The Noise Collector for sparse recovery in high dimensions

Miguel Moscoso[a,2], **Alexei Novikov**[b], **George Papanicolaou**[c,2], **and Chrysoula Tsogka**[d]

[a]Department of Mathematics, Universidad Carlos III de Madrid, Leganes, Madrid 28911, Spain; [b]Department of Mathematics, Pennsylvania State University, University Park, PA 16802; [c]Department of Mathematics, Stanford University, Stanford, CA 94305; [d]Department of Applied Mathematics, University of California, Merced, CA 95343

This manuscript was compiled on June 28, 2020

The ability to detect sparse signals from noisy high-dimensional data is a top priority in modern science and engineering. It is well known, that a sparse solution of the linear system $\mathcal{A}\rho = b_0$ can be found efficiently with an ℓ_1-norm minimization approach if the data is noiseless. However, detection of the signal from data corrupted by noise is still a challenging problem as the solution depends, in general, on a regularization parameter whose optimal value is not easy to choose. We propose a new efficient approach that does not require any parameter estimation. We introduce a *no-phantom* weight τ and the Noise Collector matrix $\mathcal{C}$, and solve an augmented system $\mathcal{A}\rho + \mathcal{C}\eta = b_0 + e$, where e is the noise. We show that the ℓ_1-norm minimal solution of this system has zero false discovery rate for any level of noise, with probability that tends to one as the dimension of b_0 increases to infinity. We obtain exact support recovery if the noise is not too large, and develop a Fast Noise Collector Algorithm which makes the computational cost of solving the augmented system comparable to that of the original one. We demonstrate the effectiveness of the method in applications to passive array imaging.

high dimensional probability | convex geometry | sparsity promoting algorithms | noisy data |

We want to find sparse solutions $\boldsymbol{\rho} \in \mathbb{R}^K$ for

$$\mathcal{A}\,\boldsymbol{\rho} = \boldsymbol{b}, \qquad [1]$$

from highly incomplete measurement data $\boldsymbol{b} = \boldsymbol{b}_0 + \boldsymbol{e} \in \mathbb{R}^N$, corrupted by noise $\boldsymbol{e}$, where $1 \ll N < K$. In the noiseless case, $\boldsymbol{\rho}$ can be found exactly by solving the optimization problem (1)

$$\boldsymbol{\rho}_* = \arg\min_{\boldsymbol{\rho}} \|\boldsymbol{\rho}\|_{\ell_1}, \text{ subject to } \mathcal{A}\,\boldsymbol{\rho} = \boldsymbol{b}, \qquad [2]$$

provided the measurement matrix $\mathcal{A} \in \mathbb{R}^{N\times K}$ satisfies additional conditions, e.g., decoherence or restricted isometry properties (2, 3), and the solution vector $\boldsymbol{\rho}$ has a small number M of nonzero components or degrees of freedom. When measurements are noisy, exact recovery is no longer possible. However, the exact support of $\boldsymbol{\rho}$ can still be determined if the noise is not too strong. The most commonly used approach is to solve the ℓ_2-relaxed form of Eq. **2**

$$\boldsymbol{\rho}_\lambda = \arg\min_{\boldsymbol{\rho}} \left(\lambda\|\boldsymbol{\rho}\|_{\ell_1} + \|\boldsymbol{\mathcal{A}\rho} - \boldsymbol{b}\|_{\ell_2}^2\right), \qquad [3]$$

known as Lasso in the statistics literature (4). There are sufficient conditions for the support of $\boldsymbol{\rho}_\lambda$ to be contained within the true support, see e.g. Fuchs (5), Tropp (6), Wainwright (7), and Maleki *et al* (8) . These conditions depend on the signal-to-noise ratio (SNR), which is not known and must be estimated, and on the regularization parameter λ, which must be carefully chosen and/or adaptively changed (9). Although such an adaptive procedure improves the outcome, the resulting solutions tend to include a large number of "false positives" in practice (10). Belloni *et al* (11) proposed to solve the square-root Lasso minimization problem instead of Eq. **3**, which makes the regularization parameter λ independent of the SNR. Our contribution is a new computationally efficient method for exact support recovery, with no false positives, in noisy settings. It also does not require an estimate on SNR.

Main Results. Suppose $\boldsymbol{\rho}$ is an M-sparse solution of system **1** with no noise, where the columns of $\mathcal{A}$ have unit length. Our main result ensures that we can still recover the support of $\boldsymbol{\rho}$ when the data is noisy by looking at the support of $\boldsymbol{\rho}_\tau$ found as

$$(\boldsymbol{\rho}_\tau, \boldsymbol{\eta}_\tau) = \arg\min_{\boldsymbol{\rho},\boldsymbol{\eta}} \left(\tau\|\boldsymbol{\rho}\|_{\ell_1} + \|\boldsymbol{\eta}\|_{\ell_1}\right), \qquad [4]$$
$$\text{subject to } \mathcal{A}\boldsymbol{\rho} + \mathcal{C}\boldsymbol{\eta} = \boldsymbol{b}_0 + \boldsymbol{e},$$

with an $O(1)$ weight τ, and an appropriately chosen *Noise Collector* matrix $\mathcal{C} \in \mathbb{R}^{N\times\Sigma}$, $\Sigma \gg K$. The minimization problem **4** can be understood as a relaxation of **2**, as it works by absorbing *all* the noise, and possibly some signal, in $\mathcal{C}\boldsymbol{\eta}_\tau$.

The following theorem shows that if the signal is pure noise, and the columns of $\mathcal{C}$ are chosen independently and at random on the unit sphere $\mathbb{S}^{N-1} = \left\{x \in \mathbb{R}^N, \|x\|_{\ell_2} = 1\right\}$, then $\mathcal{C}\boldsymbol{\eta}_\tau = \boldsymbol{e}$ for any level of noise, with large probability.

Theorem 1 (No phantom signal): Suppose $\boldsymbol{b}_0 = 0$ and $\boldsymbol{e}/\|\boldsymbol{e}\|_{\ell_2}$ is uniformly distributed on $\mathbb{S}^{N-1}$. Fix $\beta > 1$, and draw $\Sigma = N^\beta$ columns for $\mathcal{C}$, independently, from the uniform distribution on $\mathbb{S}^{N-1}$. For any $\kappa > 0$ there are constants $\tau = \tau(\kappa,\beta)$ and $N_0 = N_0(\kappa,\beta)$ such that, for all $N > N_0$, $\boldsymbol{\rho}_\tau$, the solution of Eq. **4**, is zero with probability $1 - 1/N^\kappa$.

This Theorem guarantees, with large probability, a zero false discovery rate in the absence of signals with meaningful information. The key to a zero false discovery rate is the choice of a *no-phantom* weight τ. Next, we generalize this

Significance Statement

The ability to detect sparse signals from noisy, high-dimensional data is a top priority in modern science and engineering. For optimal results, current approaches need to tune parameters that depend on the level of noise, which is often difficult to estimate. We develop a new parameter-free, computationally efficient, ℓ_1-norm minimization approach that has a zero false discovery rate (no false positives) with high probability for any level of noise while it detects the exact location of sparse signals when the noise is not too large.

Author contributions: M.M., A.N., G.P., and C.T. performed research and wrote the paper.

The authors declare no conflict of interest.

[2]To whom correspondence should be addressed: papanicolaou@stanford.edu, moscoso@math.uc3m.es

result for the case in which the recorded signals carry useful information.

Theorem 2 (Zero false discoveries): Let $\boldsymbol{\rho}$ be an M-sparse solution of the noiseless system $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}_0$. Assume κ, β, the Noise Collector, and the noise are the same as in Theorem 1. In addition, assume that the columns of $\mathcal{A}$ are incoherent, in the sense that $|\langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle| \leq \frac{1}{3M}$. Then, there are constants $\tau = \tau(\kappa,\beta)$ and $N_0 = N_0(\kappa,\beta)$ such that $\mathrm{supp}(\boldsymbol{\rho}_\tau) \subseteq \mathrm{supp}(\boldsymbol{\rho})$ for all $N > N_0$ with probability $1 - 1/N^\kappa$.

This Theorem holds for any level of noise and the same value of τ as in Theorem 1. The incoherence conditions in Theorem 2 are needed to guarantee that the true signal does not create false positives elsewhere. Theorem 2 guarantees that the support of $\boldsymbol{\rho}_\tau$ is inside the support of $\boldsymbol{\rho}$. The next Theorem shows that if the noise is not too large, then $\boldsymbol{\rho}_\tau$ and $\boldsymbol{\rho}$ have exactly the same support.

Theorem 3 (Exact support recovery): Keep the same assumptions as in Theorem 2. Let $\gamma = \min_{i\in\mathrm{supp}(\boldsymbol{\rho})} |\rho_i|/\|\rho\|_{\ell_\infty}$. There are constants $\tau = \tau(\kappa,\beta)$, $c_1 = c_1(\kappa,\beta,\gamma)$, and $N_0 = N_0(\kappa,\beta)$ such that, if the noise level satisfies $\|\boldsymbol{e}\|_{\ell_2} \leq c_1 \|\boldsymbol{b}_0\|_{\ell_2}^2 \|\boldsymbol{\rho}\|_{\ell_1}^{-1}\sqrt{N}/\sqrt{\ln N}$ then, for all $N > N_0$, $\mathrm{supp}(\boldsymbol{\rho}_\tau) = \mathrm{supp}(\boldsymbol{\rho})$ with probability $1 - 1/N^\kappa$.

To elucidate an interpretation of the last Theorem consider a model case where $\mathcal{A}$ is the identity matrix and all coefficients of $\boldsymbol{b}_0 = \boldsymbol{\rho}$ are either 1 or 0. Then $\|\boldsymbol{b}_0\|_{\ell_2}^2 = \|\boldsymbol{\rho}\|_{\ell_1} = M$. In this case, an acceptable relative level of noise is

$$\|\boldsymbol{e}\|_{\ell_2}/\|\boldsymbol{b}_0\|_{\ell_2} \lesssim \sqrt{N}/\sqrt{M\ln N}. \qquad [5]$$

This means that $\|\boldsymbol{e}\|_{\ell_2} \lesssim \sqrt{N}/\sqrt{\ln N}$, and it implies that each coefficient of $\boldsymbol{b}_0$ may be corrupted by $O(1/\sqrt{\ln N})$ on average, and some coefficients of $\boldsymbol{b}_0$ may be corrupted by $O(1)$.

Motivation. We are interested in imaging sparse scenes, accurately, using limited and noisy data. Such imaging problems arise in many areas such as medical imaging (12), structural biology (13), radar (14), and geophysics (15). In imaging, the ℓ_1-norm minimization method in Eq. **2** is often used, see e.g. (16–21), as it has the desirable property of super-resolution, that is, the enhancement of the fine scale details of the images. This has been analyzed in different settings by Donoho and Elad (22), Candès and Fernandez-Granda (23), Fannjiang and Liao (24), and Borcea and Kocyigit (25), among others. We want to retain this property in our method when the data is corrupted by additive noise.

However, noise fundamentally limits the quality of the images formed with almost all computational imaging techniques. Specifically, ℓ_1-norm minimization produces images that are unstable for low SNR due to the ill-conditioning of super-resolution reconstruction schemes. The instability emerges as clutter noise in the images, or *grass*, that degrades the resolution. Our initial motivation to introduce the Noise Collector matrix $\mathcal{C}$ was to regularize the matrix $\mathcal{A}$ and, thus, to suppress the clutter in the images. We proposed in (26) to seek the minimal ℓ_1-norm solution of the augmented linear system $\mathcal{A}\boldsymbol{\rho} + \mathcal{C}\boldsymbol{\eta} = \boldsymbol{b}$. The idea was to choose the columns of $\mathcal{C}$ almost orthogonal to those of $\mathcal{A}$. Indeed, the condition number of $[\mathcal{A}\,|\,\mathcal{C}]$ becomes $O(1)$ when $O(N)$ columns of $\mathcal{C}$ are taken at random. This essentially follows from the bounds on the largest and the smallest nonzero singular values of random matrices, see e.g. Theorem 4.6.1 in (27).

The idea to create a dictionary for noise is not new. For example, the work by Laska *et al.* (28) considers a specific version of the measurement noise model so $\boldsymbol{b} = \mathcal{A}\boldsymbol{\rho} + \mathcal{C}\boldsymbol{e}$, where $\mathcal{C}$ is a matrix with fewer (orthonormal) columns than rows, and the noise vector $\boldsymbol{e}$ is sparse. $\mathcal{C}$ represents the basis in which the noise is sparse and it is assumed to be known. Then, they show that it is possible to recover sparse signals and sparse noise exactly. We stress that we do not assume here that the noise is sparse. In our work, the noise is large (SNR can be small) and is evenly distributed across the data, so it cannot be sparsely accommodated.

To suppress the clutter, our theory in (26) required exponentially many columns, so $\Sigma \lesssim e^N$. This seemed to make the Noise Collector impractical, but the numerical experiments suggested that $O(N)$ columns were enough to obtain excellent results. We address this issue here and explain why the Noise Collector matrix $\mathcal{C}$ only needs algebraically many columns. Moreover, to absorb the noise completely, and thus improve the algorithm in (26), we introduce now the no-phantom weight τ in Eq. **4**. Indeed, by weighting the columns of the Noise Collector matrix $\mathcal{C}$ with respect to those in the model matrix $\mathcal{A}$, the algorithm now produces images with no clutter at all, no matter how much noise is added to the data.

Finally, we want the Noise Collector to be efficient, with almost no extra computational cost with respect to the Lasso problem in Eq. **3**. To this end, the Noise Collector is constructed using circulant matrices that allows for efficient matrix vector multiplications using FFTs.

We now explain how the Noise Collector works and reduce our Theorems to basic estimates in high-dimensional probability.

The Noise Collector

The method has two main ingredients: the Noise Collector matrix $\mathcal{C}$ and the no-phantom weight τ. The construction of the Noise Collector matrix $\mathcal{C}$ starts with the following three key properties. Firstly, its columns should be sufficiently orthogonal to the columns of $\mathcal{A}$, so it does not absorb signals with "meaningful" information. Secondly, the columns of $\mathcal{C}$ should be uniformly distributed on the unit sphere $\mathbb{S}^{N-1}$ so that we could approximate well a typical noise vector. Thirdly, the number of columns in $\mathcal{C}$ should grow slower than exponential with N, otherwise the method is impractical.

One way to guarantee all three properties is to impose

$$|\langle \boldsymbol{a}_i, \boldsymbol{c}_j\rangle| < \frac{\alpha}{\sqrt{N}}\ \forall i,j\,,\ \text{and}\ |\langle \boldsymbol{c}_i, \boldsymbol{c}_j\rangle| < \frac{\alpha}{\sqrt{N}}\ \forall i \neq j, \qquad [6]$$

with $\alpha > 1$, and fill out $\mathcal{C}$ drawing $\boldsymbol{c}_i$ at random with rejections until the rejection rate becomes too high. Then, by construction, the columns of $\mathcal{C}$ are almost orthogonal to the columns of $\mathcal{A}$, and when the rejection rate becomes too high this implies that we can not pack more N-dimensional unit vectors into $\mathcal{C}$ and, thus, we can approximate well a typical noise vector. Finally, the Kabatjanskii-Levenstein inequality (see discussion in (29)) implies that the number Σ of columns in $\mathcal{C}$ grows at most polynomially: $\Sigma \leq N^{\alpha^2}$. The first estimate in Eq. **6** implies that any solution $\mathcal{C}\boldsymbol{\eta} = \boldsymbol{a}_i$ satisfies, for any $i \leq N$, $\|\boldsymbol{\eta}\|_{\ell_1} \gtrsim \sqrt{N}$. This estimate measures how expensive it is to approximate columns of $\mathcal{A}$, i.e. the meaningful signal, with the Noise Collector. In turn, the no-phantom weight τ should be chosen so that it is expensive to approximate noise using columns of $\mathcal{A}$. It cannot be taken too large, though, because we may lose the signal. In fact, one can prove that if $\tau \geq \sqrt{N}/\alpha$, then $\boldsymbol{\rho}_\tau \equiv 0$ for any $\boldsymbol{\rho}$ and any level of noise.

Intuitively, τ characterizes the rate at which the signal is lost as the noise increases. The most important property of the no-phantom weight τ is that it does not depend on the level of noise, so it is chosen before we start using the Noise Collector.

It is, however, more convenient for the proofs to use a probabilistic version of Eq. 6. Suppose that the columns of $\mathcal{C}$ are drawn, independently, at random. Then, the dot product of any two random unit vectors is still typically of order $1/\sqrt{N}$, see e.g. (27). If the number of columns grows polynomially, we only have to sacrifice an asymptotically negligible event where our Noise Collector does not satisfy the three key properties, and the decoherence constraints in Eq. 6 are weakened by a logarithmic factor only. This follows from basic estimates in high-dimensional probability. We will state them in the next Lemma after we interpret problem 4 geometrically.

Consider the convex hulls

$$H_1 = \left\{ x \in \mathbb{R}^N \,\middle|\, x = \sum_{i=1}^{\Sigma} \xi_i \boldsymbol{c}_i,\ \sum_{i=1}^{\Sigma} |\xi_i| \le 1 \right\}, \qquad [7]$$

$$H_2 = \left\{ x \in \mathbb{R}^N \,\middle|\, x = \sum_{i=1}^{K} \xi_i \boldsymbol{a}_i,\ \sum_{i=1}^{\Sigma} |\xi_i| \le 1 \right\}, \qquad [8]$$

and $H(\tau) = \{\xi h_1/\tau + (1-\xi)h_2, 0 \le \xi \le 1, h_i \in H_i\}$. Theorem 1 states that for a typical noise vector $\boldsymbol{e}$ we can find $\lambda_0 > 0$ such that $\boldsymbol{e} \in \lambda_0 \partial H_1$ and $\boldsymbol{e} \notin \lambda \partial H(\tau)$ for any $\lambda < \lambda_0$.

Lemma 1 (Typical width of convex hulls H_i): Suppose $\Sigma = N^\beta$, $\beta > 1$, vectors $\boldsymbol{c}_i \in \mathbb{S}^{N-1}$, $i = 1, 2, \dots, \Sigma$, are drawn at random and independently, and $\boldsymbol{e} \in \mathbb{S}^{N-1}$. Then, for any $\kappa > 0$ there are constants $c_0 = c_0(\kappa, \beta)$, $\alpha = \sqrt{(\beta-1)/2}$ and $N_0 = N_0(\kappa, \beta)$, such that for all $N \ge N_0$

$$\max(\max_{i \le K}(|\langle \boldsymbol{a}_i, \boldsymbol{e}\rangle|), \max_{i \le \Sigma}(|\langle \boldsymbol{c}_i, \boldsymbol{e}\rangle|)) < c_0\sqrt{\ln N}/\sqrt{N}, \qquad [9]$$

and

$$\alpha\sqrt{\ln N}\boldsymbol{e}/\sqrt{N} \in H_1, \qquad [10]$$

with probability $1 - 1/N^\kappa$.

We sketch the proof of estimates 9 and 10 in Section Proofs. Estimate 9 can also be derived from the Milman's version of Dvoretzky's theorem (30). Informally, inequality 9 states that H_1 and H_2 are contained in the ℓ_2-ball of radius $c_0\sqrt{\ln N}/\sqrt{N}$ except for a few spikes in statistically insignificant directions. See Figure 1-left. Inequality 10 states that H_1 contains an ℓ_2-ball of radius $\alpha\sqrt{\ln N}/\sqrt{N}$ except for a few statistically insignificant directions.

These inequalities immediately imply Theorem 1. We just need to explain how to choose the no-phantom weight τ. There will be no phantoms if H_2/τ is strictly inside the ℓ_2-ball of radius $\alpha\sqrt{\ln N}/\sqrt{N}$. This could be done if $\tau > c_0/\alpha$.

If columns of $\mathcal{A}$ are orthogonal to each other, then Theorem 2 follows from Theorem 1. We just need to project the linear system in Eq. 4 on the span of $\boldsymbol{a}_i$, $i \notin \text{supp}(\boldsymbol{\rho})$, and apply Theorem 1 to the projections. If, in addition, we assume $\boldsymbol{b}_0 = \boldsymbol{a}_1\rho_1$, then the proof of Theorem 3 is illustrated on Figure 1-right. In detail, a typical intersection of $V = \text{span}(\boldsymbol{a}_1, \boldsymbol{e})$ and $H(\tau)$ is a rounded rhombus because it is the convex hull of $\boldsymbol{a}_1/\tau$ and the ℓ_2-ball of radius $c_0\sqrt{\ln N}/\sqrt{N}$. If $\boldsymbol{a}_1\rho_1 + \boldsymbol{e} \in \lambda_0 \partial H(\tau)$, then there are two options: 1) $\boldsymbol{a}_1\rho_1 + \boldsymbol{e}$ lies on the curved boundary of the rounded rhombus, and then $\text{supp}(\boldsymbol{\rho}_\tau) = \emptyset$; 2) $\boldsymbol{a}_1\rho_1 + \boldsymbol{e}$ lies on the flat boundary of the rounded rhombus, and then $\text{supp}(\boldsymbol{\rho}_\tau) = \text{supp}(\boldsymbol{\rho})$. The second option happens if the vector $\boldsymbol{a}_1\rho_1 + \boldsymbol{e}$ intersects the flat boundary of $\partial H(\tau)$. This gives the support recovery estimate in Theorem 3.

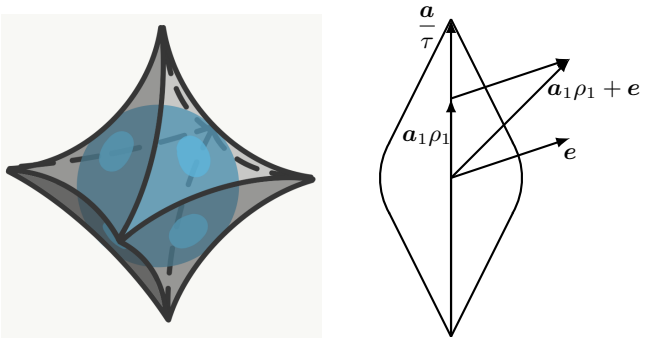

Fig. 1. Left: A convex hull H_1 is an ℓ_2 ball of radius $O(\sqrt{\ln N}/\sqrt{N})$ with few spikes. Right: An intersection of $H(\tau)$ with the span$(\boldsymbol{a}_1, \boldsymbol{e})$ is a rounded rhombus.

In the general case the columns of the combined matrix $[\mathcal{A}\,|\,\mathcal{C}]$ are incoherent. This property allows us to prove Theorems 2 and 3 in Section Proofs using known techniques, see e.g. (26). In particular, we automatically have exact recovery using (2) applied to $[\mathcal{A}\,|\,\mathcal{C}]$ if the data is noiseless.

Lemma 2 (Exact Recovery): Suppose $\boldsymbol{\rho}$ is an M-sparse solution of $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$, and there is no noise so $\boldsymbol{e} = 0$. In addition, assume that the columns of $\mathcal{A}$ are incoherent: $|\langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle| \le \frac{1}{3M}$. Then, the solution to Eq. 4 satisfies $\boldsymbol{\rho}_\tau = \boldsymbol{\rho}$ for all

$$M < \frac{2\sqrt{N}}{3c_0\tau\sqrt{\ln N}} \text{ with probability } 1 - \frac{1}{N^\kappa}. \qquad [11]$$

Fast Noise Collector Algorithm

To find the minimizer in Eq. 4, we consider a variational approach. We define the function

$$F(\boldsymbol{\rho}, \boldsymbol{\eta}, \boldsymbol{z}) = \lambda\,(\tau\|\boldsymbol{\rho}\|_{\ell_1} + \|\boldsymbol{\eta}\|_{\ell_1}) \qquad [12]$$
$$+\frac{1}{2}\|\mathcal{A}\boldsymbol{\rho} + \mathcal{C}\boldsymbol{\eta} - \boldsymbol{b}\|_{\ell_2}^2 + \langle \boldsymbol{z}, \boldsymbol{b} - \mathcal{A}\boldsymbol{\rho} - \mathcal{C}\boldsymbol{\eta}\rangle$$

for a no-phantom weight τ, and determine the solution as

$$\max_{\boldsymbol{z}} \min_{\boldsymbol{\rho}, \boldsymbol{\eta}} F(\boldsymbol{\rho}, \boldsymbol{\eta}, \boldsymbol{z}). \qquad [13]$$

The key observation is that this variational principle finds the minimum in Eq. 4 exactly for all values of the regularization parameter λ. Hence, the method has no tuning parameters. To determine the exact extremum in Eq. 13, we use the iterative soft thresholding algorithm GeLMA (31) that works as follows.

First pick a value for β and τ. For optimal results, one can calibrate τ to be the smallest constant such that Theorem 1 holds, that is, we see no phantom signals when the algorithm is fed with pure noise. In our numerical experiments we use $\beta = 1.5$ and $\tau = 2$.

Next, pick a value for the regularization parameter λ, e.g. $\lambda = 1$. Choose step sizes $\Delta t_1 < 2/\|[\mathcal{A}\,|\,\mathcal{C}]\|^2$ and $\Delta t_2 < \lambda/\|\mathcal{A}\|$.* Set $\boldsymbol{\rho}_0 = \boldsymbol{0}$, $\boldsymbol{\eta}_0 = \boldsymbol{0}$, $\boldsymbol{z}_0 = \boldsymbol{0}$, and iterate for $k \ge 0$:

$$\begin{aligned} \boldsymbol{r} &= \boldsymbol{b} - \mathcal{A}\,\boldsymbol{\rho}_k - \mathcal{C}\,\boldsymbol{\eta}_k\,, \\ \boldsymbol{\rho}_{k+1} &= \mathcal{S}_{\tau\,\lambda\Delta t_1}\,(\boldsymbol{\rho}_k + \Delta t_1\,\mathcal{A}^*(\boldsymbol{z}_k + \boldsymbol{r}))\,, \\ \boldsymbol{\eta}_{k+1} &= \mathcal{S}_{\lambda\Delta t_1}\,(\boldsymbol{\eta}_k + \Delta t_1\,\mathcal{C}^*(\boldsymbol{z}_k + \boldsymbol{r}))\,, \\ \boldsymbol{z}_{k+1} &= \boldsymbol{z}_k + \Delta t_2\,\boldsymbol{r}\,, \end{aligned} \qquad [14]$$

*Choosing two step sizes instead of the smaller one Δt_1 improves the convergence speed.

where $\mathcal{S}_r(y_i) = \text{sign}(y_i)\max\{0, |y_i| - r\}$.

The Noise Collector matrix $\mathcal{C}$ is computed by drawing $N^{\beta-1}$ normally distributed N-dimensional vectors, normalized to unit length. These are the generating vectors of the Noise Collector . From each of them, a circulant $N \times N$ matrix $\mathcal{C}_i$, $i = 1, \ldots, N^{\beta-1}$, is constructed. The Noise Collector matrix is obtained by concatenation, so $\mathcal{C} = [\mathcal{C}_1\,|\mathcal{C}_2\,|\ldots|\mathcal{C}_{N^{\beta-1}}\,]$. Exploiting the circulant structure of the matrices $\mathcal{C}_i$, we perform the matrix vector multiplications $\mathcal{C}\boldsymbol{\eta}_k$ and $\mathcal{C}^*(\boldsymbol{z}_k + \boldsymbol{r})$ in Eq. 14 using the FFT (32). This makes the complexity associated to the Noise Collector $O(N^\beta \log(N))$. Note that only the $N^{\beta-1}$ generating vectors are stored, and not the entire $N \times N^\beta$ Noise Collector matrix. In practice, we use $\beta \approx 1.5$ which makes the cost of using the Noise Collector negligible, as typically $K \gg N^{\beta-1}$. The columns of the Noise Collector matrix $\mathcal{C}$ with this circulant structure are uniformly distributed on $\mathbb{S}^{N-1}$ and they satisfy Lemma 1. This implies that the Theorems of this paper are still valid for such $\mathcal{C}$.

Application to imaging

We consider passive array imaging of point sources. The problem consists in determining the positions $\vec{\boldsymbol{z}}_j$ and the complex[†] amplitudes α_j, $j = 1, \ldots, M$, of a few point sources from measurements of polychromatic signals on an array of receivers; see Figure 2. The imaging system is characterized by the array aperture a, the distance L to the sources, the bandwidth B and the central wavelength λ_0.

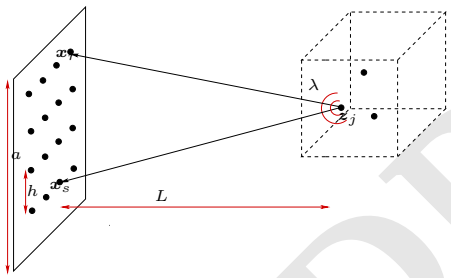

Fig. 2. General setup for passive array imaging. The source at $\vec{\boldsymbol{z}}_j$ emits a signal that is recorded at all array elements $\vec{\boldsymbol{x}}_r$, $r = 1, \ldots, N_r$.

The sources are located inside an image window IW, which is discretized with a uniform grid of points $\vec{\boldsymbol{y}}_k$, $k = 1, \ldots, K$. The unknown is the source vector $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_K]^\intercal \in \mathbb{C}^K$, whose components ρ_k correspond to the complex amplitudes of the M sources at the grid points $\vec{\boldsymbol{y}}_k$, $k = 1, \ldots, K$, with $K \gg M$. For the true source vector we have $\rho_k = \alpha_j$ if $\vec{\boldsymbol{y}}_k = \vec{\boldsymbol{z}}_j$ for some $j = 1, \ldots, M$, while $\rho_k = 0$ otherwise.

Denoting by $G(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}; \omega)$ the Green's function for the propagation of a signal of angular frequency ω from point $\vec{\boldsymbol{y}}$ to point $\vec{\boldsymbol{x}}$, we define the single-frequency Green's function vector that connects a point $\vec{\boldsymbol{y}}$ in the IW with all points $\vec{\boldsymbol{x}}_r$, $r = 1, \ldots, N_r$, on the array as

$$\boldsymbol{g}(\vec{\boldsymbol{y}}; \omega) = [G(\vec{\boldsymbol{x}}_1, \vec{\boldsymbol{y}}; \omega), G(\vec{\boldsymbol{x}}_2, \vec{\boldsymbol{y}}; \omega), \ldots, G(\vec{\boldsymbol{x}}_N, \vec{\boldsymbol{y}}; \omega)]^\intercal \in \mathbb{C}^{N_r}.$$

In three dimensions, $G(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}; \omega) = \dfrac{\exp\{\mathrm{i}\omega|\vec{\boldsymbol{x}} - \vec{\boldsymbol{y}}|/c_0\}}{4\pi|\vec{\boldsymbol{x}} - \vec{\boldsymbol{y}}|}$ if the medium is homogeneous. The data for the imaging problem are the signals $b(\vec{\boldsymbol{x}}_r, \omega_l) = \sum_{j=1}^M \alpha_j G(\vec{\boldsymbol{x}}_r, \vec{\boldsymbol{z}}_j; \omega_l)$ recorded at receiver locations $\vec{\boldsymbol{x}}_r$, $r = 1, \ldots, N_r$, at frequencies ω_l, $l = 1, \ldots, S$. These data are stacked in a column vector

$$\boldsymbol{b} = [\boldsymbol{b}(\omega_1)^\intercal, \boldsymbol{b}(\omega_2)^\intercal, \ldots, \boldsymbol{b}(\omega_S)^\intercal]^\intercal \in \mathbb{C}^N \quad ; \quad N = N_r S\,, \quad [15]$$

with $\boldsymbol{b}(\omega_l) = [b(\vec{\boldsymbol{x}}_1, \omega_l), b(\vec{\boldsymbol{x}}_2, \omega_l), \ldots, b(\vec{\boldsymbol{x}}_N, \omega_l)]^\intercal \in \mathbb{C}^{N_r}$. Then, $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$, with $\mathcal{A}$ the $N \times K$ measurement matrix whose columns $\boldsymbol{a}_k$ are the multiple-frequency Green's function vectors

$$\boldsymbol{a}_k = [\boldsymbol{g}(\vec{\boldsymbol{y}}_k; \omega_1)^\intercal, \boldsymbol{g}(\vec{\boldsymbol{y}}_k; \omega_2)^\intercal, \ldots, \boldsymbol{g}(\vec{\boldsymbol{y}}_k; \omega_S)^\intercal]^\intercal \in \mathbb{C}^N\,, \quad [16]$$

normalized to have length one. The system $\mathcal{A}\boldsymbol{\rho} = \boldsymbol{b}$ relates the unknown vector $\boldsymbol{\rho} \in \mathbb{C}^K$ to the data vector $\boldsymbol{b} \in \mathbb{C}^N$.

Next, we illustrate the performance of the Noise Collector in this imaging setup. The most important features are that (i) no calibration is necessary with respect to the level of noise, (ii) exact support recovery is obtained for relatively large levels of noise (i.e., $\|\boldsymbol{e}\|_{\ell_2} \leq c_1 \|\boldsymbol{b}_0\|_{\ell_2}^2 \sqrt{N}/(\|\boldsymbol{\rho}\|_{\ell_1}\sqrt{\ln N})$), and (iii) we have zero false discovery rates for all levels of noise, with high probability.

We consider a high frequency microwave imaging regime with central frequency $f_0 = 60$GHz corresponding to $\lambda_0 = 5$mm. We make measurements for $S = 25$ equally spaced frequencies spanning a bandwidth $B = 20$GHz. The array has $N = 25$ receivers and an aperture $a = 50$cm. The distance from the array to the center of the imaging window is $L = 50$cm. Then, the resolution is $\lambda_0 L/a = 5$mm in the cross-range (direction parallel to the array) and $c_0/B = 15$mm in range (direction of propagation). These parameters are typical in microwave scanning technology (33).

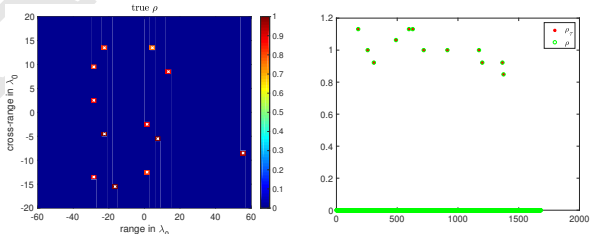

Fig. 3. Noiseless data. The exact solution is recovered for any value of λ in Algorithm 14; Left: the true image. Right: the recovered solution vector, $\boldsymbol{\rho}_\tau$, is plotted with red stars and the true solution vector, $\boldsymbol{\rho}$, with green circles.

We seek to image a source vector with sparsity $M = 12$; see the left plot in Fig. 3. The size of the imaging window is 20cm×60cm and the pixel spacing is 5mm×15mm. The number of unknowns is, therefore, $K = 1681$, and the number of data is $N = 625$. The size of the Noise Collector is taken to be $\Sigma = 10^4$, so $\beta \approx 1.5$. When the data is noiseless, we obtain exact recovery as expected; see the right plot in Fig. 3.

In Fig. 4, we display the imaging results, with and without the Noise Collector, when the data is corrupted by additive noise. The SNR $= 1$, so the ℓ_2-norms of the signals and the noise are equal. In the left plot, we show the recovered image using ℓ_1-norm minimization without the Noise Collector. There is a lot of grass in this image, with many non-zero values outside the true support. When the Noise Collector is used, the level of the grass is reduced and the image improves; see the second from the left plot. Still, there are several false discoveries because we use $\tau = 1$ in Algorithm 14.

In the third column from the left of Fig. 4 we show the image obtained with a weight $\tau = 2$ in Algorithm 14. With this

[†] We chose to work with real numbers in the previous sections for ease of presentation but the results also hold with complex numbers.

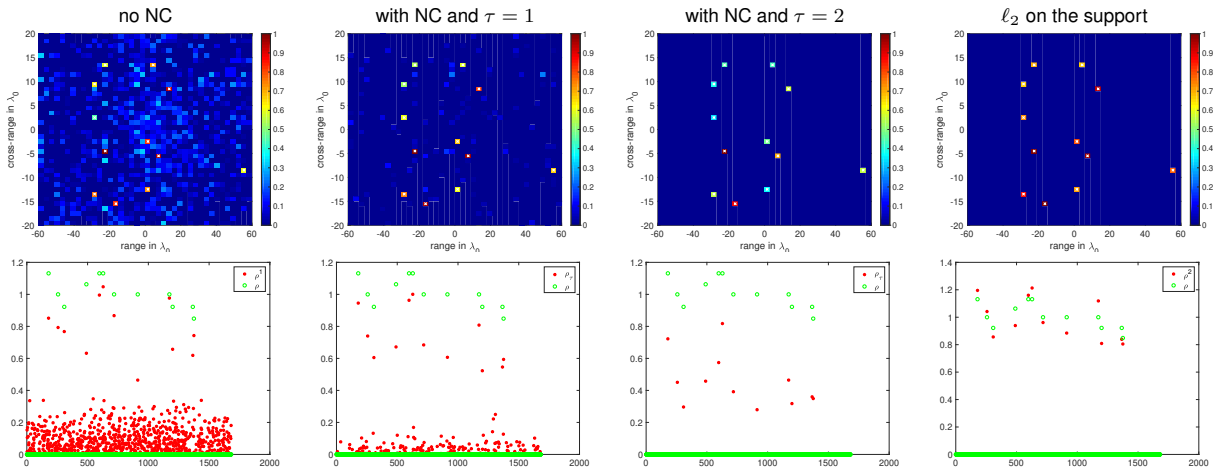

Fig. 4. High level of noise; SNR $= 1$. From left to right: ℓ_1-norm minimization without the noise collector; ℓ_1-norm minimization with a noise collector with $\Sigma = 10^4$ columns, and $\tau = 1$ in Algorithm 14; ℓ_1-norm minimization with a noise collector, and the correct $\tau = 2$ in Algorithm 14; ℓ_2-norm solution restricted to the support. In the top row we show the images, and in the bottom row the solution vector with red stars and the true solution vector with green circles.

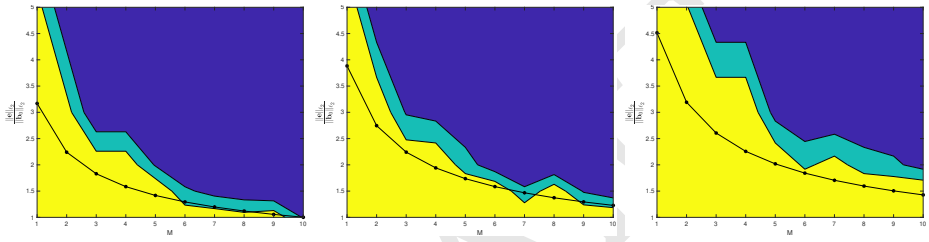

Fig. 5. Left: Algorithm performance for exact support recovery. Success corresponds to the value 1 (yellow) and failure to 0 (blue). The small phase transition zone (green) contains intermediate values. The black line is the theoretical estimate $\sqrt{N}/\sqrt{M \ln N}$. Ordinate and abscissa are the sparsity M and $\|\boldsymbol{e}\|_{\ell_2}/\|\boldsymbol{b}_0\|_{\ell_2}$. From left to right the data size is $N = 342$, $N = 625$ and $N = 961$.

weight, there are no false discoveries and the recovered support is exact. This simplifies the imaging problem dramatically, as we can now restrict the inverse problem to the true support just obtained, and then solve an overdetermined linear system using a classical ℓ_2 approach. The results are shown in the right column of Fig. 4. Note that this second step largely compensates for the signal that was lost in the first step due to the high level of noise.

In Figure 5 we illustrate the performance of the Noise Collector for different sparsity levels M and $\|\boldsymbol{e}\|_{\ell_2}/\|\boldsymbol{b}_0\|_{\ell_2}$ values. Success in recovering the true support of the unknown corresponds to the value 1 (yellow) and failure to 0 (blue). The small phase transition zone (green) contains intermediate values. The black line is the theoretical prediction Eq. 5. These results are obtained by averaging over 10 realizations of noise. We show results for three values of data size $N = 342$, $N = 625$ and $N = 961$. In our experiments the non zero components of the unknown $\boldsymbol{\rho}$ take values in $[0.6, 0.8]$ and, therefore, $\|\boldsymbol{b}_0\|_{\ell_2}/\|\boldsymbol{\rho}\|_{\ell_1} = cst/\sqrt{M}$.

Remark 1: We considered passive array imaging for ease of presentation. Same results hold for active array imaging with or without multiple scattering; see (34) for the detailed analytical setup.

Remark 2: We have considered a microwave imaging regime. Similar results can be obtained in other regimes.

Proofs

Proof of Lemma 1: Using the rotational invariance of all our probability distributions inequality 9 is true if

$$\mathbb{P}(\max_i |\langle \boldsymbol{d}_i, \boldsymbol{e}\rangle| \geq c_0\sqrt{\ln N}/\sqrt{N}) \leq 1/N^{\kappa},$$

where $\boldsymbol{d}_i$, $i = 1, 2, \ldots, K + \Sigma$ are (possibly dependent) uniformly distributed on $\mathbb{S}^{N-1}$, and we can assume $\boldsymbol{e} = (1, 0, \ldots, 0)$. Denote the event

$$\Omega_t = \left\{\max_i |\langle \boldsymbol{d}_i, \boldsymbol{e}\rangle| \geq t/\sqrt{N}\right\}.$$

$\mathbb{P}\left(|\langle \boldsymbol{d}_i, \boldsymbol{e}\rangle| \geq t/\sqrt{N}\right) \leq 2\exp(-t^2/2)$ for each $\boldsymbol{d}_i$. We obtain $\mathbb{P}(\Omega_t) \leq 2(K + \Sigma)\exp(-t^2/2)$ using the union bound. Choosing $t = c_0\sqrt{\ln N}$ for sufficiently large c_0, we get $\mathbb{P}(\Omega_t) \leq CN^{\beta}N^{-c_0^2/2} \leq N^{-\kappa}$, where $c_0^2 > 2(\beta + \kappa)$ and $N \geq N_0$. Hence, Eq. 9 holds with probability $1 - N^{-\kappa}$.

If N columns $\boldsymbol{c}_j$, $j \in S$ of $\mathcal{C}$ satisfy

$$\min_{j\in S} |\langle \boldsymbol{c}_j, \boldsymbol{e}\rangle| \geq \theta, \theta = \alpha\sqrt{\ln N}/\sqrt{N}, \qquad [17]$$

then their convex hull will contain $\theta\boldsymbol{e}$ with probability $(1/2)^N$. Therefore inequality 10 follows if 17 holds with probability $1 - 1/N^{\kappa}$. Using the rotational invariance of all our probability distributions we can assume $\boldsymbol{e} = (1, 0, \ldots, 0)$. For each $\boldsymbol{c}_i$

$$\mathbb{P}\left(|\langle \boldsymbol{c}_i, \boldsymbol{e}\rangle| \geq \frac{t}{\sqrt{N}}\right) = \frac{2}{\sqrt{2\pi}}\int_t^{\infty} \mathrm{e}^{-\frac{x^2}{2}}\, dx \geq \frac{1}{2}\mathrm{e}^{-t^2}.$$

Split the index set $1, 2, \ldots, \Sigma$ into N non-overlapping subsets S_k, $k = 1, 2, \ldots, N$ of size $N^{\beta-1}$. For each S_k

$$\mathbb{P}\left(\max_{i\in S_k} |\langle \boldsymbol{c}_i, \boldsymbol{e}\rangle| \le \frac{\alpha\sqrt{\ln N}}{\sqrt{N}}\right) \le \left(1 - \frac{1}{2N^{\alpha^2}}\right)^{N^{\beta-1}} \le \mathrm{e}^{-\frac{1}{2}N^{\frac{\beta-1}{2}}}$$

for $\alpha = \sqrt{(\beta-1)/2}$. By independence

$$\mathbb{P}\,(\mathbf{17} \text{ holds}) \ge \Pi_{k=1}^{N}\mathbb{P}(\max_{i\in S_k} |\langle \boldsymbol{c}_i, \boldsymbol{e}\rangle| \ge \alpha\sqrt{\ln N}/\sqrt{N}).$$

Then $\mathbb{P}\,(\mathbf{17} \text{ holds}) \ge (1 - \mathrm{e}^{-\frac{1}{2}N^{\frac{\beta-1}{2}}})^N \ge 1 - N\mathrm{e}^{-\frac{1}{2}N^{\frac{\beta-1}{2}}}$. Choosing N_0 sufficiently large, we obtain **10**. □

Proof of Theorem 2: When columns of $\mathcal{A}$ are not orthogonal, we will choose a τ smaller than that in Theorem 1 by a factor of two. Suppose the M-dimensional space V is the span of the column vectors $\boldsymbol{a}_j$, with j in the support of $\boldsymbol{\rho}$. Say, V is spanned by $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M$. Let $W = V^\perp$ be the orthogonal complement to V. Consider the orthogonal decomposition $\boldsymbol{a}_i = \boldsymbol{a}_i^v + \boldsymbol{a}_i^w$ for all $i \ge M+1$. Incoherence of $\boldsymbol{a}_i$ implies that $\|\boldsymbol{a}_i^w\|_{\ell_2} \ge 1/2$ for all $i \ge M+1$. Indeed, fix any $i \ge M+1$. Suppose $\boldsymbol{a}_i^v = \sum_{k=1}^{M} \xi_k \boldsymbol{a}_k$, and $|\xi_j| = \max_{k\le M} |\xi_k| = \|\boldsymbol{\xi}\|_{l_\infty}$. Thus, $\frac{1}{3M} \ge |\langle \boldsymbol{a}_j, \boldsymbol{a}_i^v\rangle| \ge |\langle \boldsymbol{a}_j, \sum_{k=1}^{M} \xi_k \boldsymbol{a}_k\rangle| \ge \|\boldsymbol{\xi}\|_{l_\infty}\left(1 - \frac{M-1}{3M}\right)$. Then $\|\boldsymbol{\xi}\|_{l_\infty} \le 1/(2M)$. So $\|\boldsymbol{a}_i^v\|_{\ell_2} \le \|\boldsymbol{\xi}\|_{\ell_1} \le M\|\boldsymbol{\xi}\|_{l_\infty} \le 1/2$, and $\|\boldsymbol{a}_i^w\|_{\ell_2} \ge \|\boldsymbol{a}_i\|_{\ell_2} - \|\boldsymbol{a}_i^v\|_{\ell_2} \ge 1/2$.

Project system **4** on W. Then, we obtain a new system **4**. The ℓ_2-norms of the columns of new $\mathcal{A}$ are at least $1/2$. Otherwise, the new system satisfies all conditions of Theorem 1. Indeed, $\boldsymbol{b}_0$ is projected to zero. All $\boldsymbol{c}_i$ and $\boldsymbol{e}/\|\boldsymbol{e}\|_{\ell_2}$ are projected to vectors uniformly distributed on $\mathbb{S}^{N-M-1}$ by the concentration of measure, see e.g. (27). If any $\boldsymbol{a}_i$, $i \ge M+1$, was used in an optimal approximation of $\boldsymbol{b}_0 + \boldsymbol{e}$, then its projection $\boldsymbol{a}_i^w$ is used in an optimal approximation of the projection of $\boldsymbol{b}_0 + \boldsymbol{e}$ on W. This is a contradiction with Lemma 1, if we choose $\tau < c_0/(2\alpha)$ and recall $\|\boldsymbol{a}_i^w\|_{\ell_2} \ge 1/2$. □

Proof of Theorem 3: Choose τ as in Theorem 2. Incoherence of $\boldsymbol{a}_i$ implies we can argue as in the proof of Theorem 2 and assume $\langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle = 0$ for $i \ne j$, $i, j \in \mathrm{supp}(\boldsymbol{\rho})$. Suppose V^i are the 2-dimensional spaces spanned by $\boldsymbol{e}$ and $\boldsymbol{a}_i$ for $i \in \mathrm{supp}(\boldsymbol{\rho})$. By Lemma 1 all $\lambda H(\tau) \cap V^i$ look like rounded rhombi depicted on Fig. 1-right, and $\lambda H_1 \cap V^i \subset B_\lambda^i$ with probability $1 - N^{-\kappa}$, where B_λ^i is a 2-dimensional ℓ_2-ball of radius $\lambda c_0\sqrt{\ln N}/\sqrt{N}$. Thus $\lambda H(\tau) \cap V^i \subset H_\lambda^i$ with probability $1 - N^{-\kappa}$, where H_λ^i is the convex hull of B_λ^i and a vector $\lambda \boldsymbol{f}_i$, $\boldsymbol{f}_i = \rho_i \|\boldsymbol{\rho}\|_{\ell_1}^{-1}\tau^{-1}\boldsymbol{a}_i$. Then $\mathrm{supp}(\boldsymbol{\rho}_\tau) = \mathrm{supp}(\boldsymbol{\rho})$, if there exists λ_0 so that $\rho_i\boldsymbol{a}_i + \boldsymbol{e}$ lies on the flat boundary of $H_{\lambda_0}^i$ for all $i \in \mathrm{supp}(\boldsymbol{\rho})$.

If $\min_{i\in\mathrm{supp}(\boldsymbol{\rho})} |\rho_i| \ge \gamma\|\boldsymbol{\rho}\|_\infty$, then there is a constant $c_2 = c_2(\gamma)$ such that if $\rho_i\boldsymbol{a}_i + \boldsymbol{e}$ lies on the flat boundary of H_λ^i for some i and some λ, then there exists λ_0 so that $\rho_i\boldsymbol{a}_i + c_2\boldsymbol{e}$ lies on the flat boundary of $H_{\lambda_0}^i$ for all $i \in \mathrm{supp}(\boldsymbol{\rho})$. Suppose V is spanned by $\boldsymbol{e}$ and $\boldsymbol{b}_0$, $H_\lambda \subset V$ is the convex hull of B_λ and $\lambda\boldsymbol{f}$, $\boldsymbol{f} = \boldsymbol{b}_0\|\boldsymbol{\rho}\|_{\ell_1}^{-1}\tau^{-1}$, where $B_\lambda \subset V$ is an ℓ_2-ball of radius $\lambda c_0\sqrt{\ln N}/\sqrt{N}$. If $\boldsymbol{b}_0 + c_2\boldsymbol{e}$ lies on the flat boundary of H_λ, then there must be an $i \in \mathrm{supp}(\boldsymbol{\rho})$ such that $\rho_i\boldsymbol{a}_i + c_2\boldsymbol{e}$ lies on the flat boundary of H_λ^i. If

$$\frac{|\langle \boldsymbol{b}_0, \boldsymbol{b}_0 + c_2\boldsymbol{e}\rangle|}{\|\boldsymbol{b}_0\|_{\ell_2}\|\boldsymbol{b}_0 + c_2\boldsymbol{e}\|_{\ell_2}} \ge \frac{c_0\sqrt{\ln N}}{\sqrt{N}\|\boldsymbol{f}\|_{\ell_2}}, \qquad [18]$$

then $\boldsymbol{b}_0 + c_2\boldsymbol{e}$ lies on the flat boundary of H_λ. Since $|\langle \boldsymbol{b}_0, \boldsymbol{e}\rangle| \le c_0\|\boldsymbol{e}\|_{\ell_2}\|\boldsymbol{b}_0\|_{\ell_2}/\sqrt{N}$ with probability $1 - N^{-\kappa}$, Eq. **18** holds if $\|\boldsymbol{e}\|_{\ell_2}/\|\boldsymbol{b}_0\|_{\ell_2} \le \|\boldsymbol{f}\|_{\ell_2}\sqrt{N}/(c_2c_0\sqrt{\ln N}) \le c_1\|\boldsymbol{b}_0\|_{\ell_2}\|\boldsymbol{\rho}\|_{\ell_1}^{-1}\sqrt{N}/\sqrt{\ln N}$. □

ACKNOWLEDGMENTS. The work of M. Moscoso was partially supported by Spanish grant MICINN FIS2016-77892-R. The work of A.Novikov was partially supported by NSF grants DMS-1515187, DMS-1813943. The work of G. Papanicolaou was partially supported by AFOSR FA9550-18-1-0519. The work of C. Tsogka was partially supported by AFOSR FA9550-17-1-0238 and FA9550-18-1-0519. We thank Marguerite Novikov for drawing the left part of Figure 1.

Data availability statement: There are no data in this paper; we present a new algorithm, its theoretical analysis, and some numerical simulations.

1. Chen SS, Donoho DL, Saunders MA (2001) Atomic decomposition by basis pursuit. *SIAM Rev* 43:12–159.
2. Donoho DL, Elad M (2003) Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proc. Natl. Acad. Sci. U.S.A.* 100(5):2197–2202.
3. Candès EJ, Tao T (2005) Decoding by linear programming. *IEEE Trans. Inf. Theory* 51:4203–4215.
4. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (methodological)* 58:267–288.
5. Fuchs JJ (2005) Recovery of exact sparse representations in the presence of bounded noise. *IEEE Trans. Inf. Theory* 51:3601–3608.
6. Tropp JA (2006) Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inf. Theory* 52:1030–1051.
7. Wainwright MJ (2009) Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inf. Theory* 55:2183–2202.
8. Maleki A, Anitori L, Yang Z, Baraniuk R (2013) Asymptotic analysis of complex lasso via complex approximate message passing (CAMP). *IEEE Trans. Inf. Theory* 59:4290–4308.
9. Zou H (2006) The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc* 101:1418–1429.
10. Sampson JN, Chatterjee N, Carroll RJ, Müller S (2013) Controlling the local false discovery rate in the adaptive lasso. *Biostatistics* 14:653–666.
11. Belloni A, Chernozhukov V, Wang L (2011) Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika* 98(4):791–806.
12. Trzasko J, Manduca A (2009) Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization. *IEEE Trans. Med. Imag* 28:106–121.
13. AlQuraishi M, McAdams HH (2011) Direct inference of protein dna interactions using compressed sensing methods. *Proc. Natl. Acad. Sci. U.S.A.* 108(36):14819–14824.
14. Baraniuk R, Steeghs P (2007) Compressive radar imaging in *2007 IEEE Radar Conference*. pp. 128–133.
15. Taylor HL, Banks SC, McCoy JF (1979) Deconvolution with the l1 norm. *Geophysics* 44:39–52.
16. Malioutov D, Cetin M, Willsky AS (2005) A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process* 53:3010–3022.
17. Romberg J (2008) Imaging via compressive sampling. *IEEE Signal Process. Mag* 25:14–20.
18. Herman MA, Strohmer T (2009) High-resolution radar via compressed sensing. *IEEE Trans. Signal Process* 57:2275–2284.
19. Tropp JA, Laska JN, Duarte MF, Romberg JK, Baraniuk RG (2010) Beyond Nyquist: Efficient sampling of sparse bandlimited signals. *IEEE Transactions on Information Theory* 56:520–544.
20. Fannjiang AC, Strohmer T, Yan P (2010) Compressed remote sensing of sparse objects. *SIAM J. Imag. Sci* 3:595–618.
21. Chai A, Moscoso M, Papanicolaou G (2013) Robust imaging of localized scatterers using the singular value decomposition and ℓ_1 optimization. *Inverse Problems* 29:025016.
22. Donoho DL (1992) Super-resolution via sparsity constraint. *SIAM J Math Anal* 23:1303–1331.
23. Candès EJ, Fernandez-Granda C (2014) Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math* 67:906–956.
24. Fannjiang AC, Liao W (2012) Coherence pattern-guided compressive sensing with unresolved grids. *SIAM J. Imag. Sci* 5:179–202.
25. Borcea L, Kocyigit I (2015) Resolution analysis of imaging with ℓ_1 optimization. *SIAM J. Imaging Sci* 8:3015–3050.
26. Moscoso M, Novikov A, Papanicolaou G, Tsogka C (2020) Imaging with highly incomplete and corrupted data. *Inverse Problems* 36:035010.
27. Vershynin R (2018) *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge Series in Statistical and Probabilistic Mathematics. (Cambridge University Press).
28. Laska JN, Davenport MA, Baraniuk RG (2009) Exact signal recovery from sparsely corrupted measurements through the pursuit of justice in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals*. (Systems and Computers, Pacific Grove, CA), pp. 1556–1560.
29. Tao T (2013) A cheap version of the kabatjanskii-levenstein bound for almost orthogonal vectors. https://terrytao.wordpress.com/2013/07/18/a-cheap-version-of-the-kabatjanskii-levenstein-bound-for-almost-orthogonal-vectors/.
30. Milman VD (1971) A new proof of dvoretzky's theorem on cross-sections of convex bodies in *Funkcional. Anal. I Priložen.* Vol. 5, pp. 28–37.
31. Moscoso M, Novikov A, Papanicolaou G, Ryzhik L (2012) A differential equations approach to l1-minimization with applications to array imaging. *Inverse Problems* 28.
32. Gray RM (2006) Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory* 2:155–239.
33. Laviada J, Arboleya-Arboleya A, Alvarez-Lopez Y, Garcia-Gonzalez C, Las-Heras F (2015) Phaseless synthetic aperture radar with efficient sampling for broadband near-field imaging: Theory and validation, ieee trans. *Antennas Propag.* 63(2):573–584.
34. Chai A, Moscoso M, Papanicolaou G (2014) Imaging strong localized scatterers with sparsity promoting optimization. *SIAM J. Imaging Sci* 7:1358–1387.