classical resolution analysis [7] that the inner products $|\langle \boldsymbol{a}_i, \boldsymbol{a}_k \rangle|$ are large for points $\boldsymbol{y}_k$ that fall inside the support of the KM point spread function, whose size is $\lambda L/a$ in cross-range (parallel to the array) and $c/B$ in range (perpendicular to the array). Given the definition of the vicinities (9), we expect the size of the vicinities to be proportional to these classical resolution limits, with an appropriate scaling factor that is inversely proportional to the sparsity $M$. This intuition is confirmed by our numerical simulations in Section 4 (see Figure 4).

Under this perspective, one could argue that Proposition 1 tells us the well-known result that a good reconstruction can be obtained for well-separated sources. Proposition 1, however, gives us more information, it provides an $\ell_1$-norm resolution theory for imaging: when vicinities do not overlap, there is a single non-zero element of the source associated within each vicinity. Permitting the columns of $\mathcal{A}$ to be almost collinear inside the vicinities allows for a fine discretization inside the vicinities and therefore the source can be recovered with very high precision. Furthermore, recovery is *exact* for noiseless data.

The assumptions in Proposition 1 are sufficient conditions but not necessary. Our numerical simulations illustrate *exact recovery* in more challenging situations, where the vicinities are not well separated (see Figure 3).

For noisy data, Proposition 1 says that it is the concept of vicinities that provides an adequate framework to look at the error between the true solution and the one provided by the $\ell_1$-norm minimization approach. Specifically, the error is controlled by the coherent misfit (10) and the incoherent remainder (11), which are shown to be small when the noise is small in $\ell_2$. This means that the reconstructed source is supported mainly in the vicinities $S_j$ of the true solution, $j \in T$, and the grass in the image is low, i.e., the part of the solution supported away from the vicinities $S_j$ is small.

Proposition 1 implies that a key to control the noise is the constant $\gamma$ defined in (12). In general, we have $\gamma = O(\sqrt{N})$. Indeed, let $\boldsymbol{y}$ be the minimum $\ell_2$-norm solution of the problem $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$ such that its support has at most size $N$. Let $\mathcal{A}_y$ be the submatrix of $\mathcal{A}$ that contains the columns that correspond to the non-zero entries of $\boldsymbol{y}$. Then, the

minimum $\ell_1$ solution $\boldsymbol{\rho}$ satisfies (by Cauchy-Schwartz $\|\boldsymbol{x}\|_{\ell_1} \leqslant \sqrt{N}\|\boldsymbol{x}\|_{\ell_2}, \forall \boldsymbol{x} \in \mathbb{C}^N$)

$$\|\boldsymbol{\rho}\|_{\ell_1} \leq \|\boldsymbol{y}\|_{\ell_1} \leqslant \sqrt{N}\|\boldsymbol{y}\|_{\ell_2} \leqslant \sqrt{N}\left\|(\mathcal{A}_y^*\mathcal{A}_y)^{-1}\mathcal{A}_y^*\right\|_{\ell_2}\|\boldsymbol{b}\|_{\ell_2}.$$

Assuming decoherence of the columns of $\mathcal{A}_y$, we conclude that $\left\|(\mathcal{A}_y^*\mathcal{A}_y)^{-1}\mathcal{A}_y^*\right\|_{\ell_2} \leqslant C$, with $C$ independent of $N$. Thus $\|\boldsymbol{\rho}\|_{\ell_1} \leqslant C\sqrt{N}\|\boldsymbol{b}\|_{\ell_2}$. A similar lower bound arises if, for example, $\mathcal{A}$ is invertible. This means that typically the quality of the image deteriorates as the number of measurements $N \to \infty$. The remedy that we propose to this is to augment the imaging matrix $\mathcal{A}$ with a "noise collector" $\mathcal{C}$ as described in the following Proposition.

**Proposition 2** *There exists a $N \times \Sigma$ noise collector matrix $\mathcal{C}$, with $\Sigma \lesssim e^N$, such that the columns of the augmented matrix $\mathcal{D} = [\mathcal{A}\,|\,\mathcal{C}]$ satisfy $\|\boldsymbol{d}_j\| = 1$,*

$$|\langle \boldsymbol{a}_i, \boldsymbol{c}_j\rangle| < \frac{1}{3M} \quad \forall i \text{ and } j\,, \tag{17}$$

$$|\langle \boldsymbol{c}_i, \boldsymbol{c}_j\rangle| < \frac{1}{3M} \quad \forall i \neq j, \tag{18}$$

*and there is a positive constant*

$$\gamma \leqslant 18M^2\,, \tag{19}$$

*such that*

$$\forall\, \boldsymbol{b},\ \exists\ \boldsymbol{\rho} \text{ such that } \mathcal{D}\boldsymbol{\rho} = \boldsymbol{b} \text{ and } \|\boldsymbol{\rho}\|_{\ell_1} \leqslant \gamma\|\boldsymbol{b}\|_{\ell_2}. \tag{20}$$

**Proof:** Let $\boldsymbol{d}_i = \boldsymbol{a}_i$, for $i = 1, \ldots, K$. We will construct iteratively a sequence of vectors $\boldsymbol{d}_{K+1} = \boldsymbol{c}_1$, $\boldsymbol{d}_{K+2} = \boldsymbol{c}_2$, ..., $\boldsymbol{d}_{K+\Sigma} = \boldsymbol{c}_\Sigma$ such that for each $s = 1 \ldots \Sigma$

$$|\langle \boldsymbol{d}_k, \boldsymbol{d}_{K+s}\rangle| \leqslant \frac{1}{3M},\ \forall k < s + K.$$

The iteration will terminate at a finite step, say, $\Sigma$. At the termination step we will have that for any $\boldsymbol{b}$, $\|\boldsymbol{b}\|_{\ell_2} = 1$ there exists $k \leqslant \Sigma + K$ such that

$$|\langle \boldsymbol{d}_k, \boldsymbol{b}\rangle| > \frac{1}{3M}. \tag{21}$$

The finite time termination is a consequence of a volume growth estimate. Namely, if (18) holds for all $i \neq j \leqslant \Sigma$, then the points $\boldsymbol{c}_i$, $i = 1, 2, \ldots \Sigma$ are centers of non-overlapping balls of radius $r$. The radius is bounded below:

$$r > \frac{1}{2}\alpha, \text{ where } \alpha = \sqrt{1 - \frac{1}{9M^2}}.$$

Thus the iteration will terminate at a finite step. Furthermore, if $r < \sqrt{2}$ then the number $\Sigma \lesssim e^{N \log \frac{\sqrt{2}}{r}}$ as the dimension $N \to \infty$, because $(r/\sqrt{2})^N\Sigma \sim 1$.

Let us finally estimate $\gamma$ in (19). Without loss of generality, we may assume $\|\boldsymbol{b}\|_{\ell_2} = 1$. By our construction, there exists $k \leqslant \Sigma + K$ such that (21) holds. Thus we can choose $\boldsymbol{d}_{n_1}$ and $c_1$ so that $|c_1| \leqslant 1$ and $\boldsymbol{b}_1 = \boldsymbol{b} - c_1\boldsymbol{d}_{n_1}$ satisfies $\|\boldsymbol{b}_1\|_{\ell_2} \leqslant \alpha$.

Using (21) inductively we can find a sequence $\{\boldsymbol{d}_{n_i}\}_{i=1}^{\infty}$, and a sequence $\{c_i\}_{i=1}^{\infty}$, so that $|c_i| \leqslant \alpha^{i-1}$ and the vectors $\boldsymbol{b}_n = \boldsymbol{b} - \sum_{i=1}^{n} c_i \boldsymbol{d}_{n_i}$ satisfy $\|\boldsymbol{b}_n\|_{\ell_2} \leqslant \alpha^n$. Therefore,

$$\boldsymbol{b} = \sum_{i=1}^{\infty} c_i \boldsymbol{d}_{n_i} \tag{22}$$

and

$$\|\boldsymbol{\rho}\|_{\ell_1} \leqslant \sum_{i=1}^{\infty} |c_i| \leqslant \sum_{i=1}^{\infty} \alpha^{i-1} = \frac{1}{1-\alpha} \leqslant 18M^2 \tag{23}$$

by the triangle inequality. □

Proposition 2 is an important result as it shows that the constant $\gamma$ in (12) can be made independent of $N$ by augmenting the columns of the linear system with columns of a noise collector matrix $\mathcal{C}$. The columns of $\mathcal{C}$ are required to be decoherent to the columns of $\mathcal{A}$ (see (17)), and decoherent between them (see (18)). Recalling that the columns of $\mathcal{A}$ for the imaging problem are Green's vectors corresponding to points in the imaging window, we stress that the columns of $\mathcal{C}$ do not admit a physical interpretation. They do not correspond to any points in the imaging window or elsewhere. Similarly, the $\Sigma$ last components of the augmented unknown vector $\boldsymbol{\rho}$ in (20) do not have a physical meaning. They correspond to fictitious auxiliary unknowns that are introduced to regularize the $\ell_1$-norm minimization problem.

The drawback of this theory is that the size of the noise collector is exponential $\Sigma \lesssim e^N$. This makes it impractical. Our numerical experiments, however, indicate great improvement in the performance of $\ell_1$-norm minimization with $\Sigma \lesssim 10K$ when the columns of $\mathcal{C}$ are selected at random (its entries are i.i.d. Gaussian random variables with mean zero and variance $1/N$). This works well for additive mean zero uncorrelated noise. For other types of noise, the idea is to construct a library that represents the values that the noise vector $\delta\boldsymbol{b}$ takes. It is the elements of this library that should be used as columns of the noise collector matrix $\mathcal{C}$. A different approach can be followed when the noise $\delta\boldsymbol{b}$ is sparse so its $\ell_1$-norm is small. Then, $\mathcal{C}$ could be simply taken as the $N \times N$ identity matrix $I$. This approach has been proposed and analyzed in [27] and provides exact recovery for sparse noise vectors $\delta\boldsymbol{b}$.

In the next section, we present numerical results to illustrate the relevance of our theory in imaging sparse sources. We focus our attention in the case of additive mean zero uncorrelated noise which is not sparse. The results show a dramatic improvement using the noise collector.

## 4. Imaging results in the framework of Propositions 1 and 2

We illustrate here the relevance of Propositions 1 and 2 in imaging. We compare $\boldsymbol{\rho}_{\ell_1}$, the $\ell_1$-norm solution of (7) and the $\ell_2$-norm Kirchhoff migration solution (16). Our results illustrate:

(i) The well-known super-resolution for $\ell_1$, 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 Proposition 1 are satisfied.

(ii) The equally well known sensitivity of $\ell_1$ to additive noise. This is made more precise in the imaging context where the constant $\gamma$ in (12) grows with the number of measurements as $\sqrt{NS}$, where $NS$ is the total number of measurements acquired by $N$ receivers at $S$ frequencies. We observe that, for a given level of noise, the $\ell_1$-norm reconstruction deteriorates as the number of measurements increases.

(iii) The noise collector matrix $\mathcal{C}$ stabilizes $\ell_1$-norm minimization in the presence of noise.

We also show how the bandwidth, the array size, and the number of sources affect the vicinities defined in (9). The numerical results are not specialized to a particular physical regime. They illustrate only the role of the Propositions 1 and 2 in solving the associated linear systems.

*Numerical method to solve the $\ell_1$ minimization problem (7)*

The solution of (7) can be found accurately and efficiently by many numerical minimization techniques. Here are some of them: orthogonal matching pursuit [9], homotopy [33, 32, 20], interior-point methods [1, 37], gradient projection [23], sub-gradient descent methods in primal and dual spaces [29, 8], and proximal gradient in combination with iterative shrinkage-thresholding [30, 31, 2]. In this work we chose to solve (7) using the Generalized Lagrange Multiplier Algorithm (GeLMA) [28], a semi-implicit version of the primal-dual method [16].

The formulation of GeLMA starts with a standard optimization argument that the solution of (7) equals the solution of the following min-max problem. Define the function

$$F(\boldsymbol{\rho}, \boldsymbol{z}) = \tau \|\boldsymbol{\rho}\|_{\ell_1} + \frac{1}{2}\|\mathcal{A}\boldsymbol{\rho} - \boldsymbol{b}\|_{\ell_2}^2 + \langle \boldsymbol{z}, \boldsymbol{b} - \mathcal{A}\boldsymbol{\rho}\rangle \tag{24}$$

for $\boldsymbol{\rho} \in \mathbb{C}^K$ and $\boldsymbol{z} \in \mathbb{C}^N$, and determine the solution of (7) as

$$\boldsymbol{\rho}_{\ell_1} = \arg_{\boldsymbol{\rho}} \max_{\boldsymbol{z}} \min_{\boldsymbol{\rho}} F(\boldsymbol{\rho}, \boldsymbol{z}). \tag{25}$$

Solutions of (24)-(25) and (7) agree for any value of the regularization parameter $\tau$ in (24) (see [28]). In practice, $\tau$ is used to adjust the thresholding level $\tau\Delta t$ of the semi-implicit discretization of (24)-(25)

$$\begin{aligned} \boldsymbol{\rho}_{k+1} &= \mathcal{S}_{\tau\Delta t}\left(\boldsymbol{\rho}_k + \Delta t\, \mathcal{A}^*(\boldsymbol{z}_k + \boldsymbol{b} - \mathcal{A}\,\boldsymbol{\rho}_k)\right), \\ \boldsymbol{z}_{k+1} &= \boldsymbol{z}_k + \Delta t\, \left(\boldsymbol{b} - \mathcal{A}\,\boldsymbol{\rho}_k\right). \end{aligned} \tag{26}$$

Here, $\mathcal{S}_t$ is the component-wise shrinkage-thresholding operator: for any $y = re^{i\phi} \in \mathbb{C}$ we have $\mathcal{S}_t(re^{i\phi}) = e^{i\phi}\max\{0, |r| - t\}$. GeLMA sets $\boldsymbol{\rho}_0 = 0$, $\boldsymbol{z}_0 = 0$, and $\Delta t = \min\{2/\|\mathcal{A}\|^2, \tau/\|\mathcal{A}\|\}$, and iterates (26) till convergence to (7).

*Imaging setup*

The images are obtained in a homogeneous medium with an active array of $N = 25$ transducers. We collect measurements corresponding to $S = 25$ frequencies equispaced in the bandwidth. Thus, the length of the data vector $\boldsymbol{b}$ is $NS = 625$. The ratios between the array size $a$ and the distance $L$ to IW, and between the bandwidth $2B$ and the central frequency $\omega_0$ vary in the numerical experiments, so the classical Rayleigh resolution limits change. The size of the IW is fixed. It is discretized using a uniform grid of $K = 3721$ points of size $\lambda_0/2$ in range and cross-range directions.

The images have been formed by solving the $\ell_1$-norm minimization problem (7) using the algorithm GeLMA (26).

*Results for noiseless data. Super-resolution and $\ell_1$-reconstructions*

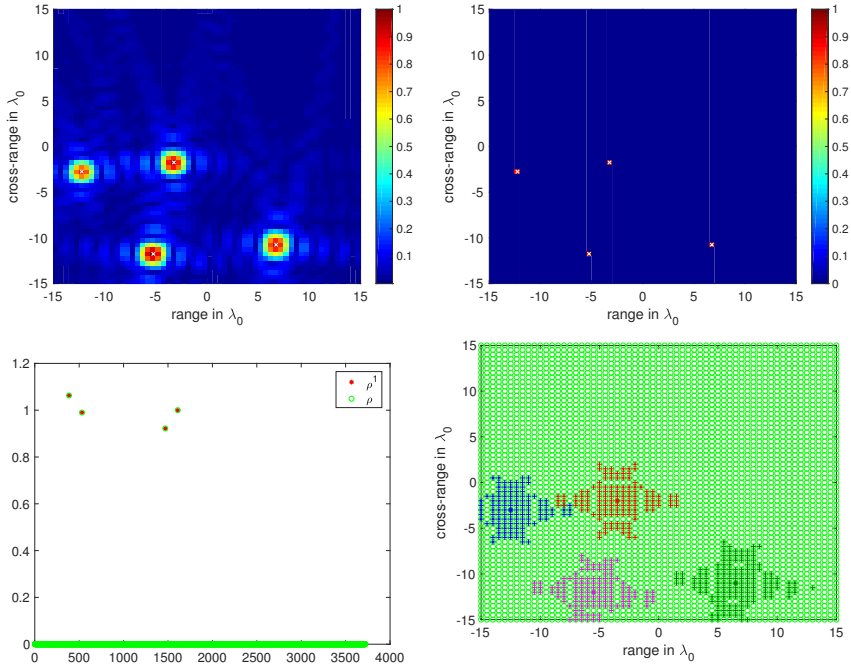


**Figure 2.** Imaging with no noise $M = 4$ sources. On the top row we show the $\ell_2$ image, $\boldsymbol{\rho}_{\ell_2}$, on the left and the $\ell_1$ image, $\boldsymbol{\rho}_{\ell_1}$, on the right. On the bottom row left image we plot the solution vector $\boldsymbol{\rho}_{\ell_1}$ with red stars and compare it with the true solution vector $\boldsymbol{\rho}$ plotted with green circles. The vicinities $S_j$, $j = 1, \ldots, M$ for the four targets are plotted with different colors in the bottom right image. In this example we use large array aperture and large bandwidth; $a/L = 1/2$ and $(2B)/\omega_0 = 1/2$.

Figures 2- 3 show the results obtained for a relatively large array and a relatively large bandwidth corresponding to ratios $a/L = 1/2$ and $(2B)/\omega_0 = 1/2$ when the data is noiseless. On the top row, from left to right we show the $\boldsymbol{\rho}_{\ell_2}$ solution (16) and the $\boldsymbol{\rho}_{\ell_1}$ solution obtained from (7). On the bottom row, the comparison between $\boldsymbol{\rho}_{\ell_1}$ (red

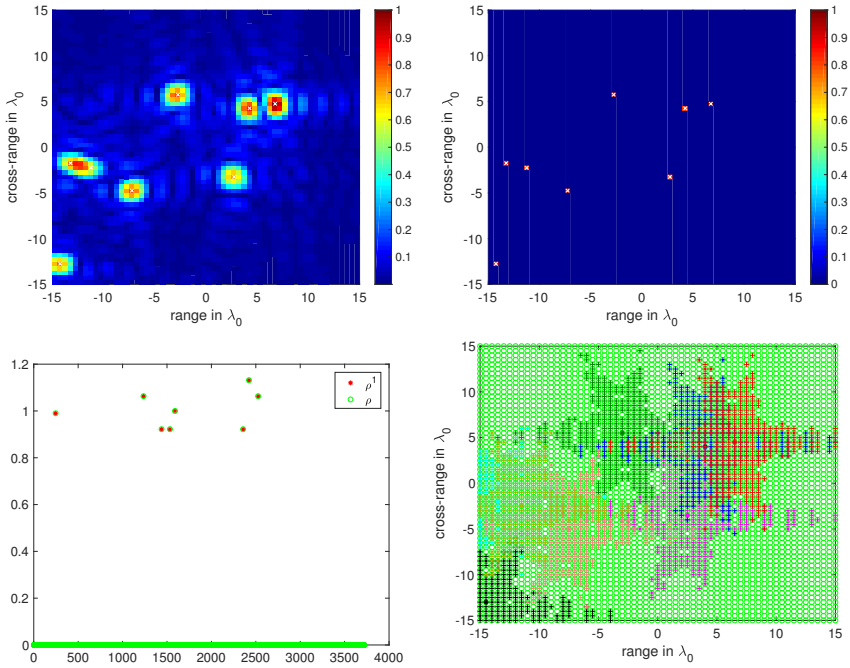


**Figure 3.** Same as Figure 2 but for $M = 8$ sources.

stars) and the true solution $\boldsymbol{\rho}$ (green circles) is shown on the left and the vicinities $S_j$ defined in (9) are plotted with different colors on the right. Figure 2 (resp. Figure 3) is for $M = 4$ (resp. $M = 8$) sources. The exact locations of the sources are indicated with white crosses in the top row images. The $M = 4$ sources in Figure 2 are very far apart: their vicinities do not overlap as it can be seen in the bottom right image. In this case, all the conditions of Proposition 1 are satisfied and we find the exact source distribution by $\ell_1$-norm minimization. The $M = 8$ sources in Figure 3 are closer, and their vicinities are larger; according to (9) the size of the vicinities increases with $M$. In fact, their vicinities overlap as it can be seen in the bottom right image. Still, the $\ell_1$-norm minimization algorithm finds the exact solution.

The classical resolution limits for this setup are $c_0/(2B) = 2\lambda_0$ in range and $\lambda_0 L/a = 2\lambda_0$ in cross-range. This means that the resolution of the $\ell_2$-norm solutions is of the order $O(2\lambda_0)$; see the top left image of Figures 2 and 3. Recall that our discretization is $\lambda_0/2$, that is four times finer than the classical resolution limit. Thus, each source roughly corresponds to a four-by-four-pixel square, which is what the $\boldsymbol{\rho}_{\ell_2}$ solutions show. Note that for $M = 8$, because two sources are quite close, the $\boldsymbol{\rho}_{\ell_2}$ solution only displays 7 sources. The ability of $\ell_1$-norm minimization to determine the location of the sources with better accuracy than the classical resolution limits is referred to as *super-resolution*.

We stress that if the IW is discretized using a very fine grid, with a grid size smaller than the classical resolution limit, then the columns of the matrix $\mathcal{A}$ are almost parallel and the decoherence condition (8) is violated. The columns that are almost parallel to those indexed by the support of the true solution are contained in the vicinities (9).

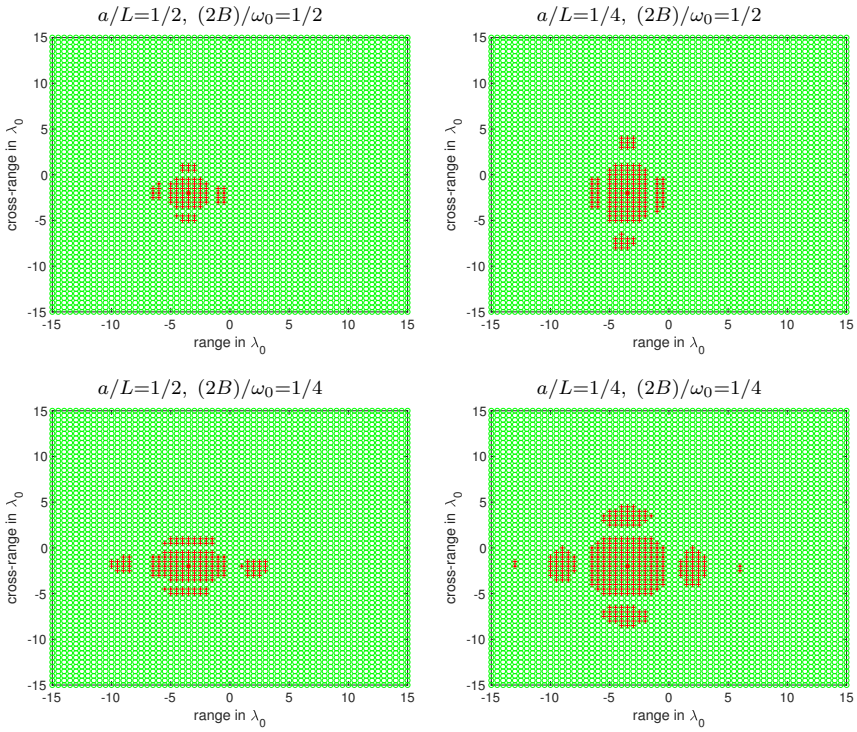


**Figure 4.** Vicinities (9) for different array and bandwidth sizes. From left to right and top to bottom the ratios $(a/L, 2B/\omega_0)$ are: $(1/2, 1/2)$, $(1/2, 1/4)$, $(1/4, 1/2)$ and $(1/4, 1/4)$.

The number of columns that belong to the vicinities depends on the imaging system. To illustrate the effect of the array and bandwidth sizes on the size of the vicinities we plot in Figure 4 the vicinity of one source for $M = 4$. From left to right and top to bottom we plot the vicinities for $[a/L, 2B/\omega_0] = [1/2, 1/2]$, $[a/L, 2B/\omega_0] = [1/2, 1/4]$, $[a/L, 2B/\omega_0] = [1/4, 1/2]$, and $[a/L, 2B/\omega_0] = [1/4, 1/4]$. As expected, the size of the vicinity is proportional to the resolution estimates $\lambda_0 L/a$ and $c_0/(2B)$ in cross-range and range, respectively.

*Results for noisy data. Stabilization of $\ell_1$-norm minimization using the noise collector matrix $\mathcal{C}$*

We add now mean zero uncorrelated noise to the data. We examine the results for different values of the signal-to-noise ratio (SNR). As we specify in the captions, our SNR is either 0dB (100% of noise) or 4dB (40% of noise). We consider first the same imaging configuration as in Figure 2 with $M = 4$ sources. The number of data is $NS = 625$ and the number of unknowns is $K = 3721$. In the top row of Figure 5 we plot the minimal $\ell_1$-norm image obtained by solving problem (7) when the SNR is 4dB. The true solution is shown with white crosses. It is apparent that, even for this

moderate level of noise, $\ell_1$-norm minimization fails to give a good image.

The problem can be alleviated using the noise collector matrix $\mathcal{C}$, as it can be seen in the results shown in the bottom row of Figure 5. To construct the noise collector matrix $\mathcal{C}$ that verifies the assumptions of Proposition 2, we take its columns $\boldsymbol{c}_j$ to be random vectors in $\mathbb{C}^{NS}$ with mean zero and variance $1/(NS)$. Their $\ell_2$-norm tends to one as $NS \to \infty$, and we check that conditions (17) and (18) are satisfied. In theory, the number of columns $\Sigma$ should be very large, of the order of $e^{NS}$, but in practice, we obtain stable results with $\Sigma$ of the order of $10^4$, which is roughly $3K$.

The solution $\boldsymbol{\rho}_{\ell_1} \in \mathbb{C}^{K+\Sigma}$ obtained with the noise collector can be decomposed into two vectors; the vector $\boldsymbol{\rho}_{iw} \in \mathbb{C}^K$ corresponding to the sought solution in the $IW$, and the vector $\boldsymbol{\rho}_{noise} \in \mathbb{C}^\Sigma$ that absorbs the noise. We display these two vectors in the bottom right plot of Figure 5. The first $K$ components correspond to $\boldsymbol{\rho}_{iw}$ and the remaining $\Sigma$ components to $\boldsymbol{\rho}_{noise}$. It is remarkable that the vector $\boldsymbol{\rho}_{iw}$ is very close to the true solution and that it contains only some small *grass*. This means that both the coherent misfit (14) and the incoherent remainder (15) are now small. This is in accordance with the theoretical error estimates (14) and (15), where $\gamma$ is now independent of the dimension of the data vector $NS$; see (19).

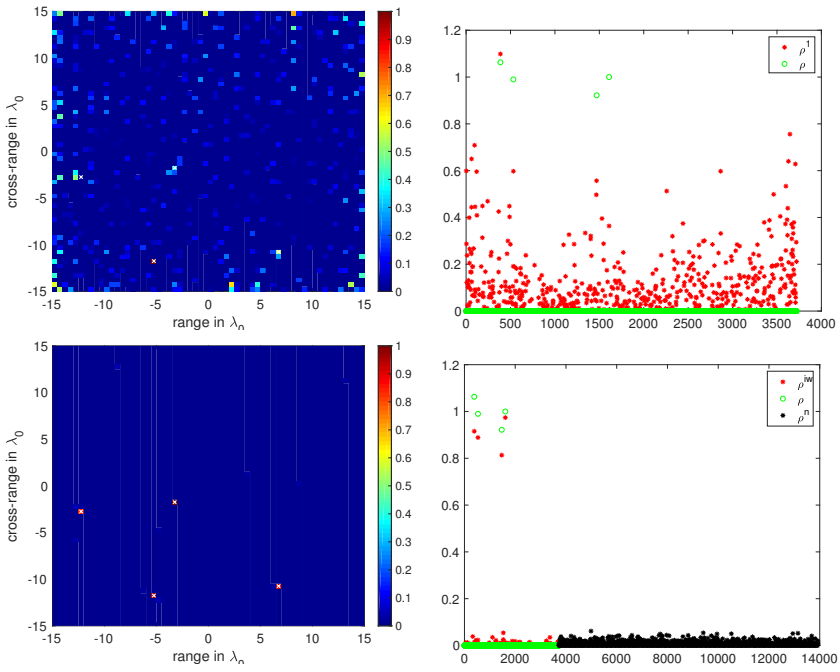


**Figure 5.** Imaging with noisy data, SNR = 4dB. The top and the bottom rows show the results without and with the noise collector, respectively. The left columns show the $\boldsymbol{\rho}_{\ell_1}$ images (the true solution is displayed with white crosses) and the right columns show the comparison (red stars) with the true solution (green circles). In the bottom right image, the first $K = 3721$ components of the solution corresponding to the IW are plotted with red stars, and the $\Sigma = 12000$ next components corresponding to the noise collector are plotted with black stars.

In the next figures, we consider an imaging setup with a large aperture $a/L = 1$ and a large bandwidth $(2B)/\omega_0 = 1$. Moreover, we increase the pixel size to $\lambda_0$ in both range and cross-range directions, so the Rayleigh resolution is of the order of a pixel. With this imaging configuration, the columns of the model matrix $\mathcal{A}$ are less coherent than in the previous numerical experiments. We plot in Figure 6 the $\ell_1$-norm image for a SNR = 4dB. With a less coherent matrix $\mathcal{A}$ the results are very good. This highlights the inherent difficulty in imaging when high resolution is required as in Figure 5 .

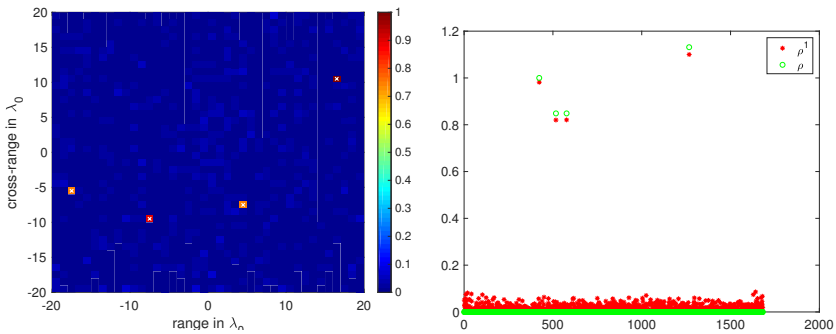


**Figure 6.** Low resolution images with a moderate level of noise in the data so SNR = 4dB. $NS = 625$ measurements. $K = 1681$ pixels in the images.

For the particular low imaging resolution configuration considered in Figure 6 we obtain good results for a large noise level corresponding to SNR = 0dB; see the top row of Figure 7 where $N\,S = 625$ as before. However, when we increase the number of measurements to $NS = 1369$, the image obtained with $\ell_1$-norm minimization turns out to be useless; see the bottom row of Figure 7. This illustrates the counter-intuitive fact that $\ell_1$-norm minimization does not always benefit from more data, at least if the data is highly contaminated with noise. This is so because the constant $\gamma$ in (12) depends on the length of the data vector $\boldsymbol{b}$ as $\sqrt{NS}$.

As before, this problem can be fixed with the noise collector as we illustrate in Figure 8. Again, the noise is effectively absorbed for both $NS = 625$ (top row) and $NS = 1369$ (bottom row) measurements using a matrix collector with a *relatively small* number of columns, many less than $e^{NS}$ as Proposition 2 suggests.

We finish with one last example that shows that the use of the noise collector makes $\ell_1$-norm minimization competitive for imaging sparse scenes because it provides stable results with super-resolution even for highly corrupted data. We consider the example with $M = 8$ sources and SNR = 0dB. The array and the bandwidth are relatively large $(a/L = 1/2,\ (2B)/\omega_0 = 1/2)$, so the classical $\ell_2$-norm resolution is of the order $O(2\lambda_0)$, as in Figure (5). In Figure 9 we show, from left to right, (i) the minimal $\ell_1$-norm solution without noise collector, which fails to give a good image, (ii) the $\ell_2$-norm solution (16), which is stable to additive noise but does not resolve nearby sources, and (iii) the minimal $\ell_1$-norm solution with the noise collector, which provides a very precise and stable image.

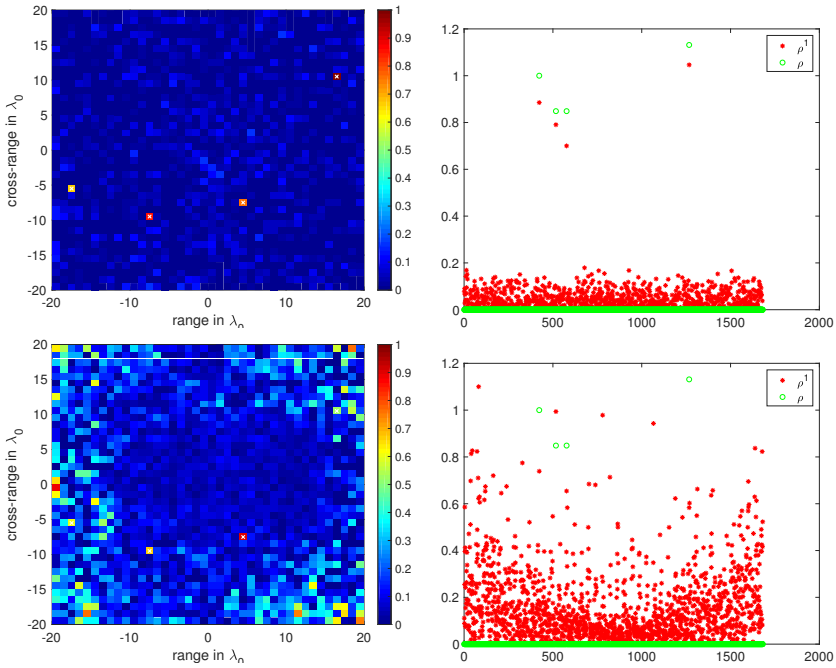


**Figure 7.** Low resolution images with a high level of noise in the data so SNR = 0dB. Top row: $NS = 625$ measurements. Bottom row: $NS = 1369$ measurements. $K = 1681$ pixels in the images.

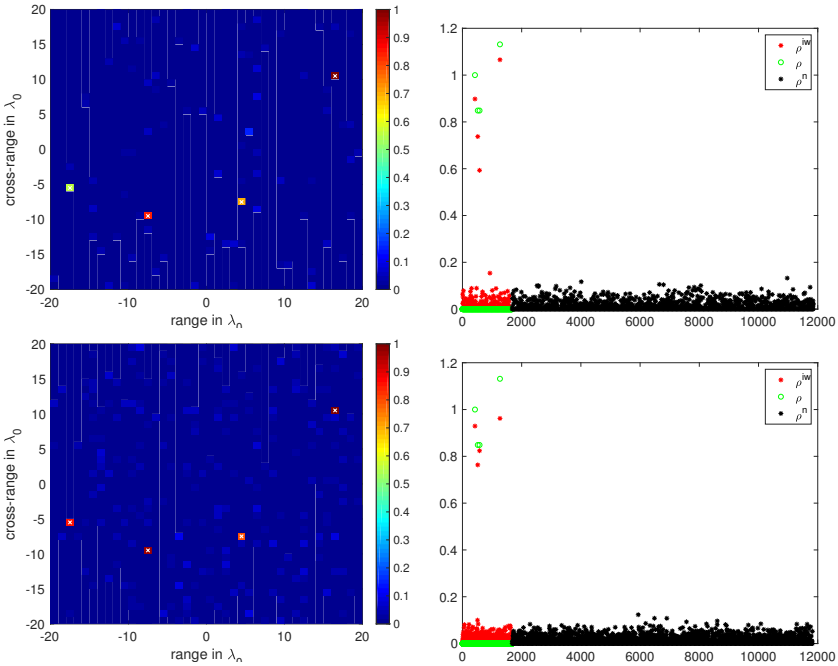


**Figure 8.** Same as Fig. 7 but with a noise collector matrix $\mathcal{C}$ with $\Sigma = 12000$ columns.

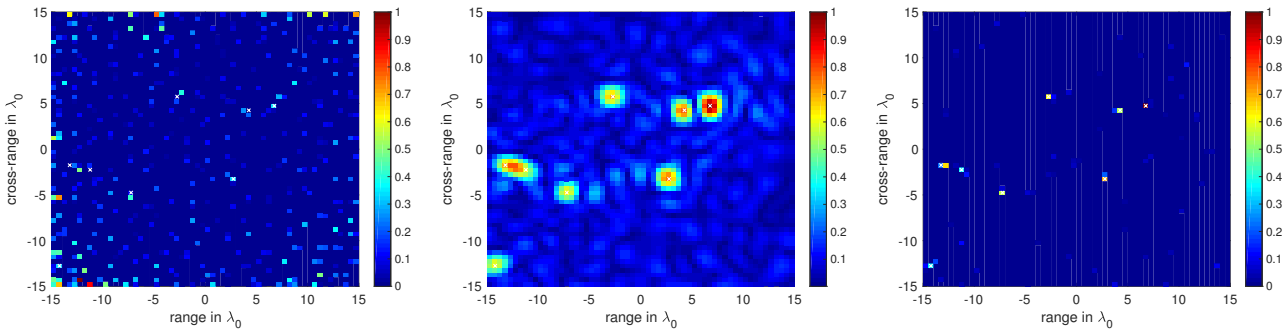


**Figure 9.** High resolution images with a high level of noise in the $NS = 625$ data, so SNR = 0dB. From left to right: plain $\boldsymbol{\rho}_{\ell_1}$ without noise collector, $\boldsymbol{\rho}_{\ell_2}$, and $\boldsymbol{\rho}_{\ell_1}$ using a noise collector. $K = 3721$ pixels in the images.

## 5. Discussion

In this paper, we consider imaging problems that can be formulated as underdetermined linear systems of the form $\mathcal{A}\,\boldsymbol{\rho}_\delta = \boldsymbol{b}_\delta$, where $\mathcal{A}$ is an $N \times K$ model matrix with $N \ll K$, and $\boldsymbol{b}_\delta$ is the $N$-dimensional data vector contaminated with noise. We assume that the solution is an M-sparse vector in $\mathbb{C}^K$, corresponding to the $K$ pixels of the IW. We consider additive noise in the data, so the data vector can be decomposed as $\boldsymbol{b}_\delta = \boldsymbol{b} + \delta\boldsymbol{b}$, where $\boldsymbol{b}$ is the data vector in the absence of noise and $\delta\boldsymbol{b}$ is the noise vector. We provide a theoretical framework that allows us to examine under what conditions the $\ell_1$-minimization problem admits a solution that is close to the exact one. We also have shown that, for our imaging problems, $\ell_1$-minimization fails when the noise level is high and the dimension $N$ of the data vector $\boldsymbol{b}_\delta$ increases. The reason is that the error is proportional to the square root of $N$.

To alleviate this problem and increase the robustness of $\ell_1$-minimization, we propose a regularization strategy. In particular, we seek the solution of $[\mathcal{A}\,|\,\mathcal{C}]\,\boldsymbol{\rho}_\delta = \boldsymbol{b}_\delta$, where the $N \times \Sigma$ matrix $\mathcal{C}$ is a noise collector. Thus, the unknown $\boldsymbol{\rho}_\delta$ is now a vector in $\mathbb{C}^{K+\Sigma}$. The first $K$ components of the unknown correspond to the distribution of sources in the IW, while the $\Sigma$ next components do not correspond to any physical quantity. They are introduced to provide a fictitious source distribution given by an appropriate linear combination of the columns of $\mathcal{C}$ that produces a good approximation to $\delta\boldsymbol{b}$. The main idea is to create a library of noises. The columns of the noise collector matrix are elements of this library and they are constructed to be incoherent with respect to the columns of $\mathcal{A}$. Theoretically, the dimension $\Sigma$ of the noise collector increases exponentially with $N$, which suggests that it may not be useful in practice. Our numerical results show, however, robustness for $\ell_1$-minimization in the presence of noise when a large enough number of columns $\Sigma \precsim 10K$ is used to build the noise collector matrix.

Our first findings on the noise collector are very encouraging. We have shown that its use improves dramatically the robustness of $\ell_1$-norm reconstructions when the

data are corrupted with additive uncorrelated noise. Many other questions ought to be addressed. Some directions of our future research concern the following aspects: what happens with other types of noise?, can we design noise collectors adaptively depending on the noise in the data?, what if the noise comes from wave propagation in a random medium?, can we design a noise collector for this case?, how much do we need to know about the noise so as to design a good noise collector?, can we retrieve this information from the data? Some of these questions will be addressed somewhere else.

## 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 MICINN grant FIS2016-77892-R. The work of A.Novikov was partially supported by NSF grants DMS-1515187, DMS-1813943. The work of C. Tsogka was partially supported by AFOSR FA9550-17-1-0238.

## References

[1] F. Alizadeh, *Interior point methods in semidefinite programming with applications to combinatorial optimization*, SIAM J. Optim. 5 (1995), pp. 13–51.

[2] Beck, Amir, and Teboulle, Marc, *A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems*, SIAM J. Img. Sci. 2 (2009), pp.183–202.

[3] G. Beylkin, *Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform*, Journal of Mathematical Physics 26 (1985), pp. 99–108.

[4] N. Bleistein, J.K. Cohen, and J.W. Stockwell Jr., *Mathematics of multidimensional seismic imaging, migration, and inversion*, Springer, New York, 2001.

[5] L. Borcea and I. Kocyigit, *Resolution analysis of imaging with $\ell_1$ optimization*, SIAM J. Imaging Sci. 8 (2015), pp. 3015–3050.

[6] L. Borcea and I. Kocyigit, *A multiple measurement vector approach to synthetic aperture radar imaging,* SIAM J. Imaging Sci. 11 (2018), pp. 770–801.

[7] L. Borcea and G. Papanicolaou and C. Tsogka, *A resolution study for imaging and time reversal in random media*, Contemporary Math. 333 (2003), pp. 63–77.

[8] Borwein, Jonathan M. and Luke, D. Russell, *Entropic regularization of the $\ell_0$ function*, Springer Optim. Appl. 49 (2011), Springer, pp. 65–92.

[9] Bruckstein, Alfred M., Donoho, David L., Elad, Michael, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM Rev. 51 (2009), pp.34–81.

[10] 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.

[11] 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.

[12] 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.

[13] E. J. Candès and C. Fernandez-Granda, *Towards a mathematical theory of super-resolution,* Communications on Pure and Applied Mathematics 67 (2014), pp. 906-956.

[14] A. Chai, M. Moscoso and G. Papanicolaou, *Robust imaging of localized scatterers using the singular value decomposition and $\ell_1$ optimization*, Inverse Problems 29 (2013), 025016.

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

[16] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, Journal of Mathematical Imaging and Vision 40 (2011), pp 120–145.

[17] D. Donoho, *Super-resolution via sparsity constraint*, SIAM Journal on Mathematical Analysis 23 (1992), pp. 1303–1331.

[18] D. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell_1$ minimization*, Proceedings of the National Academy of Sciences 100 (2003), pp. 2197–2202.

[19] 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.

[20] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, Annals of Statistics 32 (2004), pp. 407–499.

[21] A. C. Fannjiang, T. Strohmer, and P. Yan, *Compressed Remote Sensing of Sparse Objects*, SIAM J. Imaging Sci. 3 (2010), pp 595–618.

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

[23] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, *Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems,* IEEE Journal of Selected Topics in Signal Processing 1 (2007), pp.586–597.

[24] 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.

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

[26] M. A. Herman and T. Strohmer, *High-Resolution Radar via Compressed Sensing,* IEEE Transactions on Signal Processing 57 (2009), pp. 2275–2284.

[27] J. N. Laska, M. A. Davenport and R. G. Baraniuk, *Exact signal recovery from sparsely corrupted measurements through the Pursuit of Justice,* 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 2009, pp. 1556–1560.

[28] 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).

[29] A. Nedić, and D. P. Bertsekas, *Incremental subgradient methods for non-differentiable optimization*, SIAM J. Optim. 12 (2001), pp.109–138.

[30] Y. Nesterov, *A method of solving a convex programming problem with convergence rate $O(1/k^2)$*, Soviet Mathematics Doklady 27 (1983), pp. 372–376.

[31] Y. Nesterov, *Gradient methods for minimizing composite objective function*, Math. Program., Ser. B 140 (2013), pp. 125–161.

[32] M. R. Osborne, B. Presnell, and B.A. Turlach, *A new approach to variable selection in least squares problems*, IMA Journal of Numerical Analysis 20 (2000), pp. 389-403.

[33] R. Tibshirani, *Regression shrinkage and selection via the Lasso*, Journal of the Royal Statistical Society, Series B 58 (1996), pp. 267–288.

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

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

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

[37] M. H. Wright, *The interior-point revolution in optimization: history, recent developments, and lasting consequences*, Bull. Amer. Math. Soc. (N.S) 42 (2005), pp.39–56.

## Appendix A. Proof of Proposition 1

We will now prove auxiliary lemmas that we will use in the proof of Propostion 1.

**Lemma 1** *Let $\mathcal{B}$ be an $M \times M$ Hermitian matrix such that $b_{ii} = 1$, and $|b_{ij}| \leqslant c$ for all $i \neq j$. Assume $(M-1)c < 1$, then any eigenvalue $\lambda$ of $\mathcal{B}$ satisfies*

$$1 - (M-1)c \leqslant \lambda \leqslant 1 + (M-1)c. \tag{A.1}$$

**Proof:** Suppose $\mathcal{B}\boldsymbol{\rho} = \lambda\boldsymbol{\rho}$. By the triangle inequality for any row $i$ we have

$$|\rho_i| - \left|\sum_{j\neq i} b_{ij}\rho_j\right| \leqslant |\lambda\rho_i| \leqslant |\rho_i| + \left|\sum_{j\neq i} b_{ij}\rho_j\right|.$$

Since $\left|\sum_{j\neq i} b_{ij}\rho_j\right| \leqslant (M-1)c$, we obtain (A.1).

□

**Lemma 2** *Suppose $\gamma$ is defined by (12). Let $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ be minimizers of $\|\boldsymbol{\eta}\|_{\ell_1}$, subject to $\mathcal{A}\boldsymbol{\eta} = \boldsymbol{b}_1$ and $\mathcal{A}\boldsymbol{\eta} = \boldsymbol{b}_2$, respectively. Then, there exists $\boldsymbol{\xi}$ such that $\mathcal{A}\boldsymbol{\xi} = \boldsymbol{b}_1$,*

$$\|\boldsymbol{\xi}\|_{\ell_1} \leqslant \|\boldsymbol{\rho}_1\|_{\ell_1} + 2\gamma\|\boldsymbol{b}_1 - \boldsymbol{b}_2\|_{\ell_2}, \tag{A.2}$$

*and*

$$\|\boldsymbol{\xi} - \boldsymbol{\rho}_2\|_{\ell_1} \leqslant \gamma\|\boldsymbol{b}_1 - \boldsymbol{b}_2\|_{\ell_2}. \tag{A.3}$$

**Proof:** Let us first show that

$$|\|\boldsymbol{\rho}_1\|_{\ell_1} - \|\boldsymbol{\rho}_2\|_{\ell_1}| \leqslant \gamma\|\boldsymbol{b}_1 - \boldsymbol{b}_2\|_{\ell_2}. \tag{A.4}$$

Assume, for definiteness, that $\|\boldsymbol{\rho}_1\|_{\ell_1} > \|\boldsymbol{\rho}_2\|_{\ell_1}$. Then,

$$|\|\boldsymbol{\rho}_1\|_{\ell_1} - \|\boldsymbol{\rho}_2\|_{\ell_1}| = \|\boldsymbol{\rho}_1\|_{\ell_1} - \|\boldsymbol{\rho}_2\|_{\ell_1}.$$

Suppose $\boldsymbol{\rho}_3$ is a minimizer of $\|\boldsymbol{\eta}\|_{\ell_1}$, subject to $\mathcal{A}\boldsymbol{\eta} = \boldsymbol{b}_1 - \boldsymbol{b}_2$. Since $\mathcal{A}(\boldsymbol{\rho}_2 + \boldsymbol{\rho}_3) = \boldsymbol{b}_1$, and $\boldsymbol{\rho}_1$ is a minimizer of $\|\boldsymbol{\eta}\|_{\ell_1}$, subject to $\mathcal{A}\boldsymbol{\eta} = \boldsymbol{b}_1$, it follows $\|\boldsymbol{\rho}_1\|_{\ell_1} \leqslant \|\boldsymbol{\rho}_2 + \boldsymbol{\rho}_3\|_{\ell_1}$. By (12) and the triangle inequality

$$\|\boldsymbol{\rho}_2 + \boldsymbol{\rho}_3\|_{\ell_1} \leqslant \|\boldsymbol{\rho}_2\|_{\ell_1} + \gamma\|\boldsymbol{b}_1 - \boldsymbol{b}_2\|_{\ell_2}.$$

Thus, (A.4) holds.

Let $\boldsymbol{\xi} = \boldsymbol{\rho}_2 + \boldsymbol{\rho}_3$, where $\boldsymbol{\rho}_3$ is a minimizer of $\|\boldsymbol{\eta}\|_{\ell_1}$, subject to $\mathcal{A}\boldsymbol{\eta} = \boldsymbol{b}_1 - \boldsymbol{b}_2$. Then, $\mathcal{A}\boldsymbol{\xi} = \boldsymbol{b}_1$, and inequality (A.3) follows from (12). Using (A.4), (A.3), and the triangle inequality we obtain

$$\|\boldsymbol{\xi}\|_{\ell_1} \leqslant \|\boldsymbol{\rho}_2\|_{\ell_1} + \|\boldsymbol{\xi} - \boldsymbol{\rho}_2\|_{\ell_1} \leqslant \|\boldsymbol{\rho}_1\|_{\ell_1} + 2\gamma\|\boldsymbol{b}_1 - \boldsymbol{b}_2\|_{\ell_2}.$$

□

**Lemma 3** *Suppose $\mathcal{A}\boldsymbol{\rho} = \mathcal{A}\boldsymbol{\xi} = \boldsymbol{b}$, where $\boldsymbol{\rho}$ is $M$-sparse, and $\boldsymbol{\xi}$ is arbitrary. Assume vicinities (9) do not overlap. Then,*

$$\mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\xi}) \leqslant \frac{1}{2}\mathbf{In}(\boldsymbol{\rho}, \boldsymbol{\xi}). \tag{A.5}$$

*In particular,*

$$\|\boldsymbol{\rho}\|_{\ell_1} \leqslant \|\boldsymbol{\xi}\|_{\ell_1}. \tag{A.6}$$

**Proof:** For any $\mu \in \mathbb{C}^M$, we have

$$0 = \langle \mathcal{A}_T \left(\mathcal{A}_T^* \mathcal{A}_T\right)^{-1} \boldsymbol{\mu}, \mathcal{A}(\boldsymbol{\rho} - \boldsymbol{\xi})\rangle = \langle \boldsymbol{\mu}, \left(\mathcal{A}_T^* \mathcal{A}_T\right)^{-1} \mathcal{A}_T^* \mathcal{A}(\boldsymbol{\rho} - \boldsymbol{\xi})\rangle$$

since $0 = \mathcal{A}(\boldsymbol{\rho} - \boldsymbol{\xi})$. By Lemma 1, the largest eigenvalue of $\left(\mathcal{A}_T^* \mathcal{A}_T\right)^{-1}$ is smaller than $3/2$. Thus,

$$\left| \sum_{j\in T} \bar{\mu}_j \rho_j - \sum_{j\in T} \sum_{k\in S_j} \langle \boldsymbol{a}_j, \boldsymbol{a}_k\rangle \bar{\mu}_j \xi_k \right| \leqslant \frac{3}{2} \sum_{j\in T} \sum_{k\notin \Upsilon} \left| \langle \boldsymbol{a}_j, \boldsymbol{a}_k\rangle \bar{\mu}_j \xi_k \right|, \ \Upsilon = \cup_{j\in T} S_j.$$

Choose $\mu_j$, so that $|\mu_j| = 1$ and

$$\left| \sum_{j\in T} \bar{\mu}_j \rho_j - \sum_{j\in T} \sum_{k\in S_j} \langle \boldsymbol{a}_j, \boldsymbol{a}_k\rangle \bar{\mu}_j \xi_k \right| = \mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\xi}).$$

We can estimate

$$\mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\xi}) \leqslant \frac{3}{2}\frac{1}{3M} \sum_{j\in T} \sum_{k\notin \Upsilon} |\xi_k| \leqslant \frac{1}{2} \sum_{k\notin \Upsilon} |\xi_k|.$$

which is equivalent to (A.5). Observe that (see (10))

$$\|\boldsymbol{\rho}\|_{\ell_1} - \sum_{k\in \Upsilon} |\xi_k| \leqslant \mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\xi}).$$

□

**Proof:** [Proof of Proposition 1] If $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_\delta$ are minimizers of $\|\boldsymbol{\eta}\|_{\ell_1}$, subject to $\mathcal{A}\boldsymbol{\eta} = \boldsymbol{b}$ and $\mathcal{A}\boldsymbol{\eta} = \boldsymbol{b}_\delta$, respectively, we can apply Lemma 2 and conclude there exists $\xi$ such that $\mathcal{A}\boldsymbol{\xi} = \boldsymbol{b}$,

$$\|\boldsymbol{\xi}\|_{\ell_1} \leqslant \|\boldsymbol{\rho}\|_{\ell_1} + 2\gamma\delta, \tag{A.7}$$

and

$$\|\boldsymbol{\xi} - \boldsymbol{\rho}_\delta\|_{\ell_1} \leqslant \gamma\delta. \tag{A.8}$$

Since

$$\|\boldsymbol{\rho}\|_{\ell_1} \leqslant \mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\xi}) + \sum_{j\in T} |\xi_j|,$$

by Lemma 3 we have

$$\|\boldsymbol{\rho}\|_{\ell_1} \leqslant \frac{1}{2} \sum_{j\notin T} |\xi_j| + \sum_{j\in T} |\xi_j| = \|\boldsymbol{\xi}\|_{\ell_1} - \frac{1}{2}\mathbf{In}(\boldsymbol{\rho}, \boldsymbol{\xi}). \tag{A.9}$$

Comparing (A.9) and (A.7) we conclude

$$\mathbf{In}(\boldsymbol{\rho}, \boldsymbol{\xi}) \leqslant 4\gamma\delta. \tag{A.10}$$

By the triangle inequality and (A.8), we have

$$\mathbf{In}(\boldsymbol{\rho}, \boldsymbol{\rho}_\delta) \leqslant \mathbf{In}(\boldsymbol{\rho}, \boldsymbol{\xi}) + \|\boldsymbol{\xi} - \boldsymbol{\rho}_\delta\|_{\ell_1} \leqslant 5\gamma\delta.$$

Hence, we have obtained (15). From (A.5) and (A.10), we obtain

$$\mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\xi}) \leqslant 2\gamma\delta.$$

By the triangle inequality and (A.8), we have

$$\mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\rho}_\delta) \leqslant \mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\xi}) + \|\boldsymbol{\xi} - \boldsymbol{\rho}_\delta\|_{\ell_1} \leqslant 3\gamma\delta.$$

If the noise level $\delta = 0$, then $\mathbf{Co}(\boldsymbol{\rho}, \boldsymbol{\xi}) = \mathbf{In}(\boldsymbol{\rho}, \boldsymbol{\xi}) = 0$. It means $\mathrm{supp}(\boldsymbol{\rho}_\delta) \subset \Upsilon$. Since $\mathcal{A}\boldsymbol{\rho}_\delta = \mathcal{A}\boldsymbol{\rho}$, we can use (A.6). Note that the inequality (A.6) becomes strict if $\Upsilon$ does not contain collinear vectors. Thus, we conclude $\boldsymbol{\rho}_\delta = \boldsymbol{\rho}$. □