

The MUSIC algorithm for sparse objects: a compressed sensing analysis

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2011 Inverse Problems 27 035013

(<http://iopscience.iop.org/0266-5611/27/3/035013>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 169.237.215.179

The article was downloaded on 17/02/2011 at 02:24

Please note that [terms and conditions apply](#).

The MUSIC algorithm for sparse objects: a compressed sensing analysis

Albert C Fannjiang

Department of Mathematics, University of California, Davis, CA 95616-8633, USA

E-mail: fannjiang@math.ucdavis.edu

Received 9 June 2010, in final form 18 January 2011

Published 16 February 2011

Online at stacks.iop.org/IP/27/035013

Abstract

The multiple signal classification (MUSIC) algorithm, and its extension for imaging sparse *extended* objects, with noisy data is analyzed by compressed sensing (CS) techniques. A thresholding rule is developed to augment the standard MUSIC algorithm. The notion of restricted isometry property (RIP) and an upper bound on the restricted isometry constant (RIC) are employed to establish sufficient conditions for the exact localization by MUSIC with or without noise. In the noiseless case, the sufficient condition gives an upper bound on the numbers of random sampling and incident directions necessary for exact localization. In the noisy case, the sufficient condition assumes additionally an upper bound for the noise-to-object ratio in terms of the RIC and the dynamic range of objects. This bound points to the super-resolution capability of the MUSIC algorithm. Rigorous comparison of performance between MUSIC and the CS minimization principle, basis pursuit denoising (BPDN), is given. In general, the MUSIC algorithm guarantees to recover, with high probability, s scatterers with $n = \mathcal{O}(s^2)$ random sampling and incident directions and sufficiently high frequency. For the favorable imaging geometry where the scatterers are distributed on a transverse plane MUSIC guarantees to recover, with high probability, s scatterers with a median frequency and $n = \mathcal{O}(s)$ random sampling/incident directions. Moreover, for the problems of spectral estimation and source localizations both BPDN and MUSIC guarantee, with high probability, to identify exactly the frequencies of random signals with the number $n = \mathcal{O}(s)$ of sampling times. However, in the absence of abundant realizations of signals, BPDN is the preferred method for spectral estimation. Indeed, BPDN can identify the frequencies approximately with just *one* realization of signals with the recovery error at worst linearly proportional to the noise level. Numerical results confirm that BPDN outperforms MUSIC in the well-resolved case while the opposite is true for the under-resolved case, giving abundant evidence for the super-resolution capability of the MUSIC algorithm. Another advantage of MUSIC over BPDN is the former's flexibility with grid spacing and the guarantee of *approximate* localization of sufficiently

separated objects in an arbitrarily refined grid. The localization error is bounded from above by $\mathcal{O}(\lambda s)$ for general configurations and by $\mathcal{O}(\lambda)$ for objects distributed in a transverse plane, in line with physical intuition.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The multiple signal classification (MUSIC) algorithm is a well-known method in signal processing for estimating the individual frequencies of multiple time-harmonic signals [7, 27]. Mathematically, MUSIC is essentially a method of characterizing the range of the covariance matrix of the signals (see section 6 for details).

MUSIC was originally developed to estimate the direction of arrival for source localization [25]. Later, the MUSIC algorithm was extended to imaging of point scatterers [9, 10, 17]. A proof of a sufficient condition for the exact recovery of the object support in the noiseless case is given in [22] (see also [21]) which is reproduced in proposition 1 below. The performance guarantee is general but qualitative in nature. Neither does it take into account the effect of noise which is important for assessing the super-resolution effect.

The main purpose of this paper is to give a quantitative performance evaluation for the MUSIC algorithm in terms of how many data are needed and how they may be collected in order to exactly recover the locations of a given (large) number of objects, be they sources, scatterers or frequencies as well as how much noise the MUSIC algorithm can tolerate. Our approach is based on recent advances in compressed sensing theory ([2, 3, 24] and references therein) and applications to imaging (see [12, 13, 16] and references therein).

A main result for localizing scatterers obtained in this paper has the following flavor (more details later). Let ξ_{\max} and ξ_{\min} be, respectively, the strengths of the strongest and weakest (nonzero) scatterers, δ_s^\pm the (upper/lower) restricted isometry constants (RIC) of order s and ε the level of noise in the data. If the noise-to-scatterer ratio (NSR) obeys the upper bound

$$\frac{\varepsilon}{\xi_{\min}} < \sqrt{(1 + \delta_s^+)^2 \frac{\xi_{\max}^2}{\xi_{\min}^2} + (1 - \delta_s^-)^2 \tilde{\Delta}} - (1 + \delta_s^+) \frac{\xi_{\max}}{\xi_{\min}} \quad (1)$$

where

$$\tilde{\Delta} = \frac{1}{2} - \frac{1}{2} \frac{1}{\sqrt{\sqrt{2}\gamma_s + 1}}, \quad \gamma_s = 1 - \frac{\delta_{s+1}^- (1 + \delta_s^+)}{2 + \delta_s^+ - \delta_{s+1}^-}$$

then the MUSIC imaging function J^ε (defined in (36)) with the thresholding rule

$$\{\mathbf{r} \in \mathcal{K} : J^\varepsilon(\mathbf{r}) \geq 2\gamma_s^{-2}\} \quad (2)$$

recovers exactly the locations of the s scatterers (cf theorem 2, section 3). Compressed sensing theory comes into play in addressing the dependence of RIC on the frequency, the number and distribution of random sampling directions (or sensors), the number of scatterers and the inter-scatterer distances.

In the under-resolved case, the δ_s^- tends to 1 and Γ_s tends to zero, rendering the right-hand side of (1) approximately

$$\frac{(1 - \delta_s^-)^2 \tilde{\Delta}}{2(1 + \delta_s^+) \xi_{\max} / \xi_{\min}} \quad (3)$$

where ξ_{\max} / ξ_{\min} is the dynamic range of the scatterers. For a NSR smaller than (3) the s scatterers can still be perfectly localized by the MUSIC algorithm with the thresholding

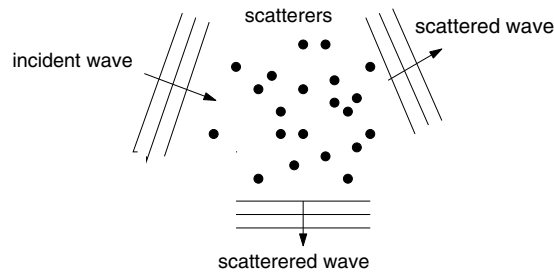


Figure 1. Scattering measurement: for each incident wave, the scattering amplitudes in multiple directions are measured.

rule (2) where the threshold is approximately

$$\frac{2}{(1 - \delta_{s+1}^-)^2} \left(\frac{1 + \delta_s^+}{2 + \delta_s^+} \right)^2 \gg 1.$$

Previous observation [23] and our numerical results (section 8) lend support to this super-resolution effect of the MUSIC algorithm.

First let us review the inverse scattering problem and the MUSIC imaging method.

1.1. Inverse scattering

Consider the scattering of the incident plane wave

$$u^i(\mathbf{r}) = e^{i\omega\mathbf{r}\cdot\hat{\mathbf{d}}} \quad (4)$$

by the variable refractive index $n(\mathbf{r}) = \sqrt{1 + \xi(\mathbf{r})}$ where $\hat{\mathbf{d}}$ is the incident direction (figure 1). The resulting total wave field u satisfies the Helmholtz equation

$$\Delta u(\mathbf{r}) + \omega^2(1 + \xi(\mathbf{r}))u(\mathbf{r}) = 0. \quad (5)$$

We assume that the wave speed is unity and hence the frequency equals the wave number ω .

We write the total field $u = u^i + u^s$ as the sum of the incident field u^i and the scattered field u^s . Since $(\nabla^2 + \omega^2)u^i = 0$, the scattered field satisfies

$$-(\nabla^2 + \omega^2)u^s = \omega^2\xi u, \quad (6)$$

subject to the radiation condition

$$\lim_{|\mathbf{r}| \rightarrow \infty} |\mathbf{r}|^{(d-1)/2} (\hat{\mathbf{r}} \cdot \nabla u^s(\mathbf{r}) - i\omega u^s(\mathbf{r})) = 0, \quad \hat{\mathbf{r}} = \mathbf{r}/|\mathbf{r}|, \quad d = 2, 3$$

which distinguishes the scattered field from the incident field. Invoking Green's function G , satisfying the radiation condition, of the free-space Helmholtz operator $-(\nabla^2 + \omega^2)$, we convert (5) into the Lippmann–Schwinger integral equation

$$u^s(\mathbf{r}) = \omega^2 \int_{\mathbb{R}^d} \xi(\mathbf{r}') (u^i(\mathbf{r}') + u^s(\mathbf{r}')) G(\mathbf{r}, \mathbf{r}') d\mathbf{r}'. \quad (7)$$

As a result of the radiation condition, Green's function $G(\mathbf{r}, \mathbf{r}')$ has the following far-field asymptotic [22]:

$$G(\mathbf{r}, \mathbf{r}') = \frac{e^{i\omega|\mathbf{r}|}}{4\pi|\mathbf{r}|^{(d-1)/2}} (e^{-i\omega\mathbf{r}'\cdot\hat{\mathbf{r}}} + \mathcal{O}(|\mathbf{r}|^{-1})), \quad \hat{\mathbf{r}} = \mathbf{r}/|\mathbf{r}|, \quad |\mathbf{r}| \gg 1$$

and therefore by (7) the scattered field has the asymptotic

$$u^s(\mathbf{r}) = \frac{e^{i\omega|\mathbf{r}|}}{|\mathbf{r}|^{(d-1)/2}} (A(\hat{\mathbf{r}}, \hat{\mathbf{d}}) + \mathcal{O}(|\mathbf{r}|^{-1})), \quad (8)$$

where the scattering amplitude A is given by the integral formula

$$A(\hat{\mathbf{r}}, \hat{\mathbf{d}}) = \frac{\omega^2}{4\pi} \int_{\mathbb{R}^d} d\mathbf{r}' \xi(\mathbf{r}') u(\mathbf{r}') e^{-i\omega \mathbf{r}' \cdot \hat{\mathbf{r}}}. \quad (9)$$

In inverse scattering, the scattering amplitude is the observable data and the main objective then is to reconstruct ξ from the knowledge of the scattering amplitude. Note that since u in (9) is part of the unknown, the inverse scattering problem is nonlinear. In physical terms, the nonlinearity is the consequence of multiple scattering between different parts of medium inhomogeneities. In the Born scattering regime, the total field u on the right-hand side of (9) can be replaced by the incident field u^i , linearizing the inverse scattering problem.

Next we recall the MUSIC algorithm as applied to localization of point scatterers.

1.2. MUSIC for point scatterers

Let $S = \{\mathbf{r}_j : j = 1, \dots, s\}$ be the locations of the scatterers. Let $\xi_j \neq 0$, $j = 1, \dots, s$, be the strength of the scatterers. We will make the Born approximation first and discuss how to lift this restriction at the end of the section (remark 2). For the discrete medium $\xi(\mathbf{r}) = \sum_j \xi_j \delta(\mathbf{r} - \mathbf{r}_j)$ and the scattering amplitude (9) becomes the finite sum

$$A(\hat{\mathbf{r}}, \hat{\mathbf{d}}) = \frac{\omega^2}{4\pi} \sum_{j=1}^s \xi_j u^i(\mathbf{r}_j) e^{-i\omega \mathbf{r}_j \cdot \hat{\mathbf{r}}} \quad (10)$$

under the Born approximation.

Let $\hat{\mathbf{d}}_l$, $l = 1, \dots, m$, and $\hat{\mathbf{s}}_k$, $k = 1, \dots, n$, be, respectively, the incident and sampling directions. For each incident field $\hat{\mathbf{d}}_l$, $l = 1, \dots, m$, the scattering amplitude is measured in all n directions $\hat{\mathbf{s}}_k$, $k = 1, \dots, n$. The whole measurement data consist of the scattering amplitudes for all pairs of $(\hat{\mathbf{d}}_l, \hat{\mathbf{s}}_k)$.

Define the data matrix $\mathbf{Y} = (Y_{k,l}) \in \mathbb{C}^{n \times m}$ as

$$Y_{k,l} \sim A(\hat{\mathbf{s}}_k, \hat{\mathbf{d}}_l), \quad k = 1, \dots, n, \quad l = 1, \dots, m \quad (11)$$

where we keep open the option of normalizing \mathbf{Y} in order to simplify the setup. The data matrix is related to the object matrix

$$\mathbf{X} = \text{diag}(\xi_j) \in \mathbb{C}^{s \times s}, \quad j = 1, \dots, s,$$

by the measurement matrices Φ and Ψ as

$$\mathbf{Y} = \Phi \mathbf{X} \Psi^* \quad (12)$$

where Φ and Ψ are, respectively,

$$\Phi_{k,j} = \frac{1}{\sqrt{n}} e^{-i\omega \hat{\mathbf{s}}_k \cdot \mathbf{r}_j} \in \mathbb{C}^{n \times s} \quad (13)$$

$$\Psi_{l,j} = \frac{1}{\sqrt{n}} e^{-i\omega \hat{\mathbf{d}}_l \cdot \mathbf{r}_j} \in \mathbb{C}^{m \times s} \quad (14)$$

after proper normalization. Both (13) and (14) are normalized to have columns of unit 2-norm. We extend the formulation (12)–(14) to the case of sparse *extended* objects in appendix A.

Note that both Φ and Ψ are unknown and (12) can be inverted only after the locations of scatterers are determined. This is what the MUSIC algorithm is designed to accomplish.

The standard version of the MUSIC algorithm deals with the case of $n = m$ and $\hat{\mathbf{s}}_k = \hat{\mathbf{d}}_k$, $k = 1, \dots, n$, as stated in the following result.

Table 1. The idealized MUSIC algorithm when the noise and round-off error are absent.

Idealized MUSIC algorithm

- Input \mathbf{Y}, Φ .
- Compute the orthogonal projector \mathcal{P} onto the null space of \mathbf{Y}^* .
- Plot $J(\mathbf{r}) = |\mathcal{P}\phi_{\mathbf{r}}|^{-2}, \mathbf{r} \in \mathcal{K}$.
- Identify the singularities of J as the object locations.

Table 2. MUSIC algorithm with thresholding (66) when the data are contaminated by noise.

MUSIC algorithm with thresholding for equation (22)

- Input $\mathbf{Y}^\epsilon, \Phi$ and the sparsity s .
- Compute the orthogonal projector \mathcal{P}^ϵ onto the noise subspace of $\mathcal{Y}^\epsilon = \mathbf{Y}^\epsilon \mathbf{Y}^{\epsilon*}$.
- Compute the function $J^\epsilon(\mathbf{r}) = |\mathcal{P}^\epsilon \phi_{\mathbf{r}}|^{-2}, \mathbf{r} \in \mathcal{K}$.
- Output the set of points corresponding to the s highest values of J^ϵ or equivalently the set $\{\mathbf{r} \in \mathcal{K} : J^\epsilon(\mathbf{r}) \geq 2\gamma_s^{-2}\}$ where $\gamma_s = 1 - \frac{\delta_{s+1}^-(1+\delta_s^+)}{2+\delta_s^+ - \delta_{s+1}^-}$ as the object locations.

Proposition 1 [21, 22]. *Let $\{\hat{\mathbf{s}}_k = \hat{\mathbf{d}}_k, k \in \mathbb{N}\}$ be a countable set of directions such that any real analytic function on the unit sphere that vanishes in $\hat{\mathbf{s}}_k, \forall k \in \mathbb{N}$, vanishes identically. Let $\mathcal{K} \subset \mathbb{R}^3$ be a compact subset containing \mathcal{S} . Then n_0 exists such that for any $n \geq n_0$ the following characterization holds for every $\mathbf{r} \in \mathcal{K}$:*

$$\mathbf{r} \in \mathcal{S} \text{ if and only if } \phi_{\mathbf{r}} \equiv \frac{1}{\sqrt{n}}(e^{-i\omega\hat{\mathbf{s}}_1 \cdot \mathbf{r}}, e^{-i\omega\hat{\mathbf{s}}_2 \cdot \mathbf{r}}, \dots, e^{-i\omega\hat{\mathbf{s}}_n \cdot \mathbf{r}})^T \in \text{Ran}(\Phi). \quad (15)$$

Moreover, the ranges of Φ and \mathbf{Y} coincide.

Remark 1. As a consequence, $\mathbf{r} \in \mathcal{S}$ if and only if $\mathcal{P}\phi_{\mathbf{r}} = 0$ where \mathcal{P} is the orthogonal projection onto the null space of \mathbf{Y}^* (Fredholm alternative) and the locations of the scatterers can be identified by the singularities of the imaging function [7]

$$J(\mathbf{r}) = \frac{1}{|\mathcal{P}\phi_{\mathbf{r}}|^2}. \quad (16)$$

Therefore in the case without noise and round-off errors we can state the idealized form of the MUSIC algorithm as given in table 1.

Once the locations are exactly recovered, then both Φ and Ψ are known explicitly and the strength $\xi_j, j = 1, \dots, s$, of scatterers can be determined by inverting the linear equation (12) which is an over-determined system, given the support of scatterers.

In the realistic case when the noise is present in the data, the singularities generally disappear from J and a thresholding rule is necessary to complete the algorithm. The thresholding rule is part of the contribution of this paper (see table 2).

Remark 2. The assumptions of proposition 1 can be relaxed: instead of $\hat{\mathbf{s}}_k = \hat{\mathbf{d}}_k, \forall k$, it suffices to have $\Psi \in \mathbb{C}^{m \times s}$ which has rank s .

In light of this observation, it is also straightforward to extend the performance guarantee for the Born scattering case to the multiple-scattering case. In the latter case, Ψ consists of entries which are the total field evaluated at \mathbf{r}_j for the incident direction $\hat{\mathbf{d}}_l$, i.e.

$$\Psi_{l,j} = u^*(\mathbf{r}_j; \hat{\mathbf{d}}_l). \quad (17)$$

What is really needed is that $\Psi \in \mathbb{C}^{m \times s}$ has rank s since then $\mathbf{Z} = \mathbf{X}\Psi^*$ and \mathbf{X} shares the same support (see more on this in section 2). Generically this is true for sufficiently large m as we will show below.

Define the incident and full field vectors at the locations of the scatterers:

$$\begin{aligned} U^i(\hat{\mathbf{d}}_l) &= (u^i(\mathbf{r}_1; \hat{\mathbf{d}}_l), \dots, u^i(\mathbf{r}_s; \hat{\mathbf{d}}_l))^T \in \mathbb{C}^s \\ U(\hat{\mathbf{d}}_l) &= (u(\mathbf{r}_1; \hat{\mathbf{d}}_l), \dots, u(\mathbf{r}_s; \hat{\mathbf{d}}_l))^T \in \mathbb{C}^s. \end{aligned}$$

Denote

$$\mathbf{G} = [(1 - \delta_{ij})G(\mathbf{r}_j, \mathbf{r}_i)] \in \mathbb{C}^{s \times s}. \quad (18)$$

The discrete version of the Lippmann–Schwinger equation (i.e. the Foldy–Lax equation) can be written as

$$U(\hat{\mathbf{d}}_l) = U^i(\hat{\mathbf{d}}_l) + \omega^2 \mathbf{G} \mathbf{X} U(\hat{\mathbf{d}}_l), \quad l = 1, \dots, m. \quad (19)$$

The δ_{ij} terms in (18) represent the singular self-energy terms of point scatterers and should be removed for self-consistency.

Denote $\mathbf{U}^i = [U^i(\hat{\mathbf{d}}_1), \dots, U^i(\hat{\mathbf{d}}_m)] \in \mathbb{C}^{s \times m}$ and $\mathbf{U} = [U(\hat{\mathbf{d}}_1), \dots, U(\hat{\mathbf{d}}_m)] \in \mathbb{C}^{s \times m}$. Suppose that ω^{-2} is not an eigenvalue of $\mathbf{G} \mathbf{X}$. Then we can invert equation (19) to obtain

$$\mathbf{U} = (\mathbf{I} - \omega^2 \mathbf{G} \mathbf{X})^{-1} \mathbf{U}^i.$$

Hence \mathbf{U} has rank s if \mathbf{U}^i does. Indeed, for sufficiently high frequency ω and m randomly selected incident directions with sufficiently large ratio \sqrt{m}/s , \mathbf{U}^i has rank s with high probability (see propositions 3 and 4).

For some special imaging geometry it is possible to reduce the number of incident and sampling directions to $\mathcal{O}(s)$ (section 4).

1.3. Outline

Proposition 1 says that if the number of sampling directions is sufficiently large then the locations of the s scatterers can be identified by the s singularities of J . However, the condition is only qualitative in the sense that an estimate for the threshold n_0 is not given. It would be of obvious interest to know, e.g., how n_0 scales with s when s is large and when the conventional wisdom ($n_0 = s + 1$) derived from counting dimensions is true. Also how much noise can the MUSIC algorithm tolerate?

However, unless additional constraints are imposed on the measurement scheme (the frequency, the incident and sampling directions, etc), it is unlikely to make progress toward obtaining a useful estimate which is the objective of this study. In [9] a geometric constraint on the configuration of sensors and objects has been pointed out for exact recovery in the absence of noise. Moreover, it seems possible that a nonvanishing portion of s randomly distributed scatterers may not be exactly recovered in the presence of machine error no matter how large n is (figure 5, middle panel, and figure 7, right panel).

Let us briefly sketch our approach and results: we shall discretize the problem by using a finite grid for the computation domain \mathcal{K} and put the problem in a probabilistic setting by using random sampling directions. Moreover, we consider noisy data and aim for a result for stable recovery by MUSIC. For the case of well-resolved grids, we show by using the compressed sensing techniques that for the NSR obeying (1) and with high probability, $n_0 = \mathcal{O}(s^2)$ for general configuration of s objects and $n_0 = \mathcal{O}(s)$ for objects distributed on a transverse plane. For the case of under-resolved grids, we seek sufficient conditions for approximate, instead of exact, localization of objects and we show that for sufficiently small NSR and with high probability, the localization error is $\mathcal{O}(\lambda s)$ with $n = \mathcal{O}(s^2)$ for a general object configuration and the localization error is $\mathcal{O}(\lambda)$ with $n = \mathcal{O}(s)$ for objects distributed in a transverse plane.

Our plan for the rest of the paper is to first give a sensitivity analysis for MUSIC and derive the condition for exact recovery with noisy data under which the MUSIC algorithm based on

the perturbed data matrix can still recover exactly the object support (section 2). Next, we review the basic notion of compressed sensing (CS) theory and show how it naturally lends itself to a proof of exact localization by MUSIC (section 3). We show that with generic, random sampling and sufficiently high frequency the MUSIC algorithm can, with high probability, recover s scatterers with $n = \mathcal{O}(s^2)$ sampling and incident directions (corollary 2). Then we consider a favorable imaging geometry where the scatterers are distributed on a transverse plane (section 4). We show that with median frequency the MUSIC algorithm can recover, with high probability, s scatterers with $n = \mathcal{O}(s)$ sampling and incident directions (corollary 3). Next we analyze the performance guarantee of the compressed sensing principle, basis pursuit denoising (BPDN) (section 5) and show that in the generic situation BPDN with sufficiently high frequency can recover s scatterers with $n = \mathcal{O}(s^2)$ sampling directions and just one incident wave (remark 11) while for the favorable geometry of planar objects BPDN can recover s scatterers with $n = \mathcal{O}(s)$ sampling directions and one incident wave (remark 10). In section 6 we return to the original applications of MUSIC and perform the compressed sensing analysis of the performance of MUSIC as applied to spectral estimation and source localization. We show that the MUSIC algorithm can, with high probability, identify exactly the frequencies of random signals with the number $n = \mathcal{O}(s)$ of sampling times (theorem 5 and remark 12). We discuss MUSIC in the setting with an arbitrarily fine grid and give error bounds in section 7. Numerical tests are given in section 8 where the super-resolution capability of MUSIC and the noise sensitivity are studied. We conclude in section 9 with a discussion on the pros and cons of MUSIC versus the orthogonal matching pursuit (OMP) as applied to reconstruction based on multiple measurement vectors. We give an extension of the MUSIC algorithm to the case of extended objects in appendix A and a proof of the performance guarantee for the complex-setting BPDN under the asymmetrical RIP in appendix B.

2. Sensitivity analysis

For quantitative performance analysis of the MUSIC algorithm, we will work with the discrete setting and assume that \mathcal{K} is a discrete set of N , typically large, number of points, i.e. the computation grid. The discrete setting appears naturally in applying MUSIC to imaging of extended scatterers (see appendix A). Moreover, we consider the extension $\tilde{\Phi}$ of Φ which includes not only the columns $\phi_{\mathbf{r}_j}$, $j = 1, \dots, s$, representing the locations of the objects but also the columns representing all the points in \mathcal{K} . Hence, $\tilde{\Phi} \in \mathbb{C}^{n \times N}$ and as usual $\tilde{\Phi}$ is normalized so that the columns have unit 2-norm. The ordering of the columns of $\tilde{\Phi}$ is not important for our purpose as long as they correspond to the points in \mathcal{K} in a well-defined manner. $\tilde{\Psi} \in \mathbb{C}^{n \times N}$ is similarly defined. Also the extension $\tilde{\mathbf{X}} \in \mathbb{C}^{N \times N}$ of \mathbf{X} is defined by filling in zeros in all the entries outside the object support.

In terms of these notations, we can write

$$\mathbf{Y} = \tilde{\Phi} \tilde{\mathbf{X}} \tilde{\Psi}^* = \sum_{j=1}^N \tilde{\Phi}_j \otimes \tilde{\Psi}_j^* \xi_j. \quad (20)$$

By a slight abuse of notation, we shall use \mathcal{S} to denote the locations of objects in the physical domain as well as the corresponding index set. Likewise \mathcal{S}^c denotes the complement set of \mathcal{S} in the computation grid \mathcal{K} as well as the total index set $\{1, \dots, N\}$. In the same vein, $\tilde{\Phi}_{\mathcal{S}}$ denotes the column submatrix of $\tilde{\Phi}$ restricted to the index set \mathcal{S} . Hence, $\tilde{\Phi}_{\mathcal{S}} = \tilde{\Phi}$ and $\tilde{\Psi}_{\mathcal{S}} = \tilde{\Psi}$.

First, let us reformulate condition (15) for exact recovery as follows.

Note that

$$\Phi\Phi^\dagger\phi_{\mathbf{r}}$$

is the orthogonal projection of $\phi_{\mathbf{r}}$ onto the range of Φ where Φ^\dagger is the pseudo-inverse of Φ . Hence,

$$\mathcal{P}\phi_{\mathbf{r}} = (\mathbf{I} - \Phi\Phi^\dagger)\phi_{\mathbf{r}}.$$

If $\phi_{\mathbf{r}}$ for $\mathbf{r} \in S^c$ is independent of the columns of Φ , then

$$\phi_{\mathbf{r}}^*\Phi\Phi^\dagger\phi_{\mathbf{r}} < \|\phi_{\mathbf{r}}\|^2$$

and vice versa. Therefore (15) is equivalent to

$$\Gamma_S \equiv \min_{\mathbf{r} \in S^c} \|\phi_{\mathbf{r}}\|_2^{-1} \|\mathcal{P}\phi_{\mathbf{r}}\|_2 = \sqrt{1 - \max_{\mathbf{r} \in S^c} \|\phi_{\mathbf{r}}\|^{-2} \phi_{\mathbf{r}}^* \Phi\Phi^\dagger \phi_{\mathbf{r}}} > 0. \quad (21)$$

The number Γ_S gives a measure of how ‘independent’ ϕ_z is from the range of Φ uniformly in $\mathbf{r} \in S^c$.

Now we give a sensitivity analysis for MUSIC with respect to perturbation in the data matrix \mathbf{Y} in terms of Γ_S and other parameters. We want to show what else is needed, in addition to (15), to guarantee exact recovery of the support of scatterers when the data matrix is perturbed.

The general data matrix considered in this paper has the form $\mathbf{Y}^\varepsilon = \mathbf{Y} + \mathbf{E}$ where $\mathbf{Y} = \Phi\mathbf{X}\Psi^* \in \mathbb{C}^{n \times m}$, $m \geq s$, the number of objects. Set $\mathbf{Z} = \mathbf{X}\Psi^* \in \mathbb{C}^{s \times m}$ such that $\mathbf{Y} = \Phi\mathbf{Z}$ where \mathbf{Z} is assumed to have rank s .

We shall treat \mathbf{Z} as the new object matrix and consider perturbed data matrices of the form

$$\mathbf{Y}^\varepsilon = \Phi\mathbf{Z} + \mathbf{E}. \quad (22)$$

Note that the locations of objects represented by \mathbf{Z} are identical to those represented by $\mathbf{X} = \text{diag}(\xi_j)$.

Set

$$\mathcal{Y}^\varepsilon = \mathbf{Y}^\varepsilon \mathbf{Y}^{\varepsilon*} = \mathcal{Y} + \mathcal{E} \quad (23)$$

where

$$\mathcal{Y} = \Phi\mathbf{Z}\mathbf{Z}^*\Phi^* \in \mathbb{C}^{n \times n} \quad (24)$$

$$\mathcal{E} = \mathbf{E}\mathbf{Z}^*\Phi^* + \Phi\mathbf{Z}\mathbf{E}^* + \mathbf{E}\mathbf{E}^* \in \mathbb{C}^{n \times n} \quad (25)$$

are both self-adjoint. Note that the range of \mathcal{Y} is the same as the range of Φ and under the assumption of (15) equals the span of $\{\phi_{\mathbf{r}} : \mathbf{r} \in S\}$.

Let $\{\mathbf{v}_j : j = 1, \dots, s\}$ and $\{\mathbf{v}_j : j = s+1, \dots, n\}$, respectively, be the set of orthonormal bases for the range and null space of \mathcal{Y} . Let $\mathbf{Q}_1 \in \mathbb{C}^{n \times s}$ and $\mathbf{Q}_2 \in \mathbb{C}^{n \times (n-s)}$, respectively, be the matrices whose columns are exactly $\{\mathbf{v}_j : j = 1, \dots, s\}$ and $\{\mathbf{v}_j : j = s+1, \dots, n\}$. Let $\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2] \in \mathbb{C}^{n \times n}$.

Let $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ be the singular values of \mathcal{Y} . Denote the *smallest nonzero* singular value of \mathcal{Y} by σ_{\min} and set $\sigma_{\max} = \sigma_1$. If \mathbf{Y} has rank s , then $\sigma_{\min} = \sigma_s$. We partition $\mathbf{Q}^*\mathcal{E}\mathbf{Q}$ as follows:

$$\mathbf{Q}^*\mathcal{E}\mathbf{Q} = \begin{bmatrix} \mathcal{E}_{11} & \mathcal{E}_{12} \\ \mathcal{E}_{21} & \mathcal{E}_{22} \end{bmatrix} \quad (26)$$

where $\mathcal{E}_{11} \in \mathbb{C}^{s \times s}$, $\mathcal{E}_{12} \in \mathbb{C}^{s \times (n-s)}$, $\mathcal{E}_{21} \in \mathbb{C}^{(n-s) \times s}$, $\mathcal{E}_{22} \in \mathbb{C}^{(n-s) \times (n-s)}$.

The following is a slight recasting of a general result of matrix perturbation theory [26].

Proposition 2 (Theorem 2.7, chapter V [26]). *If*

$$\frac{\sqrt{\|\mathcal{E}_{12}\|_2\|\mathcal{E}_{21}\|_2}}{\sigma_{\min} - \|\mathcal{E}_{11}\|_2 - \|\mathcal{E}_{22}\|_2} < \frac{1}{2} \quad (27)$$

then $\mathbf{F} \in \mathbb{C}^{(n-s) \times s}$ exists with

$$\|\mathbf{F}\|_2 \leq \frac{2\|\mathcal{E}_{21}\|_2}{\sigma_{\min} - \|\mathcal{E}_{11}\|_2 - \|\mathcal{E}_{22}\|_2} \quad (28)$$

such that the columns of

$$\mathbf{Q}_1^\varepsilon = (\mathbf{Q}_1 + \mathbf{Q}_2\mathbf{F})(\mathbf{I} + \mathbf{F}^*\mathbf{F})^{-1/2} \quad (29)$$

$$\mathbf{Q}_2^\varepsilon = (\mathbf{Q}_2 - \mathbf{Q}_1\mathbf{F}^*)(\mathbf{I} + \mathbf{F}\mathbf{F}^*)^{-1/2} \quad (30)$$

are, respectively, orthonormal bases for invariant subspaces of \mathcal{Y}^ε .

The representation of \mathcal{Y}^ε with respect to $\mathbf{Q}_1^\varepsilon, \mathbf{Q}_2^\varepsilon$ is, respectively,

$$\Sigma_1^\varepsilon = (\mathbf{I} + \mathbf{F}^*\mathbf{F})^{1/2} [\Sigma_1 + \mathcal{E}_{11} + \mathcal{E}_{12}\mathbf{F}] (\mathbf{I} + \mathbf{F}^*\mathbf{F})^{-1/2} \quad (31)$$

$$\Sigma_2^\varepsilon = (\mathbf{I} + \mathbf{F}\mathbf{F}^*)^{-1/2} [\mathcal{E}_{22} - \mathbf{F}\mathcal{E}_{12}] (\mathbf{I} + \mathbf{F}\mathbf{F}^*)^{1/2} \quad (32)$$

where $\Sigma_1 = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{\min})$.

Corollary 1. Let $\rho_* \in (1/5, 1/4)$ be the only real root of the cubic polynomial $p(\rho) = 1 - 8\rho + 20\rho^2 - 20\rho^3$ and suppose

$$\frac{\|\mathcal{E}\|_2}{\sigma_{\min}} < \rho_*. \quad (33)$$

Then $\text{Ran}(\mathbf{Q}_1^\varepsilon)$ is the singular subspace associated with the s largest singular values of \mathbf{Y}^ε and $\text{Ran}(\mathbf{Q}_2^\varepsilon)$ the singular subspace associated with the rest of the singular values.

Proof. It suffices to show that under (33) the smallest singular value of Σ_1 is larger than the largest singular value of Σ_2 .

Since $\rho_* < 1/4$, condition (33) implies that

$$\frac{\|\mathcal{E}\|_2}{\sigma_{\min} - 2\|\mathcal{E}\|_2} < \frac{1}{2}$$

which in turn implies (27). Note that under this condition $\|\mathbf{F}\|_2 \leq 1$. By proposition 2 we have

$$\begin{aligned} \|\Sigma_2^\varepsilon\|_2 &\leq \|\mathbf{I} + \mathbf{F}\mathbf{F}^*\|_2^{1/2} \|\mathcal{E}_{22} - \mathbf{F}\mathcal{E}_{12}\|_2 \\ &\leq \left(1 + \frac{4\|\mathcal{E}\|_2^2}{(\sigma_{\min} - 2\|\mathcal{E}\|_2)^2}\right)^{1/2} \left(\|\mathcal{E}\|_2 + \frac{2\|\mathcal{E}\|_2^2}{\sigma_{\min} - 2\|\mathcal{E}\|_2}\right). \end{aligned} \quad (34)$$

On the other hand,

$$\begin{aligned} \min_{\|\mathbf{e}\|_2=1} \|\Sigma_1^\varepsilon \mathbf{e}\|_2 &= \min_{\|\mathbf{e}\|_2=1} \|(\Sigma_1 + \mathcal{E}_{11} + \mathcal{E}_{12}\mathbf{F})\mathbf{e}\|_2 \|\mathbf{I} + \mathbf{F}^*\mathbf{F}\|_2^{-1/2} \\ &\geq \left(\sigma_{\min} - \|\mathcal{E}\|_2 - \frac{2\|\mathcal{E}\|_2^2}{\sigma_{\min} - 2\|\mathcal{E}\|_2}\right) \left(1 + \frac{4\|\mathcal{E}\|_2^2}{(\sigma_{\min} - 2\|\mathcal{E}\|_2)^2}\right)^{-1/2}. \end{aligned} \quad (35)$$

Let

$$\rho = \frac{\|\mathcal{E}\|_2}{\sigma_{\min}}.$$

Imposing that the right-hand side of (35) is greater than that of (34) leads to the inequality

$$\begin{aligned} p(\rho) &= (1 - 2\rho)^3 - 2\rho(1 - 2\rho)^2 - 4\rho^3 \\ &= 1 - 8\rho + 20\rho^2 - 20\rho^3 > 0 \end{aligned}$$

which holds for $\rho < \rho_*$, the only real root of $p(x)$. It is readily verified that $\rho_* \in (1/5, 1/4)$. \square

In view of corollary 1, it is natural to call $\text{Ran}(\mathbf{Q}_2^\varepsilon)$ the *noise* subspace and $\text{Ran}(\mathbf{Q}_1^\varepsilon)$ the *signal* (or *object*) subspace.

Let \mathcal{P}^ε be the orthogonal projection onto the noise subspace and define the MUSIC imaging function for the noisy data

$$J^\varepsilon(\mathbf{r}) = \frac{1}{|\mathcal{P}^\varepsilon \phi_{\mathbf{r}}|^2}. \quad (36)$$

We are ready to state the first main result of the paper.

Theorem 1. Let $\mathcal{Y}^\varepsilon = \mathcal{Y} + \mathcal{E}$ where \mathcal{Y} and \mathcal{E} are given by (24)–(25).

Suppose $\Gamma_S > 0$. Then the condition

$$\frac{\|\mathcal{E}\|_2}{\sigma_{\min}} < \Delta = \frac{1}{2} - \frac{1}{2\sqrt{\sqrt{2}\Gamma_S + 1}} \quad (37)$$

implies that the s highest peaks of $J^\varepsilon(\mathbf{r})$ coincide with the true locations of objects. Indeed, the object locations can be identified by the thresholding rule:

$$\{\mathbf{r} \in \mathcal{K} : J^\varepsilon(\mathbf{r}) \geq 2\Gamma_S^{-2}\}. \quad (38)$$

Proof. Let $\{\mathbf{v}_j + \delta\mathbf{v}_j : j = s + 1, \dots, n\}$ be the columns of \mathbf{Q}_2^ε . Clearly, $\delta\mathbf{v}_j$ is the $(j - s)$ th column of $\mathbf{Q}_2^\varepsilon - \mathbf{Q}_2$. Now we have

$$\begin{aligned} \|\mathbf{Q}_2^\varepsilon - \mathbf{Q}_2\|_2 &\leq \|\mathbf{Q}_1 \mathbf{F}^* (\mathbf{I} + \mathbf{F} \mathbf{F}^*)^{-1/2}\|_2 + \|\mathbf{Q}_2 (\mathbf{I} + \mathbf{F} \mathbf{F}^*)^{-1/2} - \mathbf{Q}_2\|_2 \\ &\leq \|\mathbf{Q}_1 \mathbf{F}^*\|_2 + \|(\mathbf{I} + \mathbf{F} \mathbf{F}^*)^{-1/2} - \mathbf{I}\|_2 \end{aligned}$$

whose first term is bounded by $\|\mathbf{F}\|_2$ and whose second term is bounded by

$$\begin{aligned} \|(\mathbf{I} + \mathbf{F} \mathbf{F}^*)^{-1/2} - \mathbf{I}\|_2 &= \|(\mathbf{I} + \mathbf{F} \mathbf{F}^*)^{-1/2} (\mathbf{I} + (\mathbf{I} + \mathbf{F} \mathbf{F}^*)^{1/2})^{-1} \mathbf{F} \mathbf{F}^*\|_2 \\ &\leq \frac{1}{2} \|\mathbf{F}\|_2^2. \end{aligned}$$

Hence,

$$\|\delta\mathbf{v}_j\|_2 \leq \|\mathbf{Q}_2^\varepsilon - \mathbf{Q}_2\|_2 \leq \|\mathbf{F}\|_2 + \frac{1}{2} \|\mathbf{F}\|_2^2 \leq \frac{2\rho(1 - \rho)}{(1 - 2\rho)^2}, \quad j = s + 1, \dots, n, \quad (39)$$

where $\rho = \|\mathcal{E}\|_2 / \sigma_{\min}$.

By corollary 1 $\{\mathbf{v}_j + \delta\mathbf{v}_j : j = s + 1, \dots, n\}$ is the set of singular vectors associated with the $n - s$ smallest singular values of \mathbf{Y} . By definition,

$$\begin{aligned} \|\mathcal{P}^\varepsilon \phi_{\mathbf{r}}\|_2^2 &= \sum_{k=s+1}^n |(\mathbf{v}_k^* + \delta\mathbf{v}_k^*) \phi_{\mathbf{r}}|^2 \\ &= \sum_{k=s+1}^n |\mathbf{v}_k^* \phi_{\mathbf{r}}|^2 + 2 \sum_{k=s+1}^n \Re[\mathbf{v}_k^* \phi_{\mathbf{r}} \delta\mathbf{v}_k^*] + \|(\mathbf{Q}_2^\varepsilon - \mathbf{Q}_2)^* \phi_{\mathbf{r}}\|_2^2. \end{aligned} \quad (40)$$

By assumption the first two terms on the right-hand side of (40) vanish if and only if $\mathbf{r} \in S$. By (39) the third term is bounded by

$$\|(\mathbf{Q}_2^\varepsilon - \mathbf{Q}_2)^* \phi_{\mathbf{r}}\|_2^2 \leq \frac{4\rho^2(1 - \rho)^2}{(1 - 2\rho)^4}.$$

For $\mathbf{r} \notin \mathcal{S}$,

$$\begin{aligned} \sum_{k=s+1}^n |(\mathbf{v}_k^* + \delta \mathbf{v}_k^*) \phi_{\mathbf{r}}|^2 &= \sum_{k=s+1}^n |\mathbf{v}_k^* \phi_{\mathbf{r}}|^2 - \|(\mathbf{Q}_2^\varepsilon - \mathbf{Q}_2)^* \phi_{\mathbf{r}}\|_2^2 \\ &\geq \Gamma_S^2 - \frac{4\rho^2(1-\rho)^2}{(1-2\rho)^4} \end{aligned}$$

by (39). Hence $J^\varepsilon(\mathbf{r})$ has the following behavior

$$J^\varepsilon(\mathbf{r}) \geq \frac{(1-2\rho)^4}{4\rho^2(1-\rho)^2}, \quad \mathbf{r} \in \mathcal{S} \tag{41}$$

$$J^\varepsilon(\mathbf{r}) \leq \left(\Gamma_S^2 - \frac{4\rho^2(1-\rho)^2}{(1-2\rho)^4} \right)^{-1}, \quad \mathbf{r} \in \mathcal{S}^c. \tag{42}$$

Setting

$$\frac{(1-2\rho)^4}{4\rho^2(1-\rho)^2} > \left(\Gamma_S^2 - \frac{4\rho^2(1-\rho)^2}{(1-2\rho)^4} \right)^{-1}$$

we obtain the inequality

$$\rho^2 - \rho + \frac{\Gamma_S}{4\Gamma_S + 2\sqrt{2}} > 0 \tag{43}$$

whose solution is (37). Note that $\Delta < 1/5 < \rho_*$, $\forall \Gamma_S \in [0, 1]$. □

Condition (37) would not be very useful unless $\|\mathcal{E}\|_2$ can be bounded from above and σ_{\min}, Γ_S can be bounded from below by other known or accessible quantities. This is what the compressed sensing techniques enable us to do.

3. Compressed sensing analysis

We now give a quantitative evaluation of MUSIC based on compressed sensing theory.

A fundamental notion in compressed sensing is the restrictive isometry property (RIP) due to Candès and Tao [5]. Precisely, let the sparsity s of a vector $Z \in \mathbb{C}^N$ be the number of nonzero components of Z and define the restricted isometry constants $\delta_s^- \in [0, 1]$, $\delta_s^+ \in [0, \infty)$ to be the *smallest* nonnegative numbers such that the inequality

$$(1 - \delta_s^-) \|Z\|_2^2 \leq \|\tilde{\Phi} Z\|_2^2 \leq (1 + \delta_s^+) \|Z\|_2^2 \tag{44}$$

holds for all $Z \in \mathbb{C}^N$ of sparsity at most s .

Roughly speaking this means that $\tilde{\Phi}$ acts like a near isometry, up to a scaling, when restricted to s -sparse vectors. In particular, if $\delta_{s+1}^- < 1$ then any $s+1$ columns of $\tilde{\Phi}$ are linearly independent which implies the characterization (15).

More generally, let us extend the notion of the restricted isometry constants to ones δ_S^\pm associated with a particular set \mathcal{S} , namely the smallest nonnegative numbers satisfying

$$(1 - \delta_S^-) \|Z\|_2^2 \leq \|\tilde{\Phi} Z\|_2^2 \leq (1 + \delta_S^+) \|Z\|_2^2 \tag{45}$$

for all $Z \in \mathbb{C}^N$ supported on the set \mathcal{S} . This will become important later when we analyze the case of an arbitrarily refined grid (section 7). Clearly,

$$\delta_s^\pm = \max_{|\mathcal{S}|=s} \delta_S^\pm. \tag{46}$$

Then (15) is equivalent to $\delta_{\mathcal{S}'}^- < 1$ for all \mathcal{S}' which is the union of \mathcal{S} and another point $\mathbf{r} \in \mathcal{S}^c$.

First, let us estimate the magnitude of the error term \mathcal{E} in terms of \mathbf{E} as follows.

Lemma 1. Suppose (45) holds for Φ . Then

$$\|\mathcal{E}\|_2 \leq \|\mathbf{E}\|_2^2 + 2\zeta_{\max}\sqrt{1 + \delta_S^+}\|\mathbf{E}\|_2 \leq \|\mathbf{E}\|_2^2 + 2\zeta_{\max}\sqrt{1 + \delta_S^+}\|\mathbf{E}\|_2 \quad (47)$$

where $\zeta_{\max} = \|\mathbf{Z}\|_2$ is the largest singular value of \mathbf{Z} .

For the case of $\mathbf{Z} = \mathbf{X}\Psi^*$ with Ψ satisfying (45), we have

$$\|\mathcal{E}\|_2 \leq \|\mathbf{E}\|_2^2 + 2\xi_{\max}(1 + \delta_S^+)\|\mathbf{E}\|_2 \leq \|\mathbf{E}\|_2^2 + 2\xi_{\max}(1 + \delta_S^+)\|\mathbf{E}\|_2 \quad (48)$$

where $\xi_{\max} = \max_i |\xi_i|$.

Proof. First we have

$$\|\mathcal{E}\|_2 \leq \|\mathbf{E}\|_2^2 + 2\|\Phi\mathbf{Z}\mathbf{E}^*\|_2.$$

The RIP (45) then implies that

$$\|\Phi\mathbf{Z}\mathbf{E}^*\|_2^2 \leq (1 + \delta_S^+)\|\mathbf{Z}\mathbf{E}^*\|_2^2$$

and thus

$$\begin{aligned} \|\mathcal{E}\|_2 &\leq \|\mathbf{E}\|_2^2 + 2\sqrt{1 + \delta_S^+}\|\mathbf{Z}\mathbf{E}^*\|_2 \\ &\leq \|\mathbf{E}\|_2^2 + 2\sqrt{1 + \delta_S^+}\|\mathbf{Z}\|_2\|\mathbf{E}^*\|_2. \end{aligned}$$

In the case of scattering objects $\mathbf{Z} = \mathbf{X}\Psi^*$

$$\|\mathbf{Z}\mathbf{E}^*\|_2 = \|\mathbf{E}\Psi\mathbf{X}^*\|_2 \leq \|\mathbf{E}\|_2\|\Psi\mathbf{X}^*\|_2 \leq \|\mathbf{E}\|_2\xi_{\max}\sqrt{1 + \delta_S^+}$$

provided that Ψ also satisfies the RIP (45). In this case,

$$\|\mathcal{E}\|_2 \leq \|\mathbf{E}\|_2^2 + 2\xi_{\max}(1 + \delta_S^+)\|\mathbf{E}\|_2 \quad (49)$$

and hence (48). \square

Lemma 2. The minimum nonzero singular value σ_{\min} of \mathcal{Y} obeys the lower bound

$$\sigma_{\min} \geq (1 - \delta_S^-)\zeta_{\min}^2 \geq (1 - \delta_S^-)\xi_{\min}^2 \quad (50)$$

where

$$\zeta_{\min} = \min_{\mathbf{e} \in \mathcal{C}^s} \frac{\|\mathbf{Z}^*\mathbf{e}\|_2}{\|\mathbf{e}\|_2}. \quad (51)$$

For the case of scattering objects $\mathbf{Z} = \mathbf{X}\Psi^*$ with Ψ satisfying (45), we have

$$\sigma_{\min} \geq (1 - \delta_S^-)^2\xi_{\min}^2 \geq (1 - \delta_S^-)^2\xi_{\min}^2. \quad (52)$$

Proof.

Using the max–min theorem [19]

$$\sigma_s(\mathcal{Y}) = \max_{\dim \mathcal{H}=s} \min_{\mathbf{e} \in \mathcal{H}} \frac{\|\mathcal{Y}\mathbf{e}\|_2}{\|\mathbf{e}\|_2} \quad (53)$$

with $\mathcal{H} = \text{Ran}(\Phi)$, we obtain

$$\sigma_{\min}(\mathcal{Y}) \geq \min_{\substack{\mathbf{e} \in \text{Ran}(\Phi) \\ \|\mathbf{e}\|_2=1}} \|\Phi\mathbf{Z}\mathbf{Z}^*\Phi^*\mathbf{e}\|_2. \quad (54)$$

Let $\{\mathbf{u}_j : j = 1, \dots, s\}$ be the eigenvectors of $\Phi\mathbf{Z}\mathbf{Z}^*\Phi^*$ associated with the nonzero eigenvalues $\{\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_s^2\}$ and form the orthonormal basis of \mathcal{H} . Write $\mathbf{e} = \sum_{j=1}^s e_j \mathbf{u}_j$ where $\sum_j |e_j|^2 = 1$. We have

$$\Phi\mathbf{Z}\mathbf{Z}^*\Phi^*\mathbf{e} = \sum_{j=1}^s \lambda_j^2 e_j \mathbf{u}_j$$

and thus

$$\|\Phi^* \mathbf{e}\|_2^2 = \sum_{j=1}^s \lambda_j^2 |e_j|^2 \geq \lambda_s^2. \tag{55}$$

It follows from (54), (45) and (55) that

$$\sigma_{\min} \geq \sqrt{1 - \delta_S^-} \|\mathbf{Z}\mathbf{Z}^* \Phi^* \mathbf{e}\|_2 \geq \sqrt{1 - \delta_S^-} \zeta_{\min}^2 \lambda_s \tag{56}$$

by (51). On the other hand, λ_s^2 is exactly the smallest eigenvalue of $\Phi^* \Phi \in \mathbb{C}^{s \times s}$ and hence by (45) is bounded from below by $1 - \delta_S^-$. Using this observation in (56) we obtain (50).

In the case of scattering objects we can bound ζ_{\min}^2 as

$$\begin{aligned} \zeta_{\min}^2 &\geq (1 - \delta_S^-) \xi_{\min}^2 \\ \zeta_{\max}^2 &\leq (1 + \delta_S^+) \xi_{\max}^2 \end{aligned}$$

by using (45) with Ψ and hence the result (52). □

Next we derive a lower bound for Γ_S in terms of RIC.

Lemma 3. Fix $S, |S| = s$. Then, the lower bound is valid

$$\Gamma_S \geq 1 - \max_{\substack{r \in S^c \\ S' = S \cup \{r\}}} \frac{\delta_{S'}^-(1 + \delta_S^+)}{2 + \delta_S^+ - \delta_{S'}^-} \geq \gamma_s \tag{57}$$

$$\gamma_s = 1 - \frac{\delta_{s+1}^-(1 + \delta_s^+)}{2 + \delta_s^+ - \delta_{s+1}^-}. \tag{58}$$

Proof. Without loss of generality, suppose $S = \{1, 2, \dots, s\}$ and consider $\phi_r = \tilde{\Phi}_{s+1}$. Let $S' = S \cup \{s + 1\}$. Our subsequent analysis is independent of these choices modulo inconsequential notational change.

Denote $\mathcal{P}\phi_r = \phi'$ and write the orthogonal decomposition

$$\phi_r = \phi' + \sum_{j=1}^s c_j \tilde{\Phi}_j. \tag{59}$$

Hence we can express ϕ' as

$$\phi' = \tilde{\Phi}_{s+1} - \sum_{j=1}^s c_j \tilde{\Phi}_j = \tilde{\Phi} \mathbf{Z}, \quad \mathbf{Z} = (-c_1, -c_2, \dots, -c_s, 1, 0, \dots, 0)^T \in \mathbb{C}^N.$$

Using (45) for sparsity S' we obtain a lower bound for $\|\phi'\|_2$:

$$(1 - \delta_{S'}^-) \left(1 + \sum_{j=1}^s |c_j|^2 \right) \leq \|\phi'\|_2^2. \tag{60}$$

On the other hand, we have by the Pythagorean theorem that

$$\|\phi_r\|_2^2 = \|\phi'\|_2^2 + \left\| \sum_{j=1}^s c_j \tilde{\Phi}_j \right\|_2^2. \tag{61}$$

Applying (45) for sparsity S to the second term on the right-hand side of (61) we obtain

$$\|\phi_r\|_2^2 - \|\phi'\|_2^2 = \left\| \sum_{j=1}^s c_j \tilde{\Phi}_j \right\|_2^2 \leq (1 + \delta_S^+) \sum_{j=1}^s |c_j|^2$$

and hence

$$\sum_{j=1}^s |c_j|^2 \geq (1 + \delta_S^+)^{-1} (\|\phi_{\mathbf{r}}\|_2^2 - \|\phi'\|_2^2). \quad (62)$$

Combining (62) and (60) we obtain

$$\|\phi'\|_2^2 \geq (1 - \delta_{S'}^-) \left(1 + \frac{1 - \|\phi'\|_2^2}{1 + \delta_S^+}\right)$$

which can be solved to yield

$$\|\phi'\|_2^2 \geq 1 - \frac{\delta_{S'}^-(1 + \delta_S^+)}{2 + \delta_S^+ - \delta_{S'}^-}. \quad (63)$$

Minimizing (63) over $\mathbf{r} \in \mathcal{S}^c$ we obtain the first inequality in (57).

The second inequality (57) follows from (46) and the observation (by differentiation) that the quantity

$$\frac{\delta_{S'}^-(1 + \delta_S^+)}{2 + \delta_S^+ - \delta_{S'}^-}$$

is an increasing function of $\delta_{S'}^- \in [0, 1]$ and $\delta_S^+ \in [0, \infty)$ separately. \square

Combining the preceding results we have the following stability criterion for exact recovery by MUSIC.

Theorem 2. Suppose $\delta_{s+1}^- < 1$ (implying (15)) and $\|\mathbf{E}\|_2 = \varepsilon$.

If the noise-to-object ratio (NOR) satisfies

$$\frac{\varepsilon}{\xi_{\min}} < \sqrt{(1 + \delta_s^+) \frac{\xi_{\max}^2}{\xi_{\min}^2} + (1 - \delta_s^-) \tilde{\Delta}} - \frac{\xi_{\max}}{\xi_{\min}} \sqrt{1 + \delta_s^+} \quad (64)$$

where

$$\tilde{\Delta} = \frac{1}{2} - \frac{1}{2} \frac{1}{\sqrt{\sqrt{2}\gamma_s + 1}}, \quad \gamma_s \text{ given by (58)}, \quad (65)$$

then the object support \mathcal{S} can be identified by the thresholding rule

$$\{\mathbf{r} \in \mathcal{K} : J^\varepsilon(\mathbf{r}) \geq 2\gamma_s^{-2}\}. \quad (66)$$

In the case of scattering objects $\mathbf{Z} = \mathbf{X}\Psi^*$ with Ψ^* satisfying the RIP (44) the thresholding rule (66) holds under the following bound on the noise-to-scatterer ratio (NSR)

$$\frac{\varepsilon}{\xi_{\min}} < \sqrt{(1 + \delta_s^+)^2 \frac{\xi_{\max}^2}{\xi_{\min}^2} + (1 - \delta_s^-)^2 \tilde{\Delta}} - (1 + \delta_s^+) \frac{\xi_{\max}}{\xi_{\min}} \quad (67)$$

where ξ_{\max}/ξ_{\min} is the dynamic range of scatterers.

Proof. By (47) and (50)

$$\tilde{\rho}^2 + 2\tilde{\rho} \frac{\xi_{\max}}{\xi_{\min}} \sqrt{1 + \delta_s^+} < \tilde{\Delta}, \quad \tilde{\rho} = \frac{\varepsilon}{\xi_{\min}}, \quad (68)$$

implies (37) in theorem 1. The sufficiency of (64) now follows from solving the quadratic inequality (68) for ρ .

The derivation of the thresholding rule (66) under the stronger condition (64) is exactly the same as that of (38). Alternatively, we can use (41) and (42) to verify the validity of the

thresholding rule (66) as follows. Under condition (37) (and hence (43)) it is straightforward to check that $2\gamma_s^{-2} \geq 2\Gamma_s^{-2}$ and is greater than the right-hand side of (42). On the other hand, (37) and lemma 3 imply that $2\gamma_s^{-2}$ is smaller than the right-hand side of (41).

The proof for the case of scattering objects is exactly the same as above. \square

Remark 3. The right-hand side of (67) decreases as the ratio

$$\frac{(1 - \delta_s^-)^2 \tilde{\Delta}}{2(1 + \delta_s^+) \xi_{\max} / \xi_{\min}} \quad (69)$$

decreases. In the under-resolved case (section 8), δ_s^- is close to 1, making the ratio (69) a small number. For a noise-to-scatterer ratio smaller than (69) the s scatterers can be perfectly localized by the MUSIC algorithm with thresholding. This super-resolution effect is demonstrated in the numerical simulations of section 8.

We state the MUSIC algorithm with the thresholding rule (66) for the noisy case as follows.

Theorem 2 provides a performance guarantee for this algorithm under condition (64). The thresholding rule (66) can be made more practical and useful as (76) in corollary 2.

A simple upper bound for the RIC can be given in terms of the notion of coherence parameter $\mu(\tilde{\Phi})$ defined as

$$\mu(\tilde{\Phi}) = \max_{i \neq j} \frac{|\sum_l \tilde{\Phi}_{li} \tilde{\Phi}_{lj}^*|}{\sqrt{\sum_l |\tilde{\Phi}_{li}|^2 \sum_l |\tilde{\Phi}_{lj}|^2}}.$$

Namely, $\mu(\tilde{\Phi})$ is the maximum of cosines of angles between any two columns.

The proof of the following well-known result is elementary and instructive.

Proposition 3. For any $r \in \mathbb{N}$, we have

$$\delta_r^\pm \leq \mu(\tilde{\Phi})(r - 1).$$

Proof. Calculating the quantity $\|\tilde{\Phi}Z\|_2^2 - \|Z\|_2^2$, we have

$$|\|\tilde{\Phi}Z\|_2^2 - \|Z\|_2^2| = \left| \sum_{i \neq j} \tilde{\Phi}_i^* \tilde{\Phi}_j Z_i^* Z_j \right| \leq \mu(\tilde{\Phi}) \sum_{i \neq j} |Z_i^* Z_j|.$$

Using the quadratic inequality $2ab \leq a^2 + b^2$ we obtain

$$\sum_{i \neq j} |Z_i^* Z_j| \leq \frac{1}{2} \sum_{i \neq j} (|Z_i|^2 + |Z_j|^2) \leq \sum_{\substack{i \neq j \\ Z_j \neq 0}} |Z_i|^2 \leq (r - 1) \|Z\|_2^2.$$

Therefore, (44) is satisfied with $\delta_r^\pm \leq \mu(\tilde{\Phi})(r - 1)$. \square

Remark 4. For $s = 2$ it follows from proposition 3 that

$$\delta_2^\pm \leq \mu(\tilde{\Phi}).$$

Since μ is almost surely less than unity for randomly selected sampling directions, the MUSIC algorithm will find the true location of the object in the absence of noise, if there is only one object.

The coherence bound for the most general setting of random sampling directions is this.

Proposition 4 [12]. *Suppose any two points in \mathcal{K} are separated by at least $\ell > 0$. Let $\hat{\mathbf{s}}_k, k = 1, \dots, n$, be independently drawn from the distribution f^s on the $(d-1)$ -dimensional sphere independently and identically. Suppose*

$$N \leq \frac{\alpha}{8} e^{K^2/2} \quad (70)$$

for any positive constant α, K . Then $\tilde{\Phi}$ satisfies the coherence bound

$$\mu(\tilde{\Phi}) < \chi^s + \frac{\sqrt{2}K}{\sqrt{n}}$$

with probability greater than $(1-\alpha)^2$ where χ^s satisfies the bound

$$\chi^s \leq c_t(1+\omega\ell)^{-1/2} \|f^s\|_{t,\infty}, \quad d=2 \quad (71)$$

$$\chi^s \leq c_1(1+\omega\ell)^{-1} \|f^s\|_{1,\infty}, \quad d=3 \quad (72)$$

where $\|\cdot\|_{t,\infty}$ is the Hölder norm of order $t > 1/2$ and the constant c_t depends only on t .

Remark 5. Replacing $\tilde{\Phi}, \hat{\mathbf{s}}_k$ and f^s in proposition 4 by $\tilde{\Psi}, \hat{\mathbf{d}}_k$ and f^i , respectively, we have the same conclusion about $\tilde{\Psi}$.

Remark 6. Constraint (70) on the number of search points in the computation grid \mathcal{K} is relatively weak and allows for essentially any grid. However, to have a small χ^s the grid spacing ℓ cannot be small compared to the wavelength.

Suppose $\omega\ell \geq C^2n$ for $d=2$ or $\omega\ell \geq C\sqrt{n}$ for $d=3$ where

$$C \geq \frac{c_1}{\sqrt{2}K} \max \{ \|f^s\|_{1,\infty}, \|f^i\|_{1,\infty} \}. \quad (73)$$

Then, according to propositions 4, 3 and remark 5, with high probability, for continuously differentiable distributions f^s, f^i we have

$$\delta_s^\pm \leq \delta_{s+1}^\pm \leq 2\sqrt{2}Ks/\sqrt{n}. \quad (74)$$

Clearly the right-hand side of (73) achieves the minimum when f^s and f^i are the uniform distribution over the unit circle for $d=2$ or the unit sphere for $d=3$.

On the other hand when the aperture for either illumination or reception is limited, the right-hand side of (73) increases and the condition becomes more stringent. This occurs, for example, when the objects are embedded in a half-space with emitters and receivers set above it. The aperture limitation typically causes the performance of the MUSIC algorithm to deteriorate (cf [1] and [18]). However, the deterioration in performance does not get worse than what is implied by (73) and may be restored by increasing the frequency at least in the case of point objects.

Theorem 2 then implies the following.

Corollary 2. *Suppose $\omega\ell \geq C^2n$ for $d=2$ or $\omega\ell \geq C\sqrt{n}$ for $d=3$ with C given by (73).*

Suppose that $\sqrt{n}/s \geq 4\sqrt{2}K$ (hence $\delta_s^\pm, \delta_{s+1}^\pm \leq 1/2$ by (74)) and that the NSR obeys

$$\frac{\varepsilon}{\xi_{\min}} < \left(\sqrt{\frac{9}{4} \frac{\xi_{\max}^2}{\xi_{\min}^2} + \frac{\tilde{\Delta}}{4}} - \frac{3}{2} \frac{\xi_{\max}}{\xi_{\min}} \right) \quad (75)$$

where $\tilde{\Delta}$ is given in (65). Then under the assumptions of proposition 4 the MUSIC algorithm with the thresholding rule

$$\left\{ \mathbf{r} \in \mathcal{K} : J^\varepsilon(\mathbf{r}) \geq \frac{128}{25} \right\} \quad (76)$$

recovers exactly the locations of s scatterers with probability at least $(1-\alpha)^2$.

The value $128/25$ is arrived from the fact that

$$\max_{\delta_s^+, \delta_{s+1}^- < 1/2} \frac{\delta_{s+1}^-(1 + \delta_s^+)}{2 + \delta_s^+ - \delta_{s+1}^-} \leq 3/8.$$

4. Planar objects: optimal recovery

Let us consider the favorable imaging geometry where all the scatterers lie on the transverse plane $z = 0$. Furthermore, we consider the idealized situation where the locations of the scatterers are a subset \mathcal{S} of a finite square lattice \mathcal{K} of spacing ℓ :

$$\mathcal{K} = \{\mathbf{r}_j : j = 1, \dots, N\} = \{(p_1\ell, p_2\ell, 0) : p_1, p_2 = 1, \dots, \sqrt{N}\}, \quad j = (p_1 - 1)\sqrt{N} + p_2. \quad (77)$$

Hence, the total number of grid points N is a perfect square.

Suppose we choose the frequency such that

$$\omega\ell = \sqrt{2}\pi. \quad (78)$$

Let $\mathbf{a}_k = (\xi_k, \eta_k)$, $k = 1, \dots, n$, be independently and uniformly distributed random variables in $[-1, 1]^2$ and set

$$\hat{\mathbf{s}}_k = \frac{1}{\sqrt{2}}(\mathbf{a}_k, \sqrt{2 - |\mathbf{a}_k|^2}). \quad (79)$$

Let the incident directions $\hat{\mathbf{d}}_l$, $l = 1, \dots, m$, be selected the same way but independently from $\hat{\mathbf{s}}_k$, $k = 1, \dots, n$. It can be proved that with $m \geq s$ the corresponding sensing matrix Ψ has rank s with probability 1.

With (78)–(79) and $j = (p_1 - 1)\sqrt{N} + p_2$ the scattering amplitude (11) for linear extended objects yields the following extended sensing matrix:

$$\tilde{\Phi}_{k,j} = e^{-\pi i \mathbf{a}_k \cdot \mathbf{p}} \in \mathbb{C}^{n \times N}. \quad (80)$$

Matrix (80) is often referred to as the random partial Fourier matrix in compressed sensing theory.

The following is a standard result about the random partial Fourier matrix [24].

Proposition 5 [24]. *Suppose \mathbf{a}_j , $j = 1, \dots, n$, are independently and uniformly distributed in $[-1, 1]^d$, $d \geq 1$. If*

$$\frac{n}{\ln n} \geq C \delta_*^{-2} s \ln^2 s \ln N \ln \frac{1}{\gamma} \quad (81)$$

for $\gamma \in (0, 1)$ and some absolute constant C , then with probability at least $1 - \gamma$ the random partial Fourier matrix defined by (80) satisfies the RIC bound

$$\delta_s^\pm < \delta_*. \quad (82)$$

Remark 7. The result holds true for sampling points \mathbf{a}_j which are i.i.d. uniform random variables (r.v.s) in the discrete set

$$\left\{ \left(\frac{k_1}{N^{1/d}}, \frac{k_2}{N^{1/d}}, \dots, \frac{k_d}{N^{1/d}} \right) : k_1, k_2, \dots, k_d \in \{-N^{1/d}, \dots, N^{1/d} - 1\} \right\} \quad (83)$$

instead of $[-1, 1]^d$ where $N^{1/d}$ is assumed to be an integer.

Assume for simplicity the plane wave incidence as before. Choosing $\delta_* = 1/2$ for sparsity $s + 1$ in proposition 5 and using theorem 2 we obtain the following result.

Corollary 3. Suppose that (78) is true and that $\hat{\mathbf{s}}_k = \hat{\mathbf{d}}_k$, $k = 1, \dots, n$, with

$$\frac{n}{\ln n} \geq 4C(s+1) \ln^2(s+1) \ln N \ln \frac{1}{\gamma}. \quad (84)$$

If the NSR obeys (75) the MUSIC algorithm with the thresholding rule (76) recovers exactly the locations of s scatterers with probability at least $1 - \gamma$.

4.1. Paraxial regime

Here we would like to extend the scattering problem to the paraxial regime for the preceding setup where, instead of n sampling directions, n point sensors located on the transverse plane $z = z_0$ measure the scattered field.

We shall make the paraxial approximation for Green's function between the object plane $z = 0$ and the sensor plane $z = z_0$:

$$G(\mathbf{s}, \mathbf{r}) = \frac{e^{i\omega z_0}}{4\pi z_0} e^{i\omega(x^2+y^2)/(2z_0)} e^{-i\omega x\xi/z_0} e^{-i\omega y\eta/z_0} e^{i\omega(\xi^2+\eta^2)/(2z_0)} \quad (85)$$

with $\mathbf{s} = (\xi, \eta, z_0)$, $\mathbf{r} = (x, y, 0)$. Denote

$$G_p(\mathbf{s}, \mathbf{r}) = e^{i\omega|x-\xi|^2/(2z_0)} e^{i\omega|y-\eta|^2/(2z_0)}.$$

Let $\mathbf{s}_k = (\xi_k, \eta_k, z_0)$, $k = 1, \dots, n$, be the locations of the transceivers. We have $\tilde{\Phi} = \tilde{\Psi}$ where the extended matrix $\tilde{\Phi}$ is given by

$$\tilde{\Phi}_{k,l} = (G(\mathbf{s}_1, \mathbf{r}_j), \dots, G(\mathbf{s}_n, \mathbf{r}_j))^T, \quad j = 1, \dots, N, \quad (86)$$

where $\mathbf{r}_j \in \mathcal{K}$ defined in (77). After proper normalization the extended sensing matrix $\tilde{\Phi}$ can be written as the product of three matrices:

$$\tilde{\Phi} = \mathbf{D}_1 \mathbf{A} \mathbf{D}_2 \quad (87)$$

where

$$\mathbf{D}_1 = \text{diag}(e^{i\omega(\xi_j^2+\eta_j^2)/(2z_0)}) \in \mathbb{C}^{n \times n}, \quad \mathbf{D}_2 = \text{diag}(e^{i\omega(x_j^2+y_j^2)/(2z_0)}) \in \mathbb{C}^{N \times N}$$

are unitary and

$$\mathbf{A} = \frac{1}{\sqrt{n}} [e^{-i\omega\xi_j x_l/z_0} e^{-i\omega\eta_j y_l/z_0}] \in \mathbb{C}^{n \times N}.$$

Now suppose that (ξ_j, η_j) , $j = 1, \dots, n$, are independently and uniformly distributed in $[-\frac{A}{2}, \frac{A}{2}]^2$ and write $(\xi_j, \eta_j) = \mathbf{a}_j \cdot A/2$ with $\mathbf{a}_j \in [-1, 1]^2$, $j = 1, \dots, n$. Write also $\mathbf{r}_l = (x_l, y_l, 0) = (\mathbf{p}l, 0)$ where $\mathbf{p} \in \mathbb{Z}^2$. Then with

$$\frac{A\ell}{\lambda z_0} = 1 \quad (88)$$

\mathbf{A} takes the form

$$\mathbf{A} = \frac{1}{\sqrt{n}} [e^{-\pi i \mathbf{a} \cdot \mathbf{p}}] \in \mathbb{C}^{n \times N} \quad (89)$$

which is exactly the random partial Fourier matrix given in (80). Here and below λ denotes the wavelength.

Since both \mathbf{D}_1 and \mathbf{D}_2 are unitary and diagonal, they leave both the ℓ^2 -norm and the sparsity of a vector unchanged. Therefore, proposition 5 and remark 7 are applicable to $\tilde{\Phi}$ given in (87).

Analogous to corollary 3 we have

Corollary 4. *Suppose that (88) is true and that there are n transceivers satisfying (84). If the NSR obeys (75), then the MUSIC algorithm with the thresholding rule (76) recovers exactly the locations of s scatterers with probability at least $1 - \gamma$.*

In the paraxial setting (88) in corollary 4 replaces condition (78) in corollary 3. Condition (88) is exactly the classical Rayleigh criterion for resolution which is, in this case, the grid spacing ℓ .

5. Comparison with basis pursuit

In the standard compressed sensing theory, one usually considers the following data model:

$$Y^\varepsilon = \tilde{\Phi}Z + E \in \mathbb{C}^n, \quad Z \in \mathbb{C}^N, \quad \|E\|_2 \leq \varepsilon \quad (90)$$

where the data and the object are vectors and employs the relaxed minimization principle called the basis pursuit denoising (BPDN) [2, 6]

$$\min_{Z' \in \mathbb{C}^N} \|Z'\|_1, \quad \text{s.t. } \|Y^\varepsilon - \tilde{\Phi}Z'\|_2 \leq \varepsilon \quad (91)$$

for reconstruction. The noiseless version $\varepsilon = 0$ of (91) is called the basis pursuit (BP). Note that BPDN uses only *one column* of the MUSIC model (22).

When Z is not s -sparse, consider the best s -sparse approximation $Z^{(s)}$ of Z . Clearly, $Z^{(s)} = Z$ if Z is s -sparse.

Denote the BPDN minimizer by \hat{Z} . When does \hat{Z} give a good approximation to the true Z ? Again, the RIP (44) gives a useful characterization [3].

Theorem 3. *Suppose the RICs of $\tilde{\Phi} \in \mathbb{C}^{n \times N}$ satisfy the inequality*

$$\frac{\sqrt{2}}{2} \delta_{2s}^+ + \left(\frac{\sqrt{2}}{2} + 1 \right) \delta_{2s}^- < 1. \quad (92)$$

Then the BPDN minimizer \hat{Z} is unique and satisfies the error bound

$$\|\hat{Z} - Z\|_2 \leq C_1 s^{-1/2} \|Z - Z^{(s)}\|_1 + C_2 \varepsilon \quad (93)$$

where

$$C_1 = \frac{2 + (\sqrt{2} - 2)\delta_{2s}^- + \sqrt{2}\delta_{2s}^+}{1 - \frac{\sqrt{2}}{2}\delta_{2s}^+ - \left(\frac{\sqrt{2}}{2} + 1\right)\delta_{2s}^-}$$

$$C_2 = \frac{4\sqrt{1 + \delta_{2s}^+}}{1 - \frac{\sqrt{2}}{2}\delta_{2s}^+ - \left(\frac{\sqrt{2}}{2} + 1\right)\delta_{2s}^-}.$$

Remark 8. The real-valued version theorem 3 with the symmetrical RIP ($\delta_s = \delta_s^\pm$) is proved in [3]. The proof for the complex-valued, asymmetrical version follows the same line of reasoning with minor modifications. For the reader's convenience and for the purpose of showing where adjustments are needed, the full proof of theorem 3 is given in appendix B.

Remark 9. Theorem 3 does not guarantee exact recovery of support when $E \neq 0$. An alternative condition in terms of the spectral norm guaranteeing BPDN's *exact localization* property is given in [4] and an application to imaging is given in [14]. Theorem 3 is invoked here because condition (92) is based on RIP and the condition on noise is more relaxed than that of [4], thus directly comparable to those of the performance guarantee for MUSIC. On the other hand, in the noiseless case ($\varepsilon = 0$), BP does *not* guarantee to recover all s -sparse

objects even under the RIP $\delta_{2s} < 1/\sqrt{2}$ according to [15]. By contrast, MUSIC guarantees to localize exactly all s -sparse objects in the absence of noise whenever $\delta_{s+1} < 1$.

Remark 10. In the case of planar objects, under the assumptions of proposition 5 (with $\delta_* = \sqrt{2} - 1$), BPDN yields the exact solution $\hat{Z} = Z$ for $n = \mathcal{O}(s)$ sampling directions (or sensors) and just one incident wave, modulo logarithmic factors. In comparison, the performance guarantee for MUSIC in corollary 3 assumes $n = \mathcal{O}(s)$ sampling *and* incident directions.

Using proposition 3 and theorem 3, we obtain

Corollary 5. *If*

$$s < \frac{1}{2} + \frac{\sqrt{2} - 1}{2\mu(\tilde{\Phi})},$$

cf (115), then (93) holds true.

Remark 11. Under the assumptions of proposition 4 with continuously differentiable f^s and f^i BPDN recovers the s -sparse object exactly $\hat{Z} = Z$ in the noiseless case $\varepsilon = 0$ for $n = \mathcal{O}(s^2)$ and sufficiently high frequency.

This is similar to the performance guarantee for MUSIC in corollary 2. However, the performance guarantee for MUSIC assumes $\mathcal{O}(s^2)$ incident waves while the performance guarantee for BPDN assumes only *one* incident wave.

6. Spectral estimation and source localization

Let us turn to the original application where the MUSIC algorithm arises, namely the source localization and the frequency estimation for multiple random signals. The two applications share almost exactly the same mathematical formulation.

Suppose the random signal $x(t)$ consists of random linear combinations of s time-harmonic components from the set

$$\{e^{-i2\pi\omega t} : \omega \in \mathcal{K}\}, \quad \mathcal{K} = \left\{ \frac{j}{N}, j = 1, \dots, N \right\}.$$

Let us write

$$x(t) = \sum_{j=1}^N a_j e^{-i2\pi\omega_j t}, \quad \omega_j = j/N \quad (94)$$

and assume that there is a fixed set \mathcal{S} (i.e. deterministic support) of s nonzero amplitudes and the elements in the complementary set \mathcal{S}^c are zero almost surely.

Consider the noisy signal model

$$y(t) = x(t) + e(t) \quad (95)$$

where $e(t)$ is the Gaussian white-noise. The task is to find out the frequency content of x by sampling $y(t)$.

Consider random sampling times $t_k, k = 1, \dots, n$, which are i.i.d. uniform r.v.s in the set $\{1, \dots, N\}$. Write $Y = (y(t_k)) \in \mathbb{C}^n, E = (e(t_k)) \in \mathbb{C}^n$ and $Z = (a_j) \in \mathbb{C}^N$. Then by (95) we have

$$Y = \tilde{\Phi}Z + E \quad (96)$$

$$\tilde{\Phi}_{k,j} = \frac{1}{\sqrt{n}} e^{-i2\pi t_k j/N} \in \mathbb{C}^{n \times N},$$

cf (90). From the one-dimensional setting of proposition 5 and remark (7) we know that if (81) is satisfied with $\delta_* = \sqrt{2} - 1$, then the RIC of $\tilde{\Phi}$ obeys bound (82) with a probability of

at least $1 - \gamma$. Applying BPDN to (96) we obtain the error bound (93) with $\mathcal{O}(s)$ data, modulo logarithmic factors.

How does MUSIC perform in this case? The standard MUSIC proceeds as follows. Let $\mathbf{R}_Y = \mathbb{E}[YY^*] \in \mathbb{C}^{n \times n}$, $\mathbf{R}_Z = \mathbb{E}[ZZ^*] \in \mathbb{C}^{N \times N}$ and $\mathbf{R}_E = \mathbb{E}[EE^*]$ be the covariance matrices of Y , Z and E , respectively.

Note that \mathbf{R}_Z is sparse and has at most rank s . \mathbf{R}_Z has rank s , for example, when a_i are zero-mean, independent random variables. In this case, $\mathbf{R}_Z = \text{diag}(\mathbb{E}|a_i|^2)$ has exactly s nonzero diagonal elements.

Suppose the noise $e(t)$ and the signal $x(t)$ are independent of each other. Then we have

$$\mathbf{R}_Y - \mathbf{R}_E = \tilde{\Phi} \mathbf{R}_Z \tilde{\Phi}^*. \quad (97)$$

Replacing the true covariances \mathbf{R}_Y and \mathbf{R}_E by their sample estimates $\hat{\mathbf{R}}_Y$ and $\hat{\mathbf{R}}_E$, we may rewrite (98) as

$$\hat{\mathbf{R}}_Y - \hat{\mathbf{R}}_E = \tilde{\Phi} \mathbf{R}_Z \tilde{\Phi}^* + \mathcal{E} \quad (98)$$

which is of the form (23) with $\mathcal{Y}^\varepsilon = \hat{\mathbf{R}}_Y - \hat{\mathbf{R}}_E$ and $\mathcal{Y} = \tilde{\Phi} \mathbf{R}_Z \tilde{\Phi}^*$. We set $\|\mathcal{E}\|_2 = \varepsilon^2$.

Analogous to theorem 1 and lemma 2 we have

Theorem 4. Let $\|\mathcal{E}\|_2 = \varepsilon^2$ and assume that \mathbf{R}_Z has rank s . If

$$\frac{\|\mathcal{E}\|_2}{\sigma_{\min}} < \frac{1}{2} - \frac{1}{2\sqrt{\sqrt{2}\Gamma_s + 1}}, \quad \sigma_{\min} = \max_{\dim \mathcal{H} = s} \min_{\mathbf{e} \in \mathcal{H}} \frac{\|\tilde{\Phi} \mathbf{R}_Z \tilde{\Phi}^* \mathbf{e}\|_2}{\|\mathbf{e}\|_2} \quad (99)$$

holds, then the frequency content \mathcal{S} can be identified by the thresholding rule

$$\{\mathbf{r} \in \mathcal{K} : J^\varepsilon(\mathbf{r}) \geq 2\Gamma_s^{-2}\}. \quad (100)$$

Lemma 4. The minimum nonzero singular value σ_{\min} of $\tilde{\Phi} \mathbf{R}_Z \tilde{\Phi}^*$ satisfies

$$\sigma_{\min} \geq (1 - \delta_s^-) \zeta_{\min}, \quad \zeta_{\min} = \min_{\mathbf{e} \in \mathbb{C}^s} \frac{\|\mathbf{R}_Z \mathbf{e}\|_2}{\|\mathbf{e}\|_2}.$$

Hence by lemma 3 we obtain the following.

Theorem 5. Assume $\delta_{s+1}^- < 1$ and the setting of theorem 4. If

$$\frac{\varepsilon}{\zeta_{\min}} < (1 - \delta_s^-) \tilde{\Delta}$$

then

$$\mathcal{S} = \left\{ \omega_j = \frac{j}{N}, j = 1, \dots, N : J^\varepsilon(\omega_j) \geq 2\gamma_s^{-2} \right\} \quad (101)$$

where $\tilde{\Delta}$ and γ_s are given by (65).

Remark 12. By proposition 5 $\mathcal{O}(n)$, modulo logarithmic factor, time samples are sufficient to recover the frequency content by MUSIC (theorem 5) or BPDN (theorem 3). However, many realizations of Y and E are needed to calculate the covariance matrices accurately and form equation (98) before the MUSIC reconstruction. Once (98) holds with sufficiently small ε , the noise structure does not affect reconstruction as long as the noise is independent of the signal.

In the absence of abundant realizations of signals, though, BPDN is the preferred method for spectral estimation. Indeed, BPDN can identify the frequency content approximately with just one realization of signals. The recovery error is at worst linearly proportional to the noise level as in (93).

The source localization problem can be treated in the same vein as follows.

Let us assume that s source points are distributed in the grid \mathcal{K} defined in (77) and each source point emits a signal governed by the paraxial Green's function (85) and recorded by the n sensors located at $\mathbf{a}_i, i = 1, \dots, n$, in the plane $z = z_0$.

Let $Z = (\xi(\mathbf{r}_j)) \in \mathbb{C}^N$ be the object vector. After proper normalization, the data vector Y can be written as (96) with the sensing matrix $\tilde{\Phi}$ of the form (87).

Results essentially identical to theorems 4 and 5 and remark 12 follow immediately.

7. Resolution and grid spacing

Being essentially a gridless method, MUSIC's flexibility with grid spacing is an advantage that the current BPDN-based imaging methods do not yet possess.

Let ℓ be a length scale to be determined below and let $\mathcal{S}_\ell = \{\mathbf{r} \in \mathcal{K} : \text{dist}(\mathbf{r}, \mathcal{S}) \leq \ell\}$ be the ℓ -neighborhood of the objects. For the problem of inverse scattering, $\text{dist}(\mathbf{r}, \mathcal{S})$ typically refers to the *physical* or Euclidean distance in the spatial domain. We would like to derive a thresholding rule which can eliminate all false alarms (i.e. artifacts) occurring outside \mathcal{S}_ℓ , no matter how refined the grid spacing is relative to the frequency.

Let $\tilde{\Phi}$ be the extension of Φ over a fine grid of spacing $\tilde{\ell}$ which may be much smaller than ω^{-1} . When $\tilde{\ell} = 0$, the computation domain \mathcal{K} is a continuum.

Generalizing definition (21), we define

$$\Gamma_S(\ell) \equiv \min_{\mathbf{r} \in \mathcal{S}_\ell^c} \|\phi_{\mathbf{r}}\|_2^{-1} \|\mathcal{P}\phi_{\mathbf{r}}\|_2 = \sqrt{1 - \max_{\mathbf{r} \in \mathcal{S}_\ell^c} \|\phi_{\mathbf{r}}\|^{-2} \phi_{\mathbf{r}}^* \Phi \Phi^\dagger \phi_{\mathbf{r}}}. \quad (102)$$

Clearly, $\Gamma_S = \Gamma_S(0^+)$. In the noiseless case, the exact recovery of \mathcal{S} by MUSIC is equivalent to $\Gamma_S(0^+) < 1$.

Extending the proof of lemma 3 to $\Gamma_S(\ell)$ we have

Lemma 5.

$$\Gamma_S(\ell) \geq \gamma_S(\ell) \equiv 1 - \max_{\mathcal{S}'} \frac{\delta_{\mathcal{S}'}^-(1 + \delta_{\mathcal{S}'}^+)}{2 + \delta_{\mathcal{S}'}^- - \delta_{\mathcal{S}'}^+} \quad (103)$$

where $\mathcal{S}' = \mathcal{S}_\ell \cup \{\mathbf{r}\}, \forall \mathbf{r} \in \mathcal{S}_\ell^c$.

Extending the analysis leading to (41)–(42) we have

$$J^\varepsilon(\mathbf{r}) \geq \frac{(1 - 2\rho)^4}{4\rho^2(1 - \rho)^2}, \quad \mathbf{r} \in \mathcal{S} \quad (104)$$

$$J^\varepsilon(\mathbf{r}) \leq \left(\Gamma_S^2(\ell) - \frac{4\rho^2(1 - \rho)^2}{(1 - 2\rho)^4} \right)^{-1}, \quad \mathbf{r} \in \mathcal{S}_\ell^c. \quad (105)$$

The following result is analogous to theorem 1.

Theorem 6. Suppose $\delta_{\mathcal{S}'}^- < 1, \mathcal{S}' = \mathcal{S} \cup \{\mathbf{r}\}, \forall \mathbf{r} \in \mathcal{S}_\ell^c$.

If

$$\frac{\|\mathcal{E}\|_2}{\sigma_{\min}} < \Delta_\ell = \frac{1}{2} - \frac{1}{2} \frac{1}{\sqrt{\sqrt{2}\Gamma_S(\ell) + 1}} \quad (106)$$

then

$$\mathcal{S} \subset \Theta = \{\mathbf{r} \in \mathcal{K} : J^\varepsilon(\mathbf{r}) \geq 2\Gamma_S^{-2}(\ell)\} \quad (107)$$

where $\Theta \cap \mathcal{S}_\ell^c = \emptyset$.

Lemmas 5 and 2 and theorem 6 then implies the following result analogous to theorem 2.

Theorem 7. Suppose $\delta_{S'}^- < 1$, $S' = S \cup \{\mathbf{r}\}$, $\forall \mathbf{r} \in S_\ell^c$. If the NSR obeying the upper bound

$$\frac{\varepsilon}{\xi_{\min}} < \sqrt{(1 + \delta_S^+)^2 \frac{\xi_{\max}^2}{\xi_{\min}^2} + (1 - \delta_S^-)^2 \tilde{\Delta}_\ell} - (1 + \delta_S^+) \frac{\xi_{\max}}{\xi_{\min}} \quad (108)$$

with

$$\tilde{\Delta}_\ell = \frac{1}{2} - \frac{1}{2} \frac{1}{\sqrt{\sqrt{2}\gamma_S(\ell) + 1}} \quad (109)$$

then

$$S \subset \Theta = \{\mathbf{r} \in \mathcal{K} : J^\varepsilon(\mathbf{r}) \geq 2\gamma_S(\ell)^{-2}\} \quad (110)$$

where $\gamma_S(\ell)$ is given in (103) and $\Theta \cap S_\ell^c = \emptyset$.

Let us now give an estimate of the length scale ℓ for (108) to be a useful upper bound for NSR. Let us focus on the general setting of proposition 4, namely arbitrarily located scatterers and random sampling directions.

We resort to the following result analogous to proposition 3. The proof is exactly the same as before and is omitted here.

Proposition 6. For any set $\mathcal{B} \subset \mathcal{K}$, $|\mathcal{B}| \leq r$, we have

$$\delta_{\mathcal{B}}^\pm \leq \mu(\tilde{\Phi}_{\mathcal{B}})(r - 1).$$

To proceed, let us tailor the estimate in proposition 4 to the current setting as follows.

Proposition 7 [12]. Suppose the physical distances between two points corresponding to any two members of $\mathcal{B} \subset \mathcal{K}$ are at least ℓ . Let $\hat{\mathbf{s}}_k$, $k = 1, \dots, n$, be independently drawn from the distribution f^s on the $(d - 1)$ -dimensional sphere independently and identically. Suppose

$$|\mathcal{K}| \leq \frac{\alpha}{8} e^{K^2/2}$$

for any positive constants α , K . Then $\tilde{\Phi}_{\mathcal{B}}$ satisfies the coherence bound

$$\mu(\tilde{\Phi}_{\mathcal{B}}) < \chi^s + \frac{\sqrt{2}K}{\sqrt{n}}$$

with probability greater than $(1 - \alpha)^2$ where χ^s satisfies the bound (71)–(72).

Suppose $\omega\ell \geq C^2n$ for $d = 2$ or $\omega\ell \geq C\sqrt{n}$ for $d = 3$ where C is given by (73) and assume that the s scatterers are separated by at least ℓ from one another. Then, according to propositions 7 and 6, with high probability, for any continuously differentiable sampling distribution f^s

$$\delta_S^\pm \leq \delta_{S'}^\pm \leq 2\sqrt{2}Ks/\sqrt{n}$$

for all $S' = S \cup \{\mathbf{r}\}$, $\mathbf{r} \in S_\ell^c$. Hence we have the following analogous result to corollary 2.

Corollary 6. Suppose $\omega\ell \geq C^2n$ for $d = 2$ or $\omega\ell \geq C\sqrt{n}$ for $d = 3$ with C given by (73).

Under the assumptions of proposition 7 (for $\mathcal{B} = S, S'$), $\sqrt{n}/s \geq 4\sqrt{2}K$ and the NSR bound (75),

$$S \subset \Theta = \left\{ \mathbf{r} \in \mathcal{K} : J^\varepsilon(\mathbf{r}) \geq \frac{128}{25} \right\} \quad (111)$$

Table 3. MUSIC algorithm with thresholding (101) which can be made explicit as (76) of corollary 2.

MUSIC algorithm with thresholding for equation (98)

- Input $\hat{\mathbf{R}}_Y, \hat{\mathbf{R}}_E, \hat{\Phi}$ and the sparsity s .
 - Compute the orthogonal projector \mathcal{P}^ε onto the noise subspace of $\mathcal{Y}^\varepsilon = \hat{\mathbf{R}}_Y - \hat{\mathbf{R}}_E$.
 - Compute the function $J^\varepsilon(\omega) = |\mathcal{P}^\varepsilon \phi_\omega|^{-2}, \omega \in \mathcal{K}$.
 - Output the set of points corresponding to the s highest values of J^ε or equivalently the set $\{\omega \in \mathcal{K} : J^\varepsilon(\omega) \geq 2\gamma_s^{-2}\}$ where $\gamma_s = 1 - \frac{\delta_{s+1}^-(1+\delta_s^+)}{2+\delta_s^+-\delta_{s+1}^-}$ as the frequency content.
-

with probability at least $(1 - \alpha)^2$ where $\Theta \cap \mathcal{S}_\ell^c = \emptyset$.

Corollary 6 asserts that under its conditions the MUSIC algorithm (cf table 2) with the explicit thresholding rule (111) can recover, with a probability of at least $(1 - \alpha)^2$, the object locations within the accuracy ℓ .

Likewise, for the spectral estimation problem discussed in section 6, consider (98) with an arbitrarily refined grid \mathcal{K} . We have the following performance guarantee (cf theorem 5).

Theorem 8. Assume $\delta_{S'}^- < 1, S' = \mathcal{S} \cup \{\omega\}, \forall \omega \in \mathcal{S}^c(\ell)$. If

$$\frac{\varepsilon}{\zeta_{\min}} < (1 - \delta_{S'}^-) \tilde{\Delta}_\ell$$

with $\tilde{\Delta}_\ell$ given in (109) then

$$S \subset \Theta = \{\omega \in \mathcal{K} : J^\varepsilon(\omega) \geq 2\gamma_S(\ell)^{-2}\} \quad (112)$$

where $\gamma_S(\ell)$ is given in (103) and $\Theta \cap \mathcal{S}_\ell^c = \emptyset$.

In other words, the MUSIC algorithm as stated in table 3 can recover, with high probability, the frequency content within the accuracy ℓ .

8. Numerical tests

In the simulations, $z_0 = 10000, \lambda = 0.1$ and the search domain is $[-250, 250]^2$ with grid spacing $\ell = 10$ on the transverse plane $z = 0$. The scatterers are independently and uniformly distributed on the grid with amplitudes independently and uniformly distributed in the range $[1, 2]$. The sensors are independently and uniformly distributed in the domain $[-A/2, A/2]^2$ with various A . The source locations are identical to the sensor locations. In the setup, condition (88) is satisfied with $A = 100$. With these parameters, the paraxial regime is about to set in (cf [16]). Note, however, that all the simulations are performed with the exact Green's function.

In our simulations we have used the Matlab codes YALL1 (acronym for Your ALgorithms for L1, available at <http://www.caam.rice.edu/optimization/L1/YALL1/>). YALL1 is an L1-minimization solver based on the alternating direction method [29].

Figure 2 compares the performances of MUSIC and BPDN in the well-resolved case $A = 100$ and the under-resolved case $A = 10$ where the aperture is only one-tenth of that satisfying (88). For figure 2 BPDN is carried out on the data matrix \mathbf{Y} with the sensors coincident with the sources, i.e. $\Phi = \Psi$. To put the problem in the proper setup for BP, we vectorize \mathbf{Y} by staking its n columns and denote the resulting \mathbb{C}^{n^2} vector by Y . We vectorize the diagonal matrix $\tilde{\mathbf{X}}$ by listing its N diagonals as a \mathbb{C}^N vector \tilde{X} . The BPDN performance of

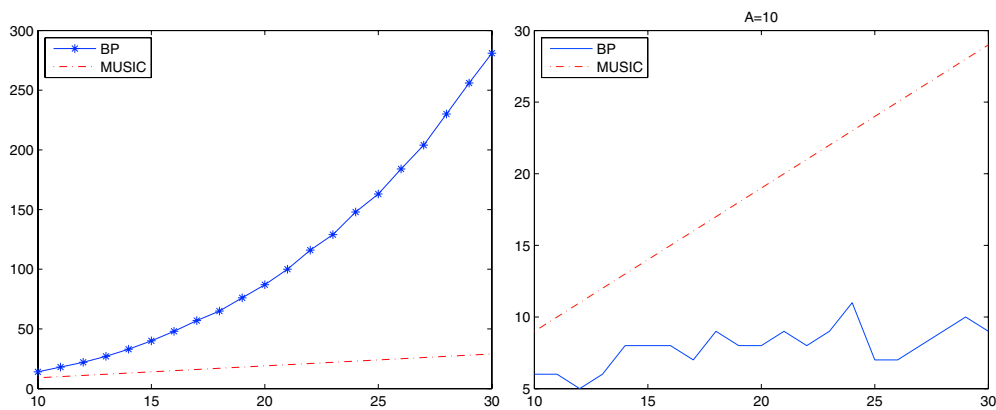


Figure 2. The number s (vertical axis) of recoverable scatterers versus the number of sensors n (horizontal axis) of MUSIC (dashed) and BPDN (solid) with both methods employing the whole data matrix in the well-resolved case $A = 100$ (left) and the under-resolved case $A = 10$ (right). In the well-resolved case, BPDN delivers a much better (quadratic-in- n) performance than MUSIC; in the under-resolved case, MUSIC outperforms BPDN whose performance tends to be unstable in this regime. The numbers of recoverable scatterers by BPDN are calculated based on successful recovery of at least 90 out of 100 independent realizations of transceivers and scatterers while the success rate of MUSIC is 100%.

this setup has been analyzed in [16]. The numbers of recoverable scatterers shown in figure 2 are computed for at least a 90% recovery rate based on 100 independent realizations of transceivers and scatterers. In both cases, MUSIC recovers $s = n - 1$ scatterers with certainty. Clearly, for the well-resolved case, BPDN has a far superior performance to MUSIC. Indeed, it can be shown that BPDN can recover $s = \mathcal{O}(n^2)$ scatterers with high probability in the well-resolved case [16]. The quadratic behavior is illustrated by the near-parabolic curve in figure 8 (left). For the under-resolved case, however, MUSIC outperforms BPDN by a significant margin, figure 8 (right). As pointed out in remark 3, the MUSIC algorithm has the super-resolution capability for a sufficiently small noise-to-scatterer ratio.

If only one column of \mathbf{Y} is used in BPDN, as discussed in section 5, then MUSIC outperforms BPDN by a wide margin even in the well-resolved case, figure 3.

We further investigate the performance of the MUSIC algorithm for the extremely under-resolved case when BPDN essentially has extremely low probabilities of exact recovery (even for $s = 1$). Figure 4 shows the success probabilities of MUSIC as a function of aperture for various n and s while figure 5 shows the success probabilities of MUSIC as a function of n for various A and s . The success rates are calculated from 1000 independent realizations of transceivers and scatterers.

Three observations about figure 4 are in order: (i) the optimal performance of $s = n - 1$ does not hold with certainty for s relatively large with respect to aperture (cf left and right panels); (ii) increasing the number of randomly selected transceivers reduces the aperture required for the same probability of recovering the same number of scatterers (left to middle panels); (iii) increasing the number of randomly selected scatterers increases the aperture required for the same probability of recovery with the same number of transceivers (middle to right panels).

Likewise, the success rates increase with the number of transceivers for any aperture and sparsity (figure 5). The most interesting plot in figure 5 is the middle panel which

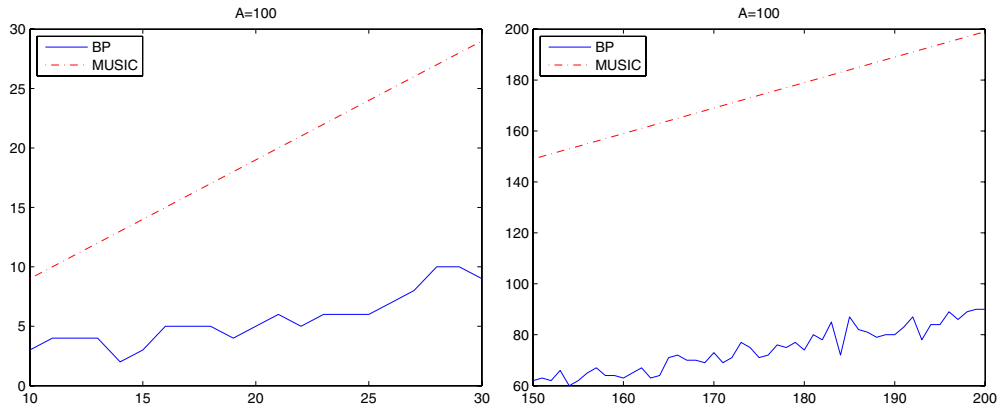


Figure 3. The number s (vertical axis) of recoverable scatterers versus the number of sensors n (horizontal axis) of MUSIC (dashed) and BPDN (solid) with BPDN employing only a *single* column of the data matrix in the well-resolved case $A = 100$ with $n \in [10, 30]$ (left) and $n \in [150, 200]$ (right). Both BPDN curves show a roughly linear behavior with a slope less than that of the MUSIC curves.

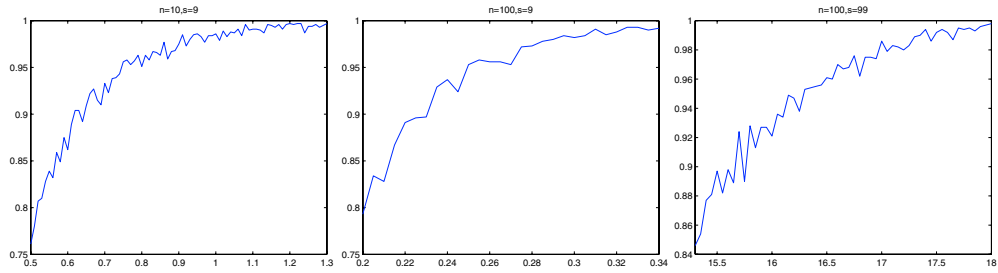


Figure 4. Success probability of the MUSIC reconstruction versus aperture for $n = 10, s = 9$ (left), $n = 100, s = 9$ (middle) and $n = 100, s = 99$ (right). Note the different aperture ranges for the three plots. The success rate is calculated from 1000 trials. Increasing the number of transceivers for the same number of scatterers reduces the aperture required for the same success rate. The reduction of aperture is about three folds (left to middle). On the other hand, a higher number of scatterers with the same number of transceivers also demands larger aperture for the same success rate. The increase in aperture is about seven times (middle to right).

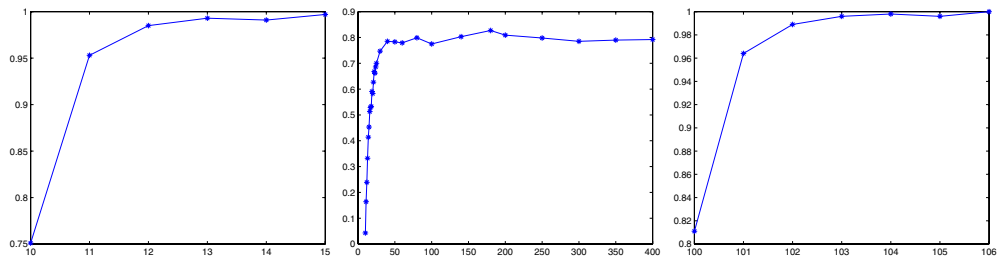


Figure 5. Success probability of MUSIC versus the number of transceivers with $A = 0.5, s = 9$ (left), $A = 0.2, s = 9$ (middle) and $A = 15, s = 99$ (right). The probabilities are calculated from 1000 independent trials.

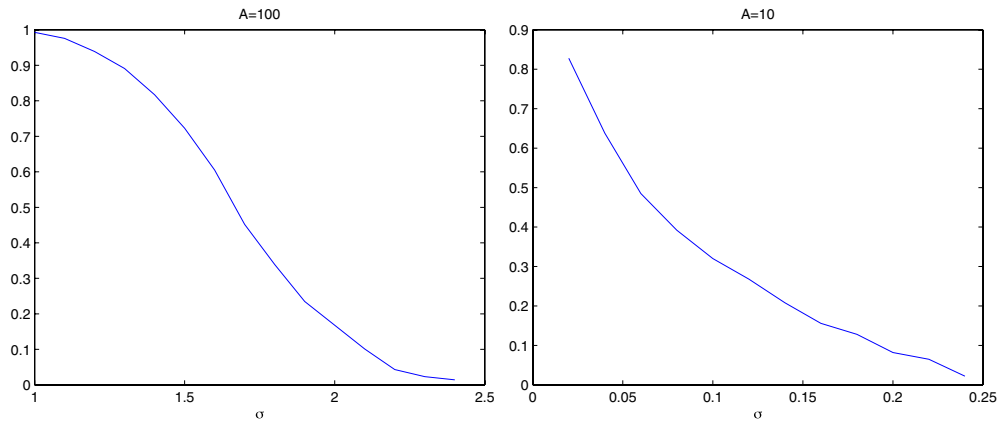


Figure 6. Success probability of MUSIC reconstruction of $s = 10$ scatterers with $n = 100$ transceivers versus the noise level σ in the well-resolved case $A = 100$ (left) and the under-resolved case $A = 10$ (right). The success rate is calculated from 1000 trials. Note the different scales of σ in the two plots. Noise sensitivity increases dramatically in the under-resolved case.

shows for $A = 0.2$, $s = 9$ that the success rate curve becomes a plateau after reaching 80%. This is *not* inconsistent with the prediction of proposition 1 since proposition 1 assumes a *fixed* configuration of scatterers while figure 5 is for random, independent realizations of scatterers. In other words, the threshold n_0 in proposition 1 may *not* be uniformly valid for all configurations of s scatterers in the under-resolved case. On the other hand, when the aperture increases by two and a half times to $A = 0.5$ and the number of transceivers increases to 15, the performance becomes uniform with respect to the scatterer configuration (left panel).

Figure 6 shows the noise sensitivity of MUSIC reconstruction of ten scatterers with 100 transceivers. Here n and s are chosen so that (81) is roughly satisfied. We add the i.i.d. noises

$$\sigma(e_1 + ie_2)Y_{\max} \quad (113)$$

to the entries of the unperturbed data matrix where e_1 and e_2 are independent, uniform r.v.s in $[-1, 1]$ and Y_{\max} is the maximum absolute value of the data entries. Hence the signal-to-noise ratio (SNR) is about $2^{-1}\sigma^{-2}$. In the well-resolved case ($A = 100$) the MUSIC reconstruction can withstand a significant amount of noise in the data matrix. Indeed, at SNR 0.5 the success rate is almost 100%, consistent with the prediction of theorem 2, and even at SNR 0.22 ($\sigma = 1.5$) the success rate can be indefinitely improved by increasing the number of transceivers (figure 7, left panel).

In the under-resolved case, however, the noise sensitivity increases significantly. Figure 6 (right panel) reminds us how fragile the superior performance of MUSIC in the under-resolved case is, cf figure 8 (right panel). Figure 7 (right panel) further indicates that in the under-resolved case the success rate may not be indefinitely improved by increasing the number of transceivers in the presence of noise.

9. Conclusions and discussions

We have developed a framework for discrete quantitative analysis of the MUSIC algorithm in the well-resolved case. Our approach is based on the RIP (44) and its variant (45) which takes into account the object configuration as well as sparsity.

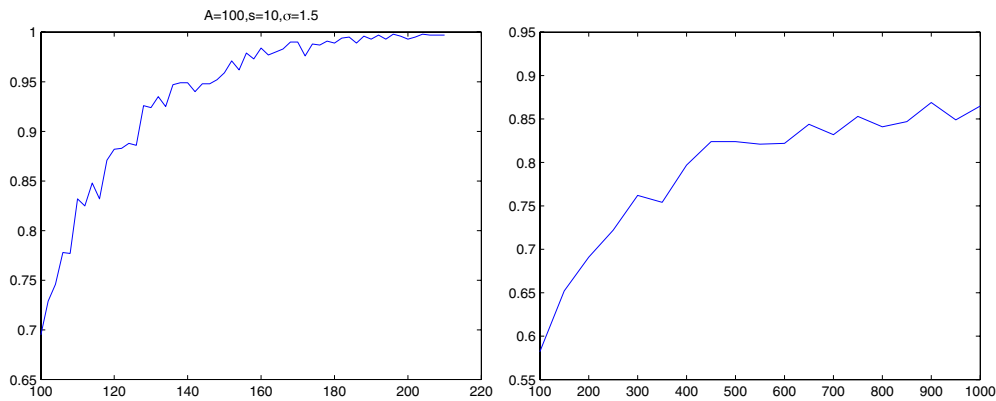


Figure 7. Success probability of MUSIC reconstruction of $s = 10$ scatterers as a function of n with $\sigma = 150\%$ in the well-resolved case $A = 100$ (left) and $\sigma = 5\%$ in the under-resolved case $A = 10$ (right). The success rate reaches the plateau of 85% near $n = 1000$ in the under-resolved case. The success rate is calculated from 1000 trials.

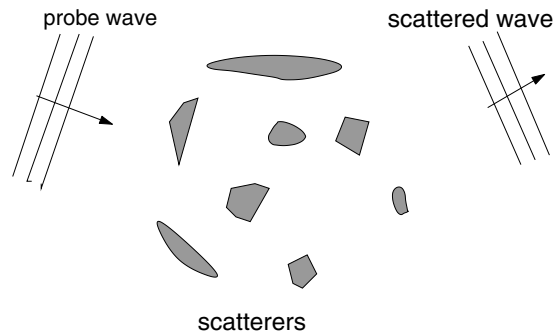


Figure 8. Scattering by extended objects.

Our first main result is a support recovery condition (theorem 2) that for the NOR obeying (64) MUSIC can exactly localize the objects with noisy data. Our result indicates the super-resolution capability of the MUSIC algorithm when the noise level is sufficiently low (remark 3).

We have provided a coherence approach to estimating RIC (propositions 4 and 3) for general object configuration in three dimensions with the grid spacing $\ell \sim \lambda s$ and the sensor number $n \sim s^2$. When the scatterers are distributed in a transverse plane, then $\ell \sim \lambda, n \sim s$ (modulo logarithmic factors) suffices. We have extended these results to the gridless setting for which ℓ is interpreted as the minimum distance between objects and only approximate localization up to the error ℓ is sought (theorem 7 and corollary 6).

Our comparative analysis shows that when the whole data matrix is employed in both BPDN and MUSIC, BPDN outperforms MUSIC in the well-resolved case in the sense that the number of objects recoverable by BPDN grows quadratically with the number of transceivers while that by MUSIC grows linearly. The MUSIC reconstruction can tolerate a significant amount of noise (figures 6 and 7, left panels). On the other hand, our numerical results show that in the under-resolved case MUSIC outperforms BPDN by a wide margin (figure 2, right panel; figures 4 and 2). However, MUSIC's super-resolution effect can only tolerate a small amount of noise (figures 6 and 7, right panels).

Even in the well-resolved case where the employment of just *one* column of the data matrix by BPDN guarantees a probabilistic recovery of objects numbered in linear proportion to the number of sensors, analogous to the performance guarantee of MUSIC, the latter outperforms the former in numerical simulations by a wide margin (figure 3).

Finally, it is worthwhile to compare the MUSIC algorithm with the other algorithms for *multiple measurement vectors* such as the orthogonal matching pursuit (OMP) [8, 28].

First, the MUSIC algorithm requires the input of the object sparsity while OMP does not. Second, the MUSIC algorithm requires the input of at least the object sparsity *plus one* measurement vectors while OMP is flexible with the number of measurement vectors. Third, the best-known sufficient condition for exact localization by OMP is [11]

$$\frac{\varepsilon}{\xi_{\min}} \leq \frac{1}{2} + \mu(\tilde{\Phi}) \left(\frac{1}{2} - s \right). \quad (114)$$

In order for the right-hand side of (114) to be positive, it is necessary that

$$s < \frac{1}{2} + \frac{1}{2\mu(\tilde{\Phi})} \quad (115)$$

which in view of proposition 4 implies $s = \mathcal{O}(\sqrt{n})$. In contrast, the best performance guarantee for MUSIC is $n \sim s$ as noted before. Fourth, like most compressed sensing techniques OMP does not perform well with an arbitrarily refined grid while MUSIC can localize the objects in such a grid within the accuracy determined by the measurement *sub-matrix* restricted to the object support.

Acknowledgments

I am grateful to Mike Yan for preparing the figures and the National Science Foundation for supporting the research through grant DMS 0908535. I thank Wenjing Liao for pointing out the result, proposition 2, which helps improve the results of the previous version of the manuscript. The research is partially supported by the NSF grant DMS-0908535.

Appendix A. Sparse extended objects

In this appendix we extend the MUSIC algorithm to image sparse extended scatterers by interpolating from grid points.

Suppose that the object function $\xi(\mathbf{r})$ has a compact support. Consider the discrete approximation by interpolating from the grid points

$$\xi_\ell(\mathbf{r}) = \ell^2 \sum_{\mathbf{q} \in \mathcal{I}} g(\mathbf{r}/\ell - \mathbf{q}) \xi(\ell\mathbf{q}) \quad \mathcal{I} \subset \mathbb{Z}^d$$

where g is some spline function and ℓ is the grid spacing. Since ξ has a compact support, \mathcal{I} is a finite set. For simplicity assume $d = 2$ and let \mathcal{I} be the finite lattice

$$\mathcal{I} = \{\mathbf{q} = (q_1, q_2) : q_1, q_2 = 1, \dots, \sqrt{N}\}$$

of total cardinality N and $\mathcal{K} = \ell\mathcal{I}$. In the case of a characteristic function g , ξ_ℓ is a piecewise constant object function. We will neglect the discretization error and assume $\xi_\ell(\mathbf{r}) = \xi(\mathbf{r})$ in the subsequent analysis.

The data matrix $\mathbf{Y} \in \mathbb{C}^{n \times m}$ is given by

$$\begin{aligned} Y_{k,l} &\sim \ell^d \sum_{\mathbf{q} \in \mathcal{I}} \xi(\ell\mathbf{q}) \int_{\mathbb{R}^d} g(\mathbf{r}'/\ell - \mathbf{q}) e^{i\omega(\hat{\mathbf{d}}_l - \hat{\mathbf{s}}_k) \cdot \mathbf{r}'} d\mathbf{r}', \\ &= \ell^d (2\pi)^{d/2} \hat{g}(\ell\omega(\hat{\mathbf{d}}_l - \hat{\mathbf{s}}_k)) \sum_{\mathbf{q} \in \mathcal{I}} \xi(\ell\mathbf{q}) e^{i\omega\ell(\hat{\mathbf{d}}_l - \hat{\mathbf{s}}_k) \cdot \mathbf{q}}. \end{aligned} \quad (\text{A.1})$$

As before we maintain the option of normalizing \mathbf{Y} . Suppose $\{\xi_j = \xi(\ell \mathbf{q}_j) : j = 1, \dots, s\}$ is the set of nonvanishing $\xi(\ell \mathbf{q})$ and let $\mathbf{X} = \text{diag}(\xi_j) \in \mathbb{C}^{s \times s}$. Dividing (A.1) by $\ell^d (2\pi)^{d/2} \hat{g}(\ell\omega(\hat{\mathbf{d}}_l - \hat{\mathbf{s}}_k))/2$ we can write the data matrix in the form (12) with the sensing matrices

$$\begin{aligned}\Phi_{kj} &= \frac{1}{\sqrt{n}} e^{-i\omega \ell \hat{\mathbf{s}}_k \cdot \mathbf{q}} \in \mathbb{C}^{n \times s} \\ \Psi_{lj} &= \frac{1}{\sqrt{n}} e^{-i\omega \ell \hat{\mathbf{d}}_l \cdot \mathbf{q}} \in \mathbb{C}^{n \times s}\end{aligned}$$

where $j = (q_1 - 1)\sqrt{N} + q_2$.

In other words, the scattering analysis for both point and extended scatterers leads to the same type of Fourier-like matrices.

Appendix B. Proof of theorem 3

The following lemma differs from the original version in [3].

Lemma 6. *We have*

$$|\Re(\tilde{\Phi}Z, \tilde{\Phi}Z')| \leq \frac{1}{2} (\delta_{s+s'}^+ + \delta_{s+s'}^-) \|Z\|_2 \|Z'\|_2$$

for all Z, Z' supported on disjoint subsets $T, T' \subset \{1, \dots, m\}$ with $|S| \leq s, |S'| \leq s'$.

Proof. Without loss of generality, suppose Z, Z' are unit vectors. Since $Z \perp Z'$, $\|Z \pm Z'\|_2^2 = 2$. Hence we have from the RIP (44)

$$2(1 - \delta_{s+s'}^-) \leq \|\tilde{\Phi}(Z \pm Z')\|_2^2 \leq 2(1 + \delta_{s+s'}^+) \quad (\text{B.1})$$

By the parallelogram identity and (B.1)

$$|\Re(\tilde{\Phi}Z, \tilde{\Phi}Z')| = \frac{1}{4} \|\tilde{\Phi}Z + \tilde{\Phi}Z'\|_2^2 - \|\tilde{\Phi}Z - \tilde{\Phi}Z'\|_2^2 \leq \frac{1}{2} (\delta_{s+s'}^+ + \delta_{s+s'}^-)$$

which proves the lemma. \square

By the triangle inequality and the fact that Z is in the feasible set we have

$$\|\tilde{\Phi}(\hat{Z} - Z)\|_2 \leq \|\tilde{\Phi}\hat{Z} - Y\|_2 + \|Y - \tilde{\Phi}Z\|_2 \leq 2\varepsilon. \quad (\text{B.2})$$

Set $\hat{Z} = Z + \Delta$ and decompose Δ into a sum of vectors $\Delta_{S_0}, \Delta_{S_1}, \Delta_{S_2}, \dots$, each of sparsity at most s . Here S_0 corresponds to the locations of the s largest coefficients of Z ; S_1 to the locations of the s largest coefficients of Δ_{S_0} ; S_2 to the locations of the next s largest coefficients of Δ_{S_0} , and so on.

Step 1. For $j \geq 2$,

$$\|\Delta_{S_j}\|_2 \leq s^{1/2} \|\Delta_{S_j}\|_\infty \leq s^{-1/2} \|\Delta_{S_{j-1}}\|_2$$

and hence

$$\sum_{j \geq 2} \|\Delta_{S_j}\|_2 \leq s^{-1/2} \sum_{j \geq 1} \|\Delta_{S_j}\|_1 \leq s^{-1/2} \|\Delta_{S_0^c}\|_1. \quad (\text{B.3})$$

This yields

$$\|\Delta_{(S_0 \cup S_1)^c}\|_2 = \left\| \sum_{j \geq 2} \Delta_{S_j} \right\|_2 \leq \sum_{j \geq 2} \|\Delta_{S_j}\|_2 \leq s^{-1/2} \|\Delta_{S_0^c}\|_1. \quad (\text{B.4})$$

Also we have

$$\|Z\|_1 \geq \|\hat{Z}\|_1 = \|Z_{S_0} + \Delta_{S_0}\|_1 + \|Z_{S_0^c} + \Delta_{S_0^c}\|_1 \geq \|Z_{S_0}\|_1 - \|\Delta_{S_0}\|_1 - \|Z_{S_0^c}\|_1 + \|\Delta_{S_0^c}\|_1$$

which implies

$$\|\Delta_{S_0^c}\|_1 \leq 2\|Z_{S_0^c}\|_1 + \|\Delta_{S_0}\|_1. \quad (\text{B.5})$$

Note that $\|Z_{S_0^c}\|_1 = \|Z - Z^{(s)}\|_1$ by definition. Applying (B.4), (B.5) and the Cauchy–Schwarz inequality gives

$$\|\Delta_{(S_0 \cup S_1)^c}\|_2 \leq \|\Delta_{S_0}\|_2 + 2e_0 \quad (\text{B.6})$$

where $e_0 \equiv s^{-1/2}\|Z - Z^{(s)}\|_1$.

Step 2. Observe

$$\begin{aligned} \|\tilde{\Phi}\Delta_{S_0 \cup S_1}\|_2^2 &= \langle \tilde{\Phi}\Delta_{S_0 \cup S_1}, \tilde{\Phi}\Delta \rangle - \left\langle \tilde{\Phi}\Delta_{S_0 \cup S_1}, \sum_{j \geq 2} \tilde{\Phi}\Delta_{S_j} \right\rangle \\ &= \Re \langle \tilde{\Phi}\Delta_{S_0 \cup S_1}, \tilde{\Phi}\Delta \rangle - \sum_{j \geq 2} \Re \langle \tilde{\Phi}\Delta_{S_0 \cup S_1}, \tilde{\Phi}\Delta_{S_j} \rangle \\ &= \Re \langle \tilde{\Phi}\Delta_{S_0 \cup S_1}, \tilde{\Phi}\Delta \rangle - \sum_{j \geq 2} [\Re \langle \tilde{\Phi}\Delta_{S_0}, \tilde{\Phi}\Delta_{S_j} \rangle + \Re \langle \tilde{\Phi}\Delta_{S_1}, \tilde{\Phi}\Delta_{S_j} \rangle]. \end{aligned} \quad (\text{B.7})$$

This calculation differs slightly from the corresponding calculation in [3].

From (B.2) and the RIP (44) it follows that

$$\|\langle \tilde{\Phi}\Delta_{S_0 \cup S_1} \rangle\| \leq \|\tilde{\Phi}\Delta_{S_0 \cup S_1}\|_2 \|\tilde{\Phi}\Delta\|_2 \leq 2\varepsilon\sqrt{1 + \delta_{2s}^+} \|\Delta_{S_0 \cup S_1}\|_2.$$

Moreover, it follows from lemma appendix B that

$$\begin{aligned} |\Re \langle \tilde{\Phi}\Delta_{S_0}, \tilde{\Phi}\Delta_{S_j} \rangle| &\leq \frac{1}{2} (\delta_{2s}^+ + \delta_{2s}^-) \|\Delta_{S_0}\|_2 \|\Delta_{S_j}\|_2 \\ |\Re \langle \tilde{\Phi}\Delta_{S_1}, \tilde{\Phi}\Delta_{S_j} \rangle| &\leq \frac{1}{2} (\delta_{2s}^+ + \delta_{2s}^-) \|\Delta_{S_0}\|_2 \|\Delta_{S_j}\|_2 \end{aligned}$$

for $j \geq 2$. Since S_0 and S_1 are disjoint

$$\|\Delta_{S_0}\|_2 + \|\Delta_{S_1}\|_2 \leq \sqrt{2}\sqrt{\|\Delta_{S_0}\|_2^2 + \|\Delta_{S_1}\|_2^2} = \sqrt{2}\|\Delta_{S_0 \cup S_1}\|_2.$$

Also

$$\begin{aligned} (1 - \delta_{2s}^-) \|\Delta_{S_0 \cup S_1}\|_2^2 &\leq \|\tilde{\Phi}\Delta_{S_0 \cup S_1}\|_2^2 \leq \|\Delta_{S_0 \cup S_1}\|_2 \\ &\times \left(2\varepsilon\sqrt{1 + \delta_{2s}^+} + \frac{1}{\sqrt{2}} (\delta_{2s}^+ + \delta_{2s}^-) \sum_{j \geq 2} \|\Delta_{S_j}\|_2 \right). \end{aligned}$$

Therefore, from (B.3) we obtain

$$\|\Delta_{S_0 \cup S_1}\|_2 \leq \alpha\varepsilon + \rho s^{-1/2} \|\Delta_{S_0^c}\|_1, \quad \alpha = \frac{2\sqrt{1 + \delta_{2s}^+}}{1 - \delta_{2s}^-}, \quad \rho = \frac{\frac{1}{\sqrt{2}} (\delta_{2s}^+ + \delta_{2s}^-)}{1 - \delta_{2s}^-}$$

and moreover by (B.5) and the definition of e_0

$$\|\Delta_{S_0 \cup S_1}\|_2 \leq \alpha\varepsilon + \rho \|\Delta_{S_0}\|_2 + 2\rho e_0.$$

Namely,

$$\|\Delta_{S_0 \cup S_1}\|_2 \leq (1 - \rho)^{-1} (\alpha\varepsilon + 2\rho e_0)$$

if (92) holds.

Finally,

$$\|\Delta\|_2 \leq \|\Delta_{S_0 \cup S_1}\|_2 + \|\Delta_{(S_0 \cup S_1)^c}\|_2 \leq 2\|\Delta_{S_0 \cup S_1}\|_2 + 2e_0 \leq 2(1 - \rho)^{-1} (\alpha\varepsilon + (1 + \rho)e_0)$$

which is what we set out to show.

References

- [1] Ammari H, Lakovleva E and Lesselier D 2005 A MUSIC algorithm for locating small inclusions buried in a half-space from scattering amplitude at a fixed frequency *SIAM Multiscale Model. Simul.* **3** 597–628
- [2] Bruckstein A M, Donoho D L and Elad M 2009 From sparse solutions of systems of equations to sparse modeling of signals *SIAM Rev.* **51** 34–81
- [3] Candès E J 2008 The restricted isometry property and its implications for compressed sensing *C. R. Acad. Sci., Paris I* **346** 589–92
- [4] Candès E J and Plan Y 2009 Near-ideal model selection by L1 minimization *Ann. Stat.* **37** 2145–77
- [5] Candès E J and Tao T 2005 Decoding by linear programming *IEEE Trans. Inform. Theory* **51** 4203–15
- [6] Chen S S, Donoho D L and Saunders M A 1998 Atomic decomposition by basis pursuit *SIAM J. Sci. Comput.* **20** 33–61
- [7] Cheney M 2001 The linear sampling method and MUSIC algorithm *Inverse Problems* **17** 591–6
- [8] Cotter S F, Rao B D, Engan K and Kreutz-Delgado K 2005 Sparse solution to linear inverse problems with multiple measurement vectors *IEEE Trans. Signal Proc.* **53** 2477–88
- [9] Devaney A J 2000 Super-resolution processing of multi-static data using time reversal and MUSIC unpublished
- [10] Devaney A J, Marengo E A and Gruber F K 2005 Time-reversal-based imaging and inverse scattering of multiply scattering point objects *J. Acoust. Soc. Am.* **118** 3129–38
- [11] Donoho D L, Elad M and Temlyakov V N 2006 Stable recovery of sparse overcomplete representations in the presence of noise *IEEE Trans. Inform. Theory* **52** 6–18
- [12] Fannjiang A 2010 Compressive inverse scattering: I. High frequency SIMO/MISO and MIMO measurements *Inverse Problems* **26** 035008
- [13] Fannjiang A 2010 Compressive inverse scattering: II. Multi-shot SISO measurements with Born scatterers *Inverse Problems* **26** 035009
- [14] Fannjiang A 2010 Exact localization and superresolution with noisy data and random illumination arXiv:1008.3146
- [15] Davies M E and Gribonval R 2009 Restricted isometry constants where ℓ_p sparse recovery can fail for $0 < p <= 1$ *IEEE Trans. Inform. Theory* **55** 2203–14
- [16] Fannjiang A, Yan P and Strohmer T 2009 Compressed remote sensing of sparse objects arXiv:0904.3994
- [17] Gruber F K, Marengo E A and Devaney A J 2004 Time-reversal imaging with multiple signal classification considering multiple scattering between the objects *J. Acoust. Soc. Am.* **115** 3042–7
- [18] Hou S, Huang K, Solna K and Zhao H 2009 A phase and space coherent direct imaging method *J. Acoust. Soc. Am.* **125** 227–38
- [19] Horn R A and Johnson C R 1985 *Matrix Analysis* (Cambridge: Cambridge University Press)
- [20] Jaffard S, Meyer Y and Ryan R D 2001 *Wavelets—Tools for Science and Technology* (Philadelphia, PA: SIAM)
- [21] Kirsch A 2002 The MUSIC-algorithm and the factorization method in inverse scattering theory for inhomogeneous media *Inverse Problems* **18** 1025–40
- [22] Kirsch A and Grinsberg N 2008 *The Factorization Method for Inverse Problems* (Oxford: Oxford University Press)
- [23] Prada C and Thomas J-L 2003 Experimental subwavelength localization of scatterers by decomposition of the time reversal operator interpreted as a covariance matrix *J. Acoust. Soc. Am.* **114** 235–43
- [24] Rauhut H 2008 Stability results for random sampling of sparse trigonometric polynomials *IEEE Trans. Inform. Theory* **54** 5661–70
- [25] Schmidt R 1986 Multiple emitter location and signal parameter estimation *IEEE Trans. Antennas Propag.* **34** 276–80
- [26] Stewart G W and Sun J-G 1990 *Matrix Perturbation Theory* (New York: Academic)
- [27] Therrien C W 1992 *Discrete Random Signals and Statistical Signal Processing* (Englewood Cliffs, NJ: Prentice-Hall)
- [28] Tropp J A, Gilbert A C and Strauss M J 2006 Algorithms for simultaneous sparse approximation: I. Greedy pursuit *Signal Process.* **86** 572–88
- [29] Yang J and Zhang Y 2009 Alternating direction algorithms for ℓ_1 -problems in compressive sensing arXiv:0912.1185