



Improved Resolution Estimate for the Two-Dimensional Super-Resolution and a New Algorithm for Direction of Arrival Estimation with Uniform Rectangular Array

Ping Liu¹ · Habib Ammari¹

Received: 17 May 2022 / Revised: 30 January 2023 / Accepted: 22 May 2023
© The Author(s) 2023

Abstract

In this paper, we develop a new technique to obtain improved estimates for the computational resolution limits in two-dimensional super-resolution problems and present a new idea for developing two-dimensional super-resolution algorithms. To be more specific, our main contributions are fourfold: (1) Our work improves the resolution estimates for number detection and location recovery in two-dimensional super-resolution problems; (2) As a consequence, we derive a stability result for a sparsity-promoting algorithm in two-dimensional super-resolution problems [or direction of arrival Problems (DOA)]. The stability result exhibits the optimal performance of sparsity promoting in solving such problems; (3) Inspired by the new techniques, we propose a new coordinate-combination-based model order detection algorithm for two-dimensional DOA estimation and theoretically demonstrate its optimal performance, and (4) we also propose a new coordinate-combination-based MUSIC algorithm for super-resolving sources in two-dimensional DOA estimation. It has excellent performance and enjoys some advantages compared to the conventional DOA algorithms.

Keywords Two-dimensional super-resolution · Direction of arrival algorithms · Resolution estimates · Stability results · Sparsity-promoting algorithm · Model order detection · MUSIC algorithm

Communicated by Carola-Bibiane Schönlieb.

This work was supported in part by the Swiss National Science Foundation Grant Nos. 200021–200307.

✉ Ping Liu
ping.liu@sam.math.ethz.ch
Habib Ammari
habib.ammari@math.ethz.ch

¹ Department of Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zurich, Switzerland

1 Introduction

It is well-known that the physical nature of wave propagation and diffraction imposes a fundamental barrier in the resolution of imaging systems, which is termed diffraction limit or resolution limit. Since the famous works of Abbe [1] and Rayleigh [57] for quantifying the resolution limit, it is widely used in practice to date that the resolution limit is near half of the wavelength (see, for instance, [5, 6]). Although this kind of resolution limit was widely used, it lacks of mathematical foundations and not very applicable to some modern imaging modalities [15, 56]. From the mathematical perspective, the resolution limit could only be set when taking into account the noise [12, 21, 23] and surpassing these classical resolution limits is very promising for imaging modalities with high signal-to-noise ratio (SNR). This understanding motivates new works on deriving more rigorous resolution limits [30, 31, 49, 50]. At the beginning of this century, the dependence of two-point resolution on the noise level has been thoroughly investigated from the perspective of statistical inference [63–65], but the resolution estimates for resolving multiple sources only achieve substantial breakthroughs in recent years due to its nonlinearity.

To understand the resolution in resolving multiple sources, in the earlier works [44–46] we have defined “computational resolution limits” for number detection and location recovery in the one- and multi-dimensional super-resolution problems and characterized them by the signal-to-noise ratio, cutoff frequency, and number of sources. In [45], we derived sharp estimates for the computational resolution limits in one dimensional super-resolution problems. We extended the estimations to multi-dimensional cases in [44], but the new estimation is not that sharp due to the techniques of projection used there. Specifically, the upper bound for the resolution increases rapidly as the source number n and space dimensionality k increases. To address this issue, this paper aims to derive better estimates for the computational resolution limits in two-dimensional super-resolution problems and provide a new idea having the potential to tackle general multi-dimensional cases. The main contributions of our work are fourfold: (1) Our work improves the resolution estimates in [44] for number detection and location recovery in two-dimensional super-resolution problems; (2) As a consequence, we derive a stability result for a sparsity-promoting algorithm in two-dimensional super-resolution problems [or direction of arrival problems (DOA)]. Although it is well-known that the total variation optimization [11] and many other convex optimization based algorithms [70] have a resolution limit near the Rayleigh limit [16, 22, 69], our stability result exhibits the optimal super-resolution ability of l_0 -minimization in solving such problems; (3) Inspired by the techniques used in the proofs, we propose a new coordinate-combination-based model order detection algorithm for two-dimensional DOA problems and demonstrate its optimal performance both theoretically and numerically, and (4) we also propose a new coordinate-combination-based MUSIC (states for MULTiple SIGNAL Classification) algorithm for super-resolving sources in two-dimensional DOA estimation. Our original algorithm enjoys certain advantages compared to the conventional DOA algo-

rithms. We also exhibit numerically the phase transition phenomenon of the algorithm, which demonstrates its excellent resolving capacity.

1.1 Existing Works on the Resolution Limit Problem

The first theory for quantifying the resolution limit was derived by Ernst Abbe [1, 72]. Since then, there have been various proposals for the resolution limit [33, 57, 61, 66], among which the famous and widely used ones are the Rayleigh limit [57] and the full width at half maximum (FWHM) [20]. However, these classical resolution limits neglect the effect of noise and hence are not mathematically rigorous [12, 21, 23]. From a mathematical perspective, there is no resolution limit when one has perfect access to the exact intensity profile of the diffraction images. Therefore, the resolution limit can only be rigorously set when taking into account the measurement noise or aberration to preclude perfect access to the diffraction images. Based on this understanding, many works were devoted to characterizing the dependence of the two-point resolution on the signal-to-noise ratio from the perspective of statistical inference [30, 31, 49, 50, 63–65]. These classical and semi-classical limits of two-point resolution have been well-studied and we refer the reader to [12, 17, 21, 41] for more detailed introductions.

For the resolution limit of superresolving multiple point sources, the problem becomes much more difficult due to the high degree of nonlinearity. To our knowledge, the first breakthrough was achieved by Donoho in 1992 [24]. He considered a grid setting where a discrete measure is supported on a lattice and regularized by a so-called Rayleigh index. The problem is to reconstruct the amplitudes of the grid points from their noisy Fourier data in $[-\Omega, \Omega]$ with Ω being the band limit. He derived both the lower and upper bounds for the minimax error in the amplitude reconstruction, emphasizing the importance of sparsity and signal-to-noise in super-resolution. But the estimate has not been improved until recent years. In recent years, due to the enormous development of super-resolution modalities in biological imaging [10, 29, 32, 59, 77] and the popularity of researches of super-resolution algorithms in applied mathematics [7, 11, 22, 25, 38, 40, 52, 53, 55, 70, 71], the inherent superresolving capacity of the imaging problem is drawing increasing interest and has been well-studied for the one-dimensional case. In [19], the authors considered resolving n -sparse point sources supported on a grid and improved the results of Donoho. They showed that, in the presence of noise with intensity σ , the minimax error in the amplitude recovery scales like $\text{SRF}^{2n-1}\sigma$ where $\text{SRF} := \frac{1}{\Delta\Omega}$ was called the super-resolution factor. The case of multi-clustered point sources was considered in [8, 37] and similar minimax error estimations were derived. In [4, 9], the authors considered the minimax error for recovering off-the-grid point sources. Based on an analysis of the “prony-type system”, they derived bounds for both amplitude and location reconstructions of the point sources. More precisely, they showed that for $\sigma \lesssim (\text{SRF})^{-2p+1}$, where p is the number of point sources in a cluster, the minimax error for the amplitude and the location recoveries scale respectively as $(\text{SRF})^{2p-1}\sigma$ and $(\text{SRF})^{2p-2}\sigma/\Omega$, while for the isolated non-clustered source, the corresponding minimax error for the amplitude and the location recoveries scale respectively as σ and σ/Ω . We also refer the reader to [12,

51] for understanding the resolution limit from the perspective of sample complexity and to [16, 69] for the resolving limit of some algorithms.

On the other hand, in order to characterize the exact resolution rather than the minimax error in recovering multiple point sources, in the earlier works [42–46] we have defined “computational resolution limits” which characterize the minimum required distance between point sources so that their number and locations can be stably resolved under certain noise level. By developing a nonlinear approximation theory in a so-called Vandermonde space, we have derived sharp bounds for computational resolution limits in the one-dimensional super-resolution problem. In particular, we have showed in [45] that the computational resolution limits for the number and location recoveries should be respectively $\frac{C_{\text{num}}}{\Omega} \left(\frac{\sigma}{m_{\text{min}}}\right)^{\frac{1}{2n-2}}$ and $\frac{C_{\text{supp}}}{\Omega} \left(\frac{\sigma}{m_{\text{min}}}\right)^{\frac{1}{2n-1}}$, where C_{num} and C_{supp} are constants and m_{min} is the minimum strength of the point sources. We have extended these estimates to multi-dimensional cases in [44] but the results are not that optimal due to the projection techniques used there. In this paper, we improve the estimates for the two-dimensional super-resolution problem by a new technique. The improvements shall be discussed in detail in Sect. 2.

1.2 Direction of Arrival Estimation

Our work also inspires new ideas for the two-dimensional direction of arrival estimation. Direction of arrival (DOA) estimation refers to the process of retrieving the direction information of several electromagnetic waves/sources from the received data of a number of antenna elements in a specific array. It is an important problem in array signal processing and finds wide applications in radar, sonar, wireless communications, etc; see, for instance, [6].

In one-dimensional DOA estimation, if the antenna elements are uniformly spaced in a line, the well-known MUSIC, ESPRIT algorithms, and other subspace methods can resolve the direction of each incident signal/source with high resolution. But for the two-dimensional DOA estimation with regular rectangular array (URA) where both azimuth and elevation angles should be determined, these subspace methods cannot be simply extended to the two-dimensional case to directly determine the azimuth and elevation angle of each source. A major idea to solve the two-dimensional DOA problem is to decompose it into two independent one-dimensional DOA estimations in which the subspace methods can be leveraged to efficiently restore the direction components of sources corresponding to x -axis and y -axis. We call the methods with this decoupling idea as one-dimensional-based algorithms throughout the paper for the convenience of discussion. It is worth emphasizing that other ways for directly obtaining the azimuth and elevation angles of each source were also considered [39, 78, 80], but the signal processing in a higher dimensional space reduced their computational efficiency.

Although the one-dimensional-based algorithms are usually much more computationally efficient, they still suffer from some issues: (1) the loss of distance separation for x -axis or y -axis components; (2) pair matching of the estimated elevation and azimuth angles. For the first issue, the x -axis (or y -axis) components of two sources may be closely spaced even though the two sources are far away in the two-dimensional

space. This causes very unstable reconstruction of the one-dimensional components and the sources. Most of the researches usually ignored these issues and some papers proposed different ways to enhance the reconstruction but the proposed methods are complicated [73, 74]. For example, in [74], the authors utilized Taylor expansion, subspace projection, and a tree structure to enhance the reconstruction when the recovered one-dimensional components are unstable. The second issue is that the pair matching of the estimated elevation and azimuth angles is very time consuming when dealing with multiple components of sources. It usually requires a complex process or two-dimensional search [18, 35, 48, 68, 80].

In this paper, we propose a new efficient one-dimensional-based algorithm for the two-dimensional DOA estimation which solves the above two issues in a simple way. First, our algorithm employs a new idea named coordinate-combination to avoid severe loss of distance separation between sources in certain region; see Sect. 5.4 for the detailed discussion. On the other hand, unlike conventional one-dimensional-based algorithms, the pair matching problem of our algorithm is a simple balanced assignment problem [54] which can be solved efficiently by many algorithms such as the Hungarian algorithm.

1.3 Organization of the Paper

The rest of the paper is organized in the following way. In Sect. 2, we present the main results on computational resolution limits for the number detection and the location recovery in the two-dimensional super-resolution problem. We also provide a stability result for a sparsity promoting algorithm. In Sect. 3, we prove the main results in Sect. 2 and discuss the generalization to higher dimensions. Inspired by the techniques in the proofs, in Sects. 4 and 5 we introduce respectively the coordinate-combination-based number detection and source recovery algorithms in two-dimensional DOA estimations. We also conduct numerical experiments to demonstrate their super-resolution capability. Section 6 presents a nonlinear approximation theory in Vandermonde space which is also a main part of the proof of our main results. Section 7 is devoted to some conclusions and future work. In the ‘‘Appendix’’, we prove a technical lemma.

2 Main Results

2.1 Model Setting

We consider the following model of a linear combination of point sources in a two-dimensional space:

$$\mu = \sum_{j=1}^n a_j \delta_{\mathbf{y}_j},$$

where δ denotes Dirac’s δ -distribution in \mathbb{R}^2 , $\mathbf{y}_j \in \mathbb{R}^2$, $1 \leq j \leq n$, which are the supports of the measure, represent the locations of the point sources and $a_j \in \mathbb{C}$, $1 \leq$

$j \leq n$, their amplitudes. We call that the measure μ is n -sparse if all a_j 's are nonzero. We denote the coordinates of each \mathbf{y}_j by $\mathbf{y}_{j,1}, \mathbf{y}_{j,2}$, and

$$m_{\min} = \min_{j=1, \dots, n} |a_j|, \quad D_{\min} = \min_{p \neq j} \|\mathbf{y}_p - \mathbf{y}_j\|_1. \tag{2.1}$$

We assume that the available measurement is the noisy Fourier data of μ in a bounded and continuous domain $[0, \Omega]^2$, that is,

$$\mathbf{Y}(\boldsymbol{\omega}) = \mathcal{F}\mu(\boldsymbol{\omega}) + \mathbf{W}(\boldsymbol{\omega}) = \sum_{j=1}^n a_j e^{i\mathbf{y}_j^\top \boldsymbol{\omega}} + \mathbf{W}(\boldsymbol{\omega}), \quad \boldsymbol{\omega} \in [0, \Omega]^2, \tag{2.2}$$

where $\mathcal{F}\mu$ denotes the Fourier transform of μ , Ω is the cutoff frequency, and \mathbf{W} is the noise. We assume that

$$|\mathbf{W}(\boldsymbol{\omega})| < \sigma, \quad \boldsymbol{\omega} \in [0, \Omega]^2,$$

where σ is the noise level. We remark that, throughout the paper, we will use bold symbols for vectors, matrices, and functions, and ordinary ones for scalar values. Especially, measurements \mathbf{Y} and noise \mathbf{W} are viewed as functions in all sections. Also, we will only use $\|\cdot\|_\infty, \|\cdot\|_1, \|\cdot\|_2$ for vectors.

We are interested in the resolution limit in superresolving a cluster of tightly spaced point sources. To be more specific, we denote by

$$B_{\delta, \infty}(\mathbf{x}) := \left\{ \mathbf{y} \mid \mathbf{y} \in \mathbb{R}^2, \|\mathbf{y} - \mathbf{x}\|_\infty < \delta \right\},$$

and assume that $\mathbf{y}_j \in B_{\frac{(n-1)\pi}{6\Omega}, \infty}(\mathbf{0}), j = 1, \dots, n$, or equivalently $\|\mathbf{y}_j\|_\infty < \frac{(n-1)\pi}{6\Omega}$. This assumption is because our techniques rely on the approximation theory in the Vandermonde space (Sect.6) and the stability is related to the distance between $e^{i\mathbf{y}_{j,1}\omega^*} + e^{i\mathbf{y}_{j,2}\omega^*}, j = 1, \dots, n$, for certain ω^* . Without this assumption, although the \mathbf{y}_j 's are well-separated on \mathbb{R} , the $e^{i\mathbf{y}_{j,1}\omega^*} + e^{i\mathbf{y}_{j,2}\omega^*}$'s may be very close. This is a common assumption when tackling the super-resolution problem, see for instance [9, 44, 45]. Since we are more interested in the case when \mathbf{y}_j 's are tightly spaced, this assumption is also reasonable.

2.2 Computational Resolution Limit for Number Detection in the Two-Dimensional Super-Resolution Problem

In this section, we estimate the super-resolving capacity of the source number detection in two-dimensional super-resolution problems. To be specific, we will define and characterize a computational resolution limit for the corresponding number detection problem. Our main results are built upon delicate analysis of the σ -admissible measure defined as follows:

Definition 2.1 Given a measurement \mathbf{Y} , we say that $\hat{\mu} = \sum_{j=1}^m \hat{a}_j \delta_{\hat{\mathbf{y}}_j}$, $\hat{\mathbf{y}}_j \in \mathbb{R}^2$ is a σ -admissible discrete measure of \mathbf{Y} if

$$|\mathcal{F}\hat{\mu}(\boldsymbol{\omega}) - \mathbf{Y}(\boldsymbol{\omega})| < \sigma, \text{ for all } \boldsymbol{\omega} \in [0, \Omega]^2.$$

Note that the set of σ -admissible measures of \mathbf{Y} characterizes all possible solutions to the inverse problem with the given measurement \mathbf{Y} . If all σ -admissible measures have at least n supports, then detecting the correct source number is possible, for example by targeting at the sparsest admissible measures. However, if there exists one σ -admissible measure with less than n supports, detecting the source number n is impossible without additional prior information. This leads to the following new definition of resolution limit, named computational resolution limit.

Definition 2.2 The computational resolution limit to the number detection problem in two dimensions is defined as the smallest nonnegative number $\mathcal{D}_{2,\text{num}}$ such that for all n -sparse measures $\sum_{j=1}^n a_j \delta_{\mathbf{y}_j}$, $\mathbf{y}_j \in B_{\frac{(n-1)\pi}{6\Omega}, \infty}(\mathbf{0})$ and the associated measurement \mathbf{Y} in (2.2), if

$$\min_{p \neq j} \|\mathbf{y}_j - \mathbf{y}_p\|_1 \geq \mathcal{D}_{2,\text{num}},$$

then there does not exist any σ -admissible measure of \mathbf{Y} with less than n supports.

The above resolution limit is termed ‘‘computational resolution limit’’ to distinguish it from the classic Rayleigh limit. Compared to the Rayleigh limit, the definition of the computational resolution limit is more rigorous from the mathematical perspective. It is related to the noise, by which it is more applicable for modern imaging techniques. In [44–46], the authors defined similar computational resolution limits and present rigorous estimations for them. Here by the following theorem, we derive an estimate to the $\mathcal{D}_{2,\text{num}}$, which substantially improves the estimate in [44] for the two-dimensional case.

Theorem 2.1 Let $n \geq 2$ and the measurement \mathbf{Y} in (2.2) be generated by an n -sparse measure $\mu = \sum_{j=1}^n a_j \delta_{\mathbf{y}_j}$, $\mathbf{y}_j \in B_{\frac{(n-1)\pi}{6\Omega}, \infty}(\mathbf{0})$. Assume that the following separation condition is satisfied

$$\min_{p \neq j, 1 \leq p, j \leq n} \|\mathbf{y}_p - \mathbf{y}_j\|_1 \geq \frac{16.6\pi(n-1)}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-2}} \tag{2.3}$$

with m_{\min} being defined in (2.1). Then there does not exist any σ -admissible measures of \mathbf{Y} with less than n supports.

Theorem 2.1 reveals that when $\min_{p \neq j, 1 \leq p, j \leq n} \|\mathbf{y}_p - \mathbf{y}_j\|_1 \geq \frac{16.6\pi(n-1)}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-2}}$, recovering exactly the source number n is possible. Compared with the Rayleigh limit $\frac{c_2\pi}{\Omega}$, where c_2 is a constant, Theorem 2.1 also indicates that resolving the source number in the sub-Rayleigh regime is theoretically possible if the SNR is sufficiently large.

Moreover, the estimate in Theorem 2.1 substantially improves the result in [44]. To be more specific, in [44], the authors considered recovering the source number and locations from the noisy measurement,

$$\mathbf{Y}(\boldsymbol{\omega}) = \mathcal{F}\mu(\boldsymbol{\omega}) + \mathbf{W}(\boldsymbol{\omega}) = \sum_{j=1}^n a_j e^{i\mathbf{y}_j^\top \boldsymbol{\omega}} + \mathbf{W}(\boldsymbol{\omega}), \quad \|\boldsymbol{\omega}\|_2 \leq \Omega,$$

where the noise $\mathbf{W}(\boldsymbol{\omega})$ is bounded as $|\mathbf{W}(\boldsymbol{\omega})| < \sigma, \|\boldsymbol{\omega}\|_2 \leq \Omega$. The upper bound estimation for the two-dimensional computational resolution limit in number detection is

$$\min_{p \neq j, 1 \leq p, j \leq n} \|\mathbf{y}_p - \mathbf{y}_j\|_2 \geq \frac{2.2e\pi n(n-1)}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-2}}. \tag{2.4}$$

Therefore, both the model and the norm of separation distances considered in [44] and our paper have no substantial difference that could make the problem in [44] inherently harder. But leveraging the new techniques here, the constant factor in the previous estimate (2.4) is improved to order n (Eq. (2.3)) now. In particular, by a simple calculation (taking into account the model difference), for $n \geq 2$, our new estimate already becomes better.

On the other hand, it is already known from [44] that the computational resolution limit for the number detection in the general k -dimensional super-resolution problem is bounded below by $\frac{C_1}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-2}}$ for some constant C_1 . Thus the $\mathcal{D}_{2,\text{num}}$ is bounded by

$$\frac{C_1}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-2}} \leq \mathcal{D}_{2,\text{num}} \leq \frac{C_2 n}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-2}}, \tag{2.5}$$

and our estimate for the upper bound is already very sharp.

The above estimates further indicate a phase transition phenomenon in the two-dimensional number detection problem. Specifically, by (2.5) we expect the presence of a line of slope $2n - 2$ in the parameter space $\log(\text{SRF}) - \log(\text{SNR})$ above which the source number can be correctly detected in each realization. This phenomenon is confirmed exactly by the number detection algorithm (Algorithm 2) later in Sect. 4.5 and illustrated in Fig. 2.

2.3 Computational Resolution Limit for Location Recovery in the Two-Dimensional Super-Resolution Problem

We next present our results on the resolution limit for the location recovery problem in two dimensions. We first introduce the following concept of δ -neighborhood of discrete measures. Define

$$B_{\delta,1}(\mathbf{x}) := \left\{ \mathbf{y} \mid \mathbf{y} \in \mathbb{R}^2, \|\mathbf{y} - \mathbf{x}\|_1 < \delta \right\}.$$

Definition 2.3 Let $\mu = \sum_{j=1}^n a_j \delta_{\mathbf{y}_j}$ be an n -sparse discrete measure in \mathbb{R}^2 and let $\delta > 0$ be such that the n balls $B_{\delta,1}(\mathbf{y}_j)$, $1 \leq j \leq n$ are pairwise disjoint. We say that $\hat{\mu} = \sum_{j=1}^n \hat{a}_j \delta_{\hat{\mathbf{y}}_j}$ is within δ -neighborhood of μ if each $\hat{\mathbf{y}}_j$ is contained in one and only one of the n balls $B_{\delta,1}(\mathbf{y}_j)$, $1 \leq j \leq n$.

According to the above definition, a measure $\hat{\mu}$ in a δ -neighborhood of μ preserves the inner structure of the collection of point sources. For a stable location (or support of measure) recovery algorithm, the output should be a measure in some δ -neighborhood of the underlying sources. Moreover, δ should tend to zero as the noise level σ tends to zero. We now introduce the computational resolution limit for the support recovery problem. For ease of exposition, we only consider measures supported in $B_{\frac{(2n-1)\pi}{12\Omega}, \infty}(\mathbf{0})$, where n is the source number.

Definition 2.4 The computational resolution limit in the two-dimensional location recovery problem is defined as the smallest non-negative number $\mathcal{D}_{2,\text{supp}}$ so that for any n -sparse measure $\mu = \sum_{j=1}^n a_j \delta_{\mathbf{y}_j}$, $\mathbf{y}_j \in B_{\frac{(2n-1)\pi}{12\Omega}, \infty}(\mathbf{0})$ and the associated measurement \mathbf{Y} in (2.2), if

$$\min_{p \neq j, 1 \leq p, j \leq n} \|\mathbf{y}_p - \mathbf{y}_j\|_1 \geq \mathcal{D}_{2,\text{supp}},$$

then there exists $\delta > 0$ such that any σ -admissible measure of \mathbf{Y} with n supports in $B_{\frac{(2n-1)\pi}{12\Omega}}(\mathbf{0})$ is within δ -neighborhood of μ .

We have the following estimate for the upper bound of $\mathcal{D}_{2,\text{supp}}$.

Theorem 2.2 Let $n \geq 2$. Let the measurement \mathbf{Y} in (2.2) be generated by an n -sparse measure $\mu = \sum_{j=1}^n a_j \delta_{\mathbf{y}_j}$, $\mathbf{y}_j \in B_{\frac{(2n-1)\pi}{12\Omega}, \infty}(\mathbf{0})$ in the two-dimensional space. Assume that

$$D_{\min} := \min_{p \neq j} \|\mathbf{y}_p - \mathbf{y}_j\|_1 \geq \frac{15.3\pi(n - \frac{1}{2})}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-1}}. \tag{2.6}$$

If $\hat{\mu} = \sum_{j=1}^n \hat{a}_j \delta_{\hat{\mathbf{y}}_j}$ supported on $B_{\frac{(2n-1)\pi}{12\Omega}, \infty}(\mathbf{0})$ is a σ -admissible measure of \mathbf{Y} , then $\hat{\mu}$ is in a $\frac{D_{\min}}{2}$ -neighborhood of μ . Moreover, after reordering the $\hat{\mathbf{y}}_j$'s, we have

$$\|\hat{\mathbf{y}}_j - \mathbf{y}_j\|_1 \leq \frac{C(n)}{\Omega} \text{SRF}^{2n-2} \frac{\sigma}{m_{\min}}, \quad 1 \leq j \leq n, \tag{2.7}$$

where $\text{SRF} := \frac{\pi}{D_{\min}\Omega}$ is the super-resolution factor and

$$C(n) = \frac{(1 + \sqrt{3})^{2n-1} 2^{5n-1} (2n - 1)^{2n-1} \pi}{3^{2n-0.5}}.$$

Theorem 2.2 demonstrates that when $\min_{p \neq j, 1 \leq p, j \leq n} \|\mathbf{y}_p - \mathbf{y}_j\|_1 \geq \frac{15.3\pi(n - \frac{1}{2})}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-1}}$, it is possible to recover stably the source locations. For sufficiently large SNR, the limit in Theorem 2.2 is less than the Rayleigh limit. This

indicates that super-resolution is possible for two-dimensional imaging problems. Moreover, for all $n \geq 2$, the estimate here is better than the one obtained in [44], which is

$$\frac{11.76e\pi n(n-1)}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-1}}.$$

Especially, the constant factor in our new estimate is of order n , while in the previous one it is of order n^2 .

It has been already shown in [44] that the computational resolution limit for the location recovery in the k -dimensional super-resolution problem is bounded below by $\frac{C_3}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-1}}$ for some constant C_3 . Thus the $\mathcal{D}_{2,\text{supp}}$ is bounded by

$$\frac{C_3}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-1}} \leq \mathcal{D}_{2,\text{supp}} \leq \frac{C_4 n}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-1}}, \tag{2.8}$$

and our estimate for the upper bound is sharp.

On the other hand, (2.8) indicates a phase transition in the location recovery problem. From (2.8) we expect that there exists a line of slope $2n - 1$ in the parameter space of $\log \text{SRF} - \log \text{SNR}$ such that the location recovery is stable in every point above the line. This is confirmed by Algorithm 4 in Sect. 5.4.2 and illustrated in Fig. 4.

Note that the bounds for the minimum separation distance to ensure a correct number detection scale like $\frac{1}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-2}}$, while for the minimum separation distance to guarantee a stable location recovery, they scale like $\frac{1}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-1}}$. How to understand the difference between the exponents in the estimates is also a very interesting question. One way is from the derivation of the lower bounds of the two resolution limits [43, 45]. In the analysis of the resolution limit in the number detection, we actually consider when the measurement from n sources can be approximated by the measurement generated by $n - 1$ sources, but in the location recovery, we consider when the measurement from n sources can be approximated by the measurement from another n sources. Thus for the number detection, we can use at most $2n - 1$ sources to construct the worst-case scenario, i.e., the measurement generated by specific $n - 1$ sources is close to the measurement from specific n sources. While for the location recovery, we can use $2n$ sources to construct the worst-case scenario, whence the location recovery has more freedom and becomes worse. This results in a difference between the exponents of the noise-to-signal ratio in both resolution estimates. The interested reader is referred to [43, 45] for detailed proofs of the lower bounds.

2.4 Stability of a Sparsity-Promoting Algorithm

Sparsity-promoting algorithms are popular methods in imaging processing and many other fields. By the results for resolution limit, we can derive a stability result for a l_0 -minimization in the two-dimensional super-resolution problems. We consider the

following l_0 -minimization problem:

$$\min_{\rho \text{ supported on } \mathcal{O}} \|\rho\|_0 \quad \text{subject to} \quad |\mathcal{F}\rho(\boldsymbol{\omega}) - \mathbf{Y}(\boldsymbol{\omega})| < \sigma, \quad \boldsymbol{\omega} \in [0, \Omega]^2, \quad (2.9)$$

where $\|\rho\|_0$ is the number of Dirac masses representing the discrete measure ρ . As a corollary of Theorems 2.1 and 2.2, we have the following stability result.

Theorem 2.3 *Let $n \geq 2$ and $\sigma \leq m_{\min}$. Let the measurement \mathbf{Y} in (2.2) be generated by an n -sparse measure $\mu = \sum_{j=1}^n a_j \delta_{\mathbf{y}_j}$, $\mathbf{y}_j \in B_{\frac{(n-1)\pi}{6\Omega}, \infty}(\mathbf{0})$ in the two-dimensional space. Assume that*

$$D_{\min} := \min_{p \neq j} \|\mathbf{y}_p - \mathbf{y}_j\|_1 \geq \frac{16.6\pi(n - \frac{1}{2})}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-1}}. \quad (2.10)$$

Let \mathcal{O} in the minimization problem (2.9) be $B_{\frac{(n-1)\pi}{6\Omega}, \infty}(\mathbf{0})$, then the solution to (2.9) contains exactly n point sources. For any solution $\hat{\mu} = \sum_{j=1}^n \hat{a}_j \delta_{\hat{\mathbf{y}}_j}$, it is in a $\frac{D_{\min}}{2}$ -neighborhood of μ . Moreover, after reordering the $\hat{\mathbf{y}}_j$'s, we have

$$\|\hat{\mathbf{y}}_j - \mathbf{y}_j\|_1 \leq \frac{C(n)}{\Omega} \text{SRF}^{2n-2} \frac{\sigma}{m_{\min}}, \quad 1 \leq j \leq n, \quad (2.11)$$

where $\text{SRF} := \frac{\pi}{D_{\min}\Omega}$ and

$$C(n) = \frac{(1 + \sqrt{3})^{2n-1} 2^{5n-1} (2n - 1)^{2n-1} \pi}{3^{2n-0.5}}.$$

Theorem 2.3 reveals that sparsity promoting over admissible solutions could resolve the source locations to the resolution limit level. It provides an insight that theoretically sparsity-promoting algorithms would have excellent performance on the two-dimensional super-resolution problems. Especially, under the separation condition (2.10), any tractable sparsity-promoting algorithms (such as total variation minimization algorithms [11]) rendering the sparsest solution could stably reconstruct all the source locations.

3 Proofs of the Main Results

The idea for proving the main results of the paper is to use some new techniques to reduce the two-dimensional super-resolution problem to an approximation problem of specific complex vectors, for which we develop a nonlinear approximation theory in Vandermonde space in Sect. 6. The reduction techniques are mainly based on three crucial observations in the following subsection.

3.1 Three Crucial Observations

We here introduce three crucial observations that reduce the two-dimensional super-resolution problem to its one-dimensional analog, by which we are able to derive the resolution limit theory of this paper. These ideas also inspire a new direction for the DOA algorithms; see Sects. 4 and 5.

Translation invariance By translation invariance we mean that if a measure $\hat{\mu} = \sum_{j=1}^q \hat{a}_j \delta_{\hat{y}_j}$ is a σ -admissible measure for the measurement \mathbf{Y} , then $\hat{\mu} = \sum_{j=1}^q \hat{a}_j \delta_{\hat{y}_j + \mathbf{v}}$ is a σ -admissible measure for the measurement $e^{i\mathbf{v}^\top \omega} \mathbf{Y}(\omega)$, for any vector $\mathbf{v} \in \mathbb{R}^2$. More precisely, we have

$$\left| \sum_{j=1}^q \hat{a}_j e^{i(\hat{y}_j + \mathbf{v})^\top \omega} - e^{i\mathbf{v}^\top \omega} \mathbf{Y}(\omega) \right| = \left| \sum_{j=1}^q \hat{a}_j e^{i\hat{y}_j^\top \omega} - \mathbf{Y}(\omega) \right| < \sigma, \quad \omega \in [0, \Omega]^2. \tag{3.1}$$

In addition, if for certain $\delta \geq 0$,

$$\left| \sum_{j=1}^q \hat{a}_j e^{i\hat{y}_j^\top \omega} - \sum_{j=1}^n a_j e^{i\mathbf{y}_j^\top \omega} \right| < \delta, \quad \omega \in [0, \Omega]^2, \tag{3.2}$$

then for any vector $\mathbf{v} \in \mathbb{R}^2$,

$$\left| \sum_{j=1}^q \hat{a}_j e^{i(\hat{y}_j + \mathbf{v})^\top \omega} - \sum_{j=1}^n a_j e^{i(\mathbf{y}_j + \mathbf{v})^\top \omega} \right| < \delta, \quad \omega \in [0, \Omega]^2.$$

Combination of coordinates The second observation is that if (3.2) holds, we have a similar estimate for the summation of combinations of $e^{i\tau \hat{y}_{j,1}}$, $e^{i\tau \hat{y}_{j,2}}$ and $e^{i\tau \mathbf{y}_{j,1}}$, $e^{i\tau \mathbf{y}_{j,2}}$ for certain τ . Specifically, we have the following lemma. Throughout the paper, we denote the combinatorial numbers by $\binom{t}{l_1}$ for certain $t \geq t_1$.

Lemma 3.1 *For any integer $t \geq 0$ and $\tau \leq \frac{\Omega}{t}$, the measurement constraint (3.2) implies*

$$\left| \sum_{j=1}^q \hat{a}_j (e^{i r_1} e^{i\tau \hat{y}_{j,1}} + e^{i r_2} e^{i\tau \hat{y}_{j,2}})^t - \sum_{j=1}^n a_j (e^{i r_1} e^{i\tau \mathbf{y}_{j,1}} + e^{i r_2} e^{i\tau \mathbf{y}_{j,2}})^t \right| < 2^t \delta, \quad r_1, r_2 \in \mathbb{R}.$$

Proof Let $\hat{d}_j = e^{ir_1} e^{i\tau\hat{y}_{j,1}} + e^{ir_2} e^{i\tau\hat{y}_{j,2}}$ and $d_j = e^{ir_1} e^{i\tau y_{j,1}} + e^{ir_2} e^{i\tau y_{j,2}}$. We have

$$\begin{aligned} & \left| \sum_{j=1}^q \hat{a}_j \hat{d}_j^t - \sum_{j=1}^n a_j d_j^t \right| = \left| \sum_{j=1}^q \hat{a}_j (e^{ir_1} e^{i\tau\hat{y}_{j,1}} + e^{ir_2} e^{i\tau\hat{y}_{j,2}})^t \right. \\ & \quad \left. - \sum_{j=1}^n a_j (e^{ir_1} e^{i\tau y_{j,1}} + e^{ir_2} e^{i\tau y_{j,2}})^t \right| \\ & = \left| \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} \binom{t}{t_1} \left(\sum_{j=1}^q \hat{a}_j e^{ir_1 t_1} e^{ir_2 t_2} e^{i\tau\hat{y}_{j,1} t_1} e^{i\tau\hat{y}_{j,2} t_2} \right. \right. \\ & \quad \left. \left. - \sum_{j=1}^n a_j e^{ir_1 t_1} e^{ir_2 t_2} e^{i\tau y_{j,1} t_1} e^{i\tau y_{j,2} t_2} \right) \right| \\ & \leq \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} \binom{t}{t_1} \left| \sum_{j=1}^q \hat{a}_j e^{i\tau\hat{y}_{j,1} t_1} e^{i\tau\hat{y}_{j,2} t_2} - \sum_{j=1}^n a_j e^{i\tau y_{j,1} t_1} e^{i\tau y_{j,2} t_2} \right| \\ & = \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} \binom{t}{t_1} \left| \sum_{j=1}^q \hat{a}_j e^{i(t_1 \tau, t_2 \tau)\hat{y}_j} - \sum_{j=1}^n a_j e^{i(t_1 \tau, t_2 \tau)y_j} \right| \\ & < \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} \binom{t}{t_1} \delta \quad \left(\text{by } \tau \leq \frac{\Omega}{t} \text{ and (3.2)} \right) \\ & = 2^t \delta. \end{aligned}$$

□

This is the key observation of the paper. It reduces the two-dimensional super-resolution problem to nearly a one-dimensional super-resolution one. Since it is about the difference between summation of combinations of $e^{i\tau\hat{y}_{j,1}}$, $e^{i\tau\hat{y}_{j,2}}$ and $e^{i\tau y_{j,1}}$, $e^{i\tau y_{j,2}}$, we refer to this observation as combination of coordinates and call the elements $e^{i\tau y_{j,1}} + e^{i\tau y_{j,2}}$ coordinate-combined elements. This coordinate-combination technique will be used in deriving new algorithms for the DOA problem in Sects. 4 and 5.

Compared to the projection techniques in [12, 44] which utilize the measurement constraint only in several one-dimensional spaces to derive stability results, our formulation utilizes more measurement constraints and consequently yields better estimates.

Preservation of the separation distance for the coordinate-combined elements The last observation is that, for θ_j 's in $[0, \frac{2\pi}{3}]^2$, the coordinate-combined elements $e^{i\theta_{j,1}} + e^{i\theta_{j,2}}$ still preserve the separation distance between the θ_j 's. This is revealed by Lemma 3.2. Note that the projection trick in [12, 44] and many conventional two-dimensional DOA algorithms do not preserve the separation distance between the original source. This causes many issues in the reconstruction and resolution estimation. Lemma 3.2 is the

main result of this paper by which we can overcome the above issues and hence find a new way to solve two-dimensional DOA problems.

Lemma 3.2 For two different vectors $\theta_j \in [0, \frac{2\pi}{3}]^2, j = 1, 2$ with $\frac{\pi}{3} \leq \theta_{j,2} - \theta_{j,1} \leq \frac{2}{3}\pi, j = 1, 2$, if $\|\theta_1 - \theta_2\|_1 \geq \Delta$, then

$$\left| e^{i\theta_{1,1}} + e^{i\theta_{1,2}} - (e^{i\theta_{2,1}} + e^{i\theta_{2,2}}) \right| \geq \frac{3}{2\pi} \Delta.$$

Proof Note that $0 \leq \theta_{j,1} < \theta_{j,2} \leq \frac{2\pi}{3}, j = 1, 2$. We prove the lemma by considering the following two cases.

Case 1 $0 \leq \theta_{1,1} \leq \theta_{2,1} < \theta_{2,2} \leq \theta_{1,2} \leq \frac{2\pi}{3}$.

In this case,

$$\begin{aligned} \left| e^{i\theta_{1,1}} + e^{i\theta_{1,2}} - (e^{i\theta_{2,1}} + e^{i\theta_{2,2}}) \right| &\geq \left| e^{i\theta_{2,1}} + e^{i\theta_{2,2}} \right| - \left| e^{i\theta_{1,1}} + e^{i\theta_{1,2}} \right| \\ &= 2 \left(\cos\left(\frac{\phi_2}{2}\right) - \cos\left(\frac{\phi_1}{2}\right) \right), \end{aligned}$$

where $\phi_j = \theta_{j,2} - \theta_{j,1}, j = 1, 2$. By the assumption made in the lemma, we have $\Delta \leq \phi_1 - \phi_2 \leq \frac{\pi}{3}$. Note also that $\frac{\pi}{6} \leq \frac{\phi_1 + \phi_2}{4} \leq \frac{\pi}{3}$. Thus

$$\begin{aligned} 2 \left(\cos\left(\frac{\phi_2}{2}\right) - \cos\left(\frac{\phi_1}{2}\right) \right) &= 4 \sin\left(\frac{\phi_1 + \phi_2}{4}\right) \sin\left(\frac{\phi_1 - \phi_2}{4}\right) \\ &\geq 4 \sin\left(\frac{\pi}{6}\right) \sin\left(\frac{\Delta}{4}\right) \geq \frac{3\Delta}{2\pi}, \end{aligned}$$

where the last inequality uses $\sin(\frac{\Delta}{4}) \geq \frac{3}{\pi} \frac{\Delta}{4}$ for $0 < \frac{\Delta}{4} \leq \frac{\pi}{4}$.

Case 2 $0 \leq \theta_{1,1} \leq \theta_{2,1} \leq \theta_{1,2} \leq \theta_{2,2} \leq \frac{2\pi}{3}$.

The idea is to calculate the angle between $e^{i\theta_{1,1}} + e^{i\theta_{1,2}}$ and $e^{i\theta_{2,1}} + e^{i\theta_{2,2}}$. By simple analysis of the angle relations between $e^{i\theta_{1,1}}, e^{i\theta_{1,2}}, e^{i\theta_{2,1}}, e^{i\theta_{2,2}}$, we obtain that the angle between $e^{i\theta_{1,1}} + e^{i\theta_{1,2}}$ and $e^{i\theta_{2,1}} + e^{i\theta_{2,2}}$ is $\frac{\theta_{2,1} - \theta_{1,1} + \theta_{2,2} - \theta_{1,2}}{2}$, which is larger than $\frac{\Delta}{2}$. Thus

$$\left| e^{i\theta_{1,1}} + e^{i\theta_{1,2}} - (e^{i\theta_{2,1}} + e^{i\theta_{2,2}}) \right| \geq \max \left(\left| e^{i\theta_{1,1}} + e^{i\theta_{1,2}} \right|, \left| e^{i\theta_{2,1}} + e^{i\theta_{2,2}} \right| \right) \sin\left(\frac{\Delta}{2}\right).$$

Since $\frac{\pi}{3} \leq \theta_{j,2} - \theta_{j,1} \leq \frac{2}{3}\pi, j = 1, 2$, we have

$$\max \left(\left| e^{i\theta_{1,1}} + e^{i\theta_{1,2}} \right|, \left| e^{i\theta_{2,1}} + e^{i\theta_{2,2}} \right| \right) \geq 1.$$

Therefore,

$$\left| e^{i\theta_{1,1}} + e^{i\theta_{1,2}} - (e^{i\theta_{2,1}} + e^{i\theta_{2,2}}) \right| \geq \sin\left(\frac{\Delta}{2}\right) \geq \frac{3\Delta}{2\pi},$$

where the last inequality uses $\sin(\frac{\Delta}{2}) \geq \frac{3}{\pi} \frac{\Delta}{2}$ for $0 < \frac{\Delta}{2} \leq \frac{\pi}{6}$. □

3.2 Proof of Theorem 2.1

Proof The proof of this theorem is by contradiction. Suppose that there exists a measure $\hat{\mu} = \sum_{j=1}^q \hat{a}_j \delta_{\hat{y}_j}$ with $q < n$ which is a σ -admissible measure of \mathbf{Y} . Then, by the measurement constraint (2.2) and $|\mathbf{W}(\boldsymbol{\omega})| < \sigma$, we have

$$\left| \sum_{j=1}^q \hat{a}_j e^{i\hat{y}_j^\top \boldsymbol{\omega}} - \sum_{j=1}^n a_j e^{i\mathbf{y}_j^\top \boldsymbol{\omega}} \right| < 2\sigma, \quad \boldsymbol{\omega} \in [0, \Omega]^2. \tag{3.3}$$

Since $\mathbf{y}_j \in [-\lambda, \lambda]^2$ with $\lambda = \frac{(n-1)\pi}{6\Omega}$, by letting $\mathbf{v} = (0, 6\lambda)^\top$, we obtain

$$\mathbf{y}_j + \mathbf{v} \in [-\lambda, \lambda] \times [5\lambda, 7\lambda]. \tag{3.4}$$

On the other hand, by (3.3) we also get

$$\left| \sum_{j=1}^q \hat{a}_j e^{i(\hat{y}_j + \mathbf{v})^\top \boldsymbol{\omega}} - \sum_{j=1}^n a_j e^{i(\mathbf{y}_j + \mathbf{v})^\top \boldsymbol{\omega}} \right| < 2\sigma, \quad \boldsymbol{\omega} \in [0, \Omega]^2.$$

Thus with a slight abuse of notation, we still denote those $\hat{y}_j + \mathbf{v}$ and $\mathbf{y}_j + \mathbf{v}$ by $\hat{\mathbf{y}}_j$, \mathbf{y}_j respectively and consider them in the rest of the proof. Note that we have

$$\mathbf{y}_j \in [-\lambda, \lambda] \times [5\lambda, 7\lambda], \quad j = 1, \dots, n.$$

Let $\tau = \frac{\Omega}{2(n-1)}$, together with $\lambda = \frac{(n-1)\pi}{6\Omega}$, we have $\tau \mathbf{y}_j \in [-\frac{\pi}{12}, \frac{\pi}{12}] \times [\frac{5\pi}{12}, \frac{7\pi}{12}]$. This yields

$$-\frac{\pi}{12} \leq \tau \mathbf{y}_{j,1} \leq \frac{\pi}{12}, \quad \frac{5\pi}{12} \leq \tau \mathbf{y}_{j,2} \leq \frac{7\pi}{12}, \quad \frac{\pi}{3} \leq \tau \mathbf{y}_{j,2} - \tau \mathbf{y}_{j,1} \leq \frac{2\pi}{3}. \tag{3.5}$$

On the other hand, let $\hat{d}_j = e^{i\tau \hat{\mathbf{y}}_{j,1}} + e^{i\tau \hat{\mathbf{y}}_{j,2}}$ and $d_j = e^{i\tau \mathbf{y}_{j,1}} + e^{i\tau \mathbf{y}_{j,2}}$. By Lemma 3.1 and (3.3) we have that

$$\left| \sum_{j=1}^q \hat{a}_j \hat{d}_j^t - \sum_{j=1}^n a_j d_j^t \right| < 2^{t+1} \sigma, \quad t = 0, 1, \dots, 2n - 2. \tag{3.6}$$

Let

$$\mathbf{b} = \left(\sum_{j=1}^q \hat{a}_j \hat{d}_j^0 - \sum_{j=1}^n a_j d_j^0, \sum_{j=1}^q \hat{a}_j \hat{d}_j^1 - \sum_{j=1}^n a_j d_j^1, \dots, \sum_{j=1}^q \hat{a}_j \hat{d}_j^{2n-2} - \sum_{j=1}^n a_j d_j^{2n-2} \right)^\top.$$

Since (3.5) holds, Lemma 3.2 yields

$$\begin{aligned} d_{\min} &:= \min_{p \neq q} |d_p - d_q| \geq \frac{3}{2\pi} \min_{p \neq q} \tau \|\mathbf{y}_p - \mathbf{y}_q\|_1 > 12.4 \left(\frac{\sigma}{m_{\min}} \right)^{\frac{1}{2n-2}} \\ &> 2\sqrt{6(1 + \sqrt{3})} \left(\frac{4}{\sqrt{3}} \frac{\sigma}{m_{\min}} \right)^{\frac{1}{2n-2}}, \end{aligned}$$

where the second to the last inequality is due to the separation condition (2.3). On the other hand, we have $|\hat{d}_p| \leq 2, p = 1, \dots, q$ and $|d_j| \leq \sqrt{3}, j = 1, \dots, n$ since (3.5) holds. Thus we can apply Theorem 6.2 and get

$$\|\mathbf{b}\|_2 \geq \frac{m_{\min} (d_{\min})^{2n-2}}{(2(1 + 2)(1 + \sqrt{3}))^{(n-1)}} > \frac{4^n \sigma}{\sqrt{3}}.$$

Note that the results and proofs of Theorem 6.2 are presented in Sect. 6, but the readers can first focus only on the results before going through their detailed proofs. However, (3.6) implies that $\|\mathbf{b}\|_2 < \frac{4^n \sigma}{\sqrt{3}}$, which is a contradiction. This proves the theorem. \square

3.3 Proof of Theorem 2.2

Proof Note that $\mathbf{y}_j, \hat{\mathbf{y}}_j$'s are in $[-\lambda, \lambda]^2$ with $\lambda = \frac{(2n-1)\pi}{12\Omega}$ and $\hat{\mu} = \sum_{j=1}^n \hat{a}_j \delta_{\hat{\mathbf{y}}_j}$ is a σ -admissible measure of \mathbf{Y} . Let $\tau = \frac{\Omega}{2n-1}$. Similarly to the proof in the above section, we can construct $\mathbf{x}_j = \mathbf{y}_j + \mathbf{v}, \hat{\mathbf{x}}_j = \hat{\mathbf{y}}_j + \mathbf{v}$ so that $\tau \hat{\mathbf{x}}_j, \tau \mathbf{x}_j \in [-\frac{\pi}{12}, \frac{\pi}{12}] \times [\frac{5\pi}{12}, \frac{7\pi}{12}]$ and

$$\left| \sum_{j=1}^n \hat{a}_j e^{i\hat{\mathbf{x}}_j^\top \boldsymbol{\omega}} - \sum_{j=1}^n a_j e^{i\mathbf{x}_j^\top \boldsymbol{\omega}} \right| < 2\sigma, \quad \boldsymbol{\omega} \in [0, \Omega]^2. \tag{3.7}$$

Thus we have

$$-\frac{\pi}{12} \leq \tau \mathbf{x}_{j,1} \leq \frac{\pi}{12}, \quad \frac{5\pi}{12} \leq \tau \mathbf{x}_{j,2} \leq \frac{7\pi}{12}, \quad \frac{\pi}{3} \leq \tau \mathbf{x}_{j,2} - \tau \mathbf{x}_{j,1} \leq \frac{2\pi}{3}, \tag{3.8}$$

$$-\frac{\pi}{12} \leq \tau \hat{\mathbf{x}}_{j,1} \leq \frac{\pi}{12}, \quad \frac{5\pi}{12} \leq \tau \hat{\mathbf{x}}_{j,2} \leq \frac{7\pi}{12}, \quad \frac{\pi}{3} \leq \tau \hat{\mathbf{x}}_{j,2} - \tau \hat{\mathbf{x}}_{j,1} \leq \frac{2\pi}{3}. \tag{3.9}$$

Moreover, it follows that

$$\begin{aligned}
 -\frac{\pi}{12} \leq \tau \mathbf{x}_{j,1} \leq \frac{\pi}{12}, \quad \frac{-7\pi}{12} \leq \tau \mathbf{x}_{j,2} - \pi \leq \frac{-5\pi}{12}, \quad \frac{\pi}{3} \\
 \leq \tau \mathbf{x}_{j,1} - (\tau \mathbf{x}_{j,2} - \pi) \leq \frac{2\pi}{3},
 \end{aligned} \tag{3.10}$$

$$\begin{aligned}
 -\frac{\pi}{12} \leq \tau \hat{\mathbf{x}}_{j,1} \leq \frac{\pi}{12}, \quad \frac{-7\pi}{12} \leq \tau \hat{\mathbf{x}}_{j,2} - \pi \leq \frac{-5\pi}{12}, \quad \frac{\pi}{3} \\
 \leq \tau \hat{\mathbf{x}}_{j,1} - (\tau \hat{\mathbf{x}}_{j,2} - \pi) \leq \frac{2\pi}{3}.
 \end{aligned} \tag{3.11}$$

Let $\hat{d}_j = e^{i\tau \hat{\mathbf{x}}_{j,1}} + e^{i\tau \hat{\mathbf{x}}_{j,2}}$, $d_j = e^{i\tau \mathbf{x}_{j,1}} + e^{i\tau \mathbf{x}_{j,2}}$ and $\hat{g}_j = e^{i(\tau \hat{\mathbf{x}}_{j,1} + e^{i(\tau \hat{\mathbf{x}}_{j,2} - \pi)})}$, $g_j = e^{i\tau \mathbf{x}_{j,1}} + e^{i(\tau \mathbf{x}_{j,2} - \pi)}$. By (3.7) and Lemma 3.1, we arrive at

$$\left| \sum_{j=1}^n \hat{a}_j \hat{d}_j^t - \sum_{j=1}^n a_j d_j^t \right| < 2^{t+1} \sigma, \quad t = 0, 1, \dots, 2n - 1, \tag{3.12}$$

$$\left| \sum_{j=1}^n \hat{a}_j \hat{g}_j^t - \sum_{j=1}^n a_j g_j^t \right| < 2^{t+1} \sigma, \quad t = 0, 1, \dots, 2n - 1. \tag{3.13}$$

Let

$$\mathbf{d} = \left(\sum_{j=1}^n \hat{a}_j \hat{d}_j^0 - \sum_{j=1}^n a_j d_j^0, \sum_{j=1}^n \hat{a}_j \hat{d}_j^1 - \sum_{j=1}^n a_j d_j^1, \dots, \sum_{j=1}^n \hat{a}_j \hat{d}_j^{2n-1} - \sum_{j=1}^n a_j d_j^{2n-1} \right)^\top,$$

and

$$\mathbf{g} = \left(\sum_{j=1}^n \hat{a}_j \hat{g}_j^0 - \sum_{j=1}^n a_j g_j^0, \sum_{j=1}^n \hat{a}_j \hat{g}_j^1 - \sum_{j=1}^n a_j g_j^1, \dots, \sum_{j=1}^n \hat{a}_j \hat{g}_j^{2n-1} - \sum_{j=1}^n a_j g_j^{2n-1} \right)^\top.$$

Equations (3.12) and (3.13) imply respectively

$$\|\mathbf{d}\|_2 < \frac{2^{2n+1} \sigma}{\sqrt{3}}, \quad \|\mathbf{g}\|_2 < \frac{2^{2n+1} \sigma}{\sqrt{3}}.$$

Note also that by (3.8), (3.9), (3.10), and (3.11), we get

$$|\hat{d}_j|, |d_j|, |\hat{g}_j|, |g_j| \leq \sqrt{3}, \quad j = 1, \dots, n.$$

Define $d_{\min} := \min_{p \neq q} |d_p - d_q|$ and $g_{\min} := \min_{p \neq q} |g_p - g_q|$. Applying Theorem 6.2, we thus have that

$$\left\| \eta_{n,n}(d_1, \dots, d_n, \hat{d}_1, \dots, \hat{d}_n) \right\|_{\infty} < \frac{(1 + \sqrt{3})^{2n-1} 2^{2n+1} \sigma}{d_{\min}^{n-1} \sqrt{3} m_{\min}}, \quad (3.14)$$

and

$$\left\| \eta_{n,n}(g_1, \dots, g_n, \hat{g}_1, \dots, \hat{g}_n) \right\|_{\infty} < \frac{(1 + \sqrt{3})^{2n-1} 2^{2n+1} \sigma}{g_{\min}^{n-1} \sqrt{3} m_{\min}}, \quad (3.15)$$

where $\eta_{n,n}(\dots)$'s are vectors defined as in (6.10). We now demonstrate that we can reorder \hat{d}_j, \hat{g}_j to have $|\hat{d}_j - d_j| < \frac{d_{\min}}{2}$ and $|\hat{g}_j - g_j| < \frac{g_{\min}}{2}$, $j = 1, \dots, n$. First, since (3.8) and (3.10) hold, by Lemma 3.2 we have

$$d_{\min} \geq \frac{3}{2\pi} \min_{p \neq q} \tau \left\| \mathbf{y}_p - \mathbf{y}_q \right\|_1 \geq 11.475 \left(\frac{\sigma}{m_{\min}} \right)^{\frac{1}{2n-1}} > 2^{3/2} (1 + \sqrt{3}) \left(\frac{2^{5/2} \sigma}{\sqrt{3} m_{\min}} \right)^{\frac{1}{2n-1}}, \quad (3.16)$$

and

$$g_{\min} \geq \frac{3}{2\pi} \min_{p \neq q} \tau \left\| \mathbf{y}_p - \mathbf{y}_q \right\|_1 \geq 11.475 \left(\frac{\sigma}{m_{\min}} \right)^{\frac{1}{2n-1}} > 2^{3/2} (1 + \sqrt{3}) \left(\frac{2^{5/2} \sigma}{\sqrt{3} m_{\min}} \right)^{\frac{1}{2n-1}}, \quad (3.17)$$

where we also use separation condition (2.6) in the above derivation. Let

$$\epsilon_d = \frac{(1 + \sqrt{3})^{2n-1} 2^{2n+1} \sigma}{d_{\min}^{n-1} \sqrt{3} m_{\min}}, \quad \epsilon_g = \frac{(1 + \sqrt{3})^{2n-1} 2^{2n+1} \sigma}{g_{\min}^{n-1} \sqrt{3} m_{\min}}.$$

By (3.16),

$$d_{\min}^{2n-1} \geq \frac{(1 + \sqrt{3})^{2n-1} 2^{3n+1} \sigma}{\sqrt{3} m_{\min}}, \quad \text{or equivalently, } d_{\min}^n \geq 2^n \epsilon_d,$$

and by (3.17),

$$g_{\min}^{2n-1} \geq \frac{(1 + \sqrt{3})^{2n-1} 2^{3n+1} \sigma}{\sqrt{3} m_{\min}}, \quad \text{or equivalently, } g_{\min}^n \geq 2^n \epsilon_g.$$

Thus the conditions of Lemma 6.8 are satisfied. By Lemma 6.8, we have that after reordering \hat{d}_j, \hat{g}_j ,

$$\left| \hat{d}_j - d_j \right| < \frac{d_{\min}}{2}, \quad \left| \hat{g}_j - g_j \right| < \frac{g_{\min}}{2},$$

and

$$\begin{aligned} |\hat{d}_j - d_j| &\leq \left(\frac{2}{d_{\min}}\right)^{n-1} \epsilon_d = \left(\frac{1}{d_{\min}}\right)^{2n-2} \frac{(1 + \sqrt{3})^{2n-1} 2^{3n}}{\sqrt{3}} \frac{\sigma}{m_{\min}}, \\ |\hat{g}_j - g_j| &\leq \left(\frac{2}{g_{\min}}\right)^{n-1} \epsilon_g = \left(\frac{1}{g_{\min}}\right)^{2n-2} \frac{(1 + \sqrt{3})^{2n-1} 2^{3n}}{\sqrt{3}} \frac{\sigma}{m_{\min}}. \end{aligned}$$

Observing

$$\begin{aligned} e^{i\tau\hat{x}_{j,1}} - e^{i\tau x_{j,1}} &= \frac{1}{2}(\hat{d}_j - d_j + \hat{g}_j - g_j), \\ e^{i\tau\hat{x}_{j,2}} - e^{i\tau x_{j,2}} &= \frac{1}{2}(\hat{d}_j - d_j - (\hat{g}_j - g_j)), \end{aligned} \tag{3.18}$$

we conclude that

$$\begin{aligned} \left| e^{i\tau\hat{x}_{j,1}} - e^{i\tau x_{j,1}} \right| + \left| e^{i\tau\hat{x}_{j,2}} - e^{i\tau x_{j,2}} \right| &\leq \left(\left(\frac{1}{d_{\min}}\right)^{2n-2} + \left(\frac{1}{g_{\min}}\right)^{2n-2} \right) \\ &\quad \frac{(1 + \sqrt{3})^{2n-1} 2^{3n}}{\sqrt{3}} \frac{\sigma}{m_{\min}}. \end{aligned} \tag{3.19}$$

On the other hand, by (3.8) and (3.9),

$$|\hat{x}_{j,1} - x_{j,1}| \leq \frac{\pi}{6} \quad \text{and} \quad |\hat{x}_{j,2} - x_{j,2}| \leq \frac{\pi}{6}.$$

We further have

$$\begin{aligned} \tau \left| \hat{x}_{j,1} - x_{j,1} \right| + \tau \left| \hat{x}_{j,2} - x_{j,2} \right| &\leq \frac{\pi}{3} \left(\left| e^{i\hat{x}_{j,1}} - e^{ix_{j,1}} \right| + \left| e^{i\hat{x}_{j,2}} - e^{ix_{j,2}} \right| \right) \\ &\leq \left(\left(\frac{1}{d_{\min}}\right)^{2n-2} + \left(\frac{1}{g_{\min}}\right)^{2n-2} \right) \frac{(1 + \sqrt{3})^{2n-1} 2^{3n} \pi}{3\sqrt{3}} \frac{\sigma}{m_{\min}}. \end{aligned}$$

Recalling that $\tau = \frac{\Omega}{2^{n-1}}$, we have

$$\begin{aligned} \left| \hat{x}_{j,1} - x_{j,1} \right| + \left| \hat{x}_{j,2} - x_{j,2} \right| &\leq \frac{2n-1}{\Omega} \left(\left(\frac{1}{d_{\min}}\right)^{2n-2} + \left(\frac{1}{g_{\min}}\right)^{2n-2} \right) \\ &\quad \frac{(1 + \sqrt{3})^{2n-1} 2^{3n} \pi}{3\sqrt{3}} \frac{\sigma}{m_{\min}}. \end{aligned}$$

Note that by (3.16), we obtain that

$$D_{\min} \leq \frac{2\pi(2n-1)}{3\Omega} d_{\min} \quad \text{and} \quad D_{\min} \leq \frac{2\pi(2n-1)}{3\Omega} g_{\min}.$$

Thus

$$\begin{aligned} \|\hat{\mathbf{x}}_j - \mathbf{x}_j\|_1 &\leq \frac{(1 + \sqrt{3})^{2n-1} 2^{3n+1} \pi (2n-1)}{3\sqrt{3}\Omega} \left(\frac{2(2n-1)}{3}\right)^{2n-2} \left(\frac{\pi}{\Omega D_{\min}}\right)^{2n-2} \frac{\sigma}{m_{\min}} \\ &= \frac{(1 + \sqrt{3})^{2n-1} 2^{5n-1} (2n-1)^{2n-1} \pi}{3^{2n-0.5}\Omega} \left(\frac{\pi}{\Omega D_{\min}}\right)^{2n-2} \frac{\sigma}{m_{\min}}. \end{aligned}$$

Since $\|\hat{\mathbf{y}}_j - \mathbf{y}_j\|_1 = \|\hat{\mathbf{x}}_j - \mathbf{x}_j\|_1$, we further get

$$\|\hat{\mathbf{y}}_j - \mathbf{y}_j\|_1 \leq \frac{(1 + \sqrt{3})^{2n-1} 2^{5n-1} (2n-1)^{2n-1} \pi}{3^{2n-0.5}\Omega} \left(\frac{\pi}{\Omega D_{\min}}\right)^{2n-2} \frac{\sigma}{m_{\min}}.$$

Since $D_{\min} \geq \frac{15.3\pi(n-0.5)}{\Omega} \left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-1}}$, together with the above estimate, we can also show that

$$\|\hat{\mathbf{y}}_j - \mathbf{y}_j\|_1 < \frac{D_{\min}}{2}.$$

This completes the proof. \square

3.4 Discussion on the Generalization to Higher Dimensions

The techniques in this paper seem to have the potential to improve the estimates of resolution limits in higher dimensions in [44]. First, it is not difficult to see that translation invariance and combination of coordinates discussed in Sect. 3.1 can be generalized to any k -dimensional space. After that, the k -dimensional super-resolution problems can be transformed into the following approximation:

$$\left| \sum_{j=1}^q \hat{a}_j \hat{d}_j^t - \sum_{j=1}^n a_j d_j^t \right| < C\sigma, \quad t = 0, 1, \dots, s(n), \quad (3.20)$$

where $\hat{a}_j \in \mathbb{C}$, $\hat{d}_j \in \mathbb{C}$, $a_j \in \mathbb{C}$, $d_j \in \mathbb{C}$ and $s(n)$ are $2n-2$ or $2n-1$. Note that d_j, \hat{d}_j 's are from the combination of coordinates as in the definition in Sect. 3.2 (after Eq. (3.5)). Problem (3.20) is still what is considered in Sect. 6 and the results there can be directly applied. The only requirement now is that the separation distance between the d_j 's should be large enough to ensure that we can detect the correct source number or stably recover the d_j 's from constraint (3.20). Thus, the main obstacle is to generalize Lemma 3.2 to the k -dimensional space, which says that when the sources \mathbf{y}_j 's are well-separated, the d_j 's should also preserve the separation distance to some extent. This will help to generalize Theorems 2.1 and 2.2 to the k -dimensional space and improve significantly the estimates of resolution limits in higher dimensions.

However, a direct generalization of Lemma 3.2 is already hard by the current method where we should analyze the geometry of the problem case by case. In the future, we

seek to find new ways to get around this delicate issue and solve the problem in all dimensional spaces in a unified way.

On the other hand, the problem in k -dimensional space is actually more than just a direct generalization of Lemma 3.2. In particular, in order to analyze the stability of the location recovery in the k -dimensional space, we must not only prove that

$$\left| e^{i\theta_{1,1}} + e^{i\theta_{1,2}} + \dots + e^{i\theta_{1,k}} - \left(e^{i\theta_{2,1}} + e^{i\theta_{2,2}} + \dots + e^{i\theta_{2,k}} \right) \right| \geq c(k)\Delta$$

holds for some constant $c(k)$ when $\|\theta_1 - \theta_2\|_1 \geq \Delta$, but also show that

$$\begin{aligned} \left| e^{i\theta_{1,1}} - e^{i\theta_{1,2}} + \dots + e^{i\theta_{1,k}} - \left(e^{i\theta_{2,1}} - e^{i\theta_{2,2}} + \dots + e^{i\theta_{2,k}} \right) \right| &\geq c(k)\Delta, \\ \left| e^{i\theta_{1,1}} + e^{i\theta_{1,2}} - \dots + e^{i\theta_{1,k}} - \left(e^{i\theta_{2,1}} + e^{i\theta_{2,2}} - \dots + e^{i\theta_{2,k}} \right) \right| &\geq c(k)\Delta, \\ &\vdots \\ \left| e^{i\theta_{1,1}} + e^{i\theta_{1,2}} + \dots - e^{i\theta_{1,k}} - \left(e^{i\theta_{2,1}} + e^{i\theta_{2,2}} + \dots - e^{i\theta_{2,k}} \right) \right| &\geq c(k)\Delta \end{aligned}$$

under the same condition. This is because we need to construct more complex values like d_j, g_j (defined after Eq. (3.11)) and use the above results to obtain estimates such as (3.16) and (3.17) in order to get the stability of the reconstruction of each coordinate by treatments like (3.18) and (3.19). Therefore, the generalization of our results to higher dimensions is quite complicated. This definitely deserves further researches. It may inspire new algorithms in solving imaging and DOA problems in k -dimensional spaces. As discussed, the generalization and treatment in the k -dimensional space are still unclear, so we can not comment further on the extensions of our algorithms (proposed in Sects. 4 and 5) in higher dimensions.

4 An Algorithm for the Model Order Detection in Two-Dimensional DOA Estimation

In this section, based on the observations made in Sect. 3.1, we propose a new algorithm, named coordinate-combination-based sweeping singular-value-thresholding number detection algorithm, for the model order detection in two-dimensional DOA estimations.

4.1 Problem Formulation

The existing two-dimensional DOA algorithms usually try to estimate the azimuth and elevation angles (θ_j, ϕ_j) 's that are shown in Fig. 1. More precisely, we consider n narrowband signals/sources impinging on an $(\Omega + 1) \times (\Omega + 1)$ uniform rectangular array (URA) with $(\Omega + 1)^2$ well calibrated and identically polarized antenna elements. The signal received by these antenna elements in a single snapshot can be expressed

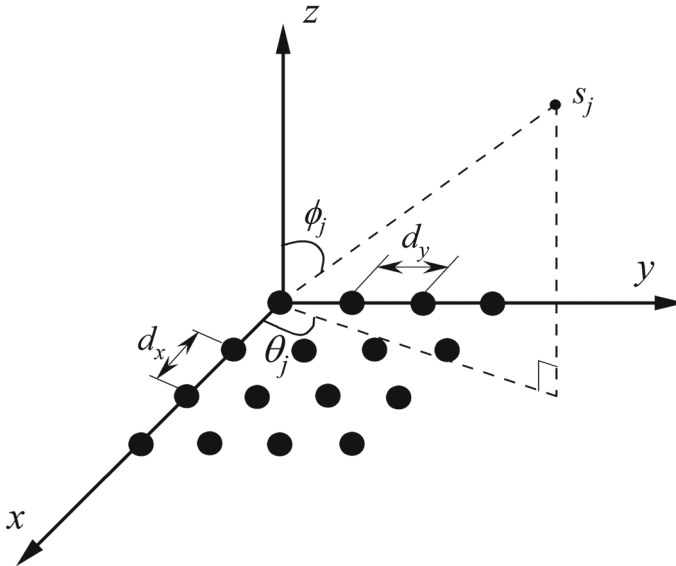


Fig. 1 The geometry of a uniform rectangular array

by

$$\mathbf{Y}(\boldsymbol{\omega}) = \sum_{j=1}^n s_j p_j e^{jkd_x \boldsymbol{\omega}_1 y_{j,1}} e^{jkd_y \boldsymbol{\omega}_2 y_{j,2}} + \mathbf{W}(\boldsymbol{\omega}), \quad \boldsymbol{\omega} \in \{0, 1, \dots, \Omega\}^2, \quad (4.1)$$

where s_j is the j th incident signal, p_j is a complex constant denoting the signal/antenna polarization mismatch, k represents the wavenumber of the carrier frequency, and d_x and d_y denote the distance between adjacent antenna element along the x -axis and y -axis, respectively. $\mathbf{y}_{j,1} = \sin \phi_j \cos \theta_j$ is the direction component of signal s_j propagating along the x -axis and $\mathbf{y}_{j,2} = \sin \phi_j \sin \theta_j$ is the one propagating along the y -axis. The ϕ_j and θ_j denote respectively the elevation and azimuth angles of s_j . $\mathbf{W}(\boldsymbol{\omega})$ is the additive noise, which is usually assumed to be white Gaussian noise. Note that p_j, k, d_x, d_y 's are known factors in practical applications.

For convenience, we consider the following simplified form of (4.1):

$$\mathbf{Y}(\boldsymbol{\omega}) = \sum_{j=1}^n a_j e^{i\mathbf{y}_j^\top \boldsymbol{\omega}} + \mathbf{W}(\boldsymbol{\omega}), \quad \boldsymbol{\omega} \in \{0, 1, \dots, \Omega\}^2, \quad (4.2)$$

where \mathbf{W} is the noise with $|\mathbf{W}(\boldsymbol{\omega})| < \sigma$, $\boldsymbol{\omega} \in \{0, 1, \dots, \Omega\}^2$, and σ being the noise level. We aim to recover stably the number of the signals and the \mathbf{y}_j 's, by which the elevation and azimuth angles are stably resolved. For a better exposition, we still consider a discrete measure $\mu = \sum_{j=1}^n a_j \delta_{\mathbf{y}_j}$ and denote the $a_j \delta_{\mathbf{y}_j}$'s as sources. The measurement (4.2) can be viewed as the noisy Fourier data of the measure μ at some discrete points. To simplify the notation, here we still denote the measurements and

noise by respectively \mathbf{Y} and \mathbf{W} and now they can be viewed as functions who only take nonzeros values at $\boldsymbol{\omega} \in \{0, 1, \dots, \Omega\}^2$. This will not result in any confusion.

In this section and the next one, we shall propose new algorithms for detecting the model order and recovering the supports of μ from the measurement (4.2). Our number detection method is based on thresholding on a Hankel matrix assembled by data from modifications of (4.2). The following subsection shall introduce the details of the Hankel matrix formulation. We refer to [2, 3, 13, 27, 28, 36, 44, 45, 58, 62, 75, 76] for other model detecting algorithms.

4.2 Hankel Matrix Construction

The Hankel matrix is constructed by the following three steps.

Measurement modification by source translation

Due to the translation invariance, suppose the sources are supported in $[-\lambda, \lambda]^2$, we consider them displacing with a vector \mathbf{v} and get that $\mathbf{x}_j = \mathbf{y}_j + \mathbf{v}$. Using a simple measurement modification technique, we obtain the measurement for the new source $\check{\mu} = \sum_{j=1}^n a_j \delta_{\mathbf{x}_j}$. Specifically, we consider

$$\begin{aligned} \mathbf{X}(\boldsymbol{\omega}) &= e^{i\mathbf{v}^\top \boldsymbol{\omega}} \mathbf{Y}(\boldsymbol{\omega}) = \sum_{j=1}^n a_j e^{i(\mathbf{y}_j + \mathbf{v})^\top \boldsymbol{\omega}} + e^{i\mathbf{v}^\top \boldsymbol{\omega}} \mathbf{W}(\boldsymbol{\omega}) \\ &= \sum_{j=1}^n a_j e^{i\mathbf{x}_j^\top \boldsymbol{\omega}} + \tilde{\mathbf{W}}(\boldsymbol{\omega}), \quad \boldsymbol{\omega} \in \{0, 1, \dots, \Omega\}^2, \end{aligned} \tag{4.3}$$

with $|\tilde{\mathbf{W}}(\boldsymbol{\omega})| < \sigma$.

Measurement modification by coordinate-combination The second procedure consists in modifying the measurement based on coordinate-combination. For $s > 0$, let $r = \frac{\Omega}{2s}$. From the measurement \mathbf{X} , we construct a list of new data given by

$$\mathbf{D}(t) = \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} \binom{t}{t_1} \mathbf{X}(\boldsymbol{\omega}_{rt_1, rt_2}), \quad t = 0, \dots, 2s,$$

where $\boldsymbol{\omega}_{rt_1, rt_2} = (rt_1, rt_2)^\top$. Note that

$$\begin{aligned} \mathbf{D}(t) &= \sum_{j=1}^n a_j (e^{i\mathbf{x}_{j,1}r} + e^{i\mathbf{x}_{j,2}r})^t + \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} \binom{t}{t_1} \tilde{\mathbf{W}}(\boldsymbol{\omega}_{rt_1, rt_2}) \\ &= \sum_{j=1}^n a_j (e^{i\mathbf{x}_{j,1}r} + e^{i\mathbf{x}_{j,2}r})^t + \hat{\mathbf{W}}(t), \end{aligned}$$

where $\hat{\mathbf{W}}(t) = \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} \binom{t}{t_1} \tilde{\mathbf{W}}(\boldsymbol{\omega}_{rt_1, rt_2})$.

Hankel matrix construction and singular value decomposition Finally, from these $\mathbf{D}(t)$'s, we assemble the following Hankel matrix

$$\mathbf{H}(s) = \begin{pmatrix} \mathbf{D}(0) & \mathbf{D}(1) & \cdots & \mathbf{D}(s) \\ \mathbf{D}(1) & \mathbf{D}(2) & \cdots & \mathbf{D}(s+1) \\ \cdots & \cdots & \ddots & \cdots \\ \mathbf{D}(s) & \mathbf{D}(s+1) & \cdots & \mathbf{D}(2s) \end{pmatrix}. \tag{4.4}$$

We observe that $\mathbf{H}(s)$ has the decomposition

$$\mathbf{H}(s) = \mathbf{B}\mathbf{A}\mathbf{B}^T + \Delta, \tag{4.5}$$

where $\mathbf{A} = \text{diag}(a_1, \dots, a_n)$ and $\mathbf{B} = (\phi_s(e^{i\mathbf{x}_1,1r} + e^{i\mathbf{x}_1,2r}), \dots, \phi_s(e^{i\mathbf{x}_n,1r} + e^{i\mathbf{x}_n,2r}))$ with $\phi_s(\omega)$ being defined as

$$\phi_s(\omega) = (1, \omega, \dots, \omega^s)^\top, \tag{4.6}$$

and

$$\Delta = \begin{pmatrix} \hat{\mathbf{W}}(0) & \hat{\mathbf{W}}(1) & \cdots & \hat{\mathbf{W}}(s) \\ \hat{\mathbf{W}}(1) & \hat{\mathbf{W}}(2) & \cdots & \hat{\mathbf{W}}(s+1) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{W}}(s) & \hat{\mathbf{W}}(s+1) & \cdots & \hat{\mathbf{W}}(2s) \end{pmatrix}. \tag{4.7}$$

We denote the singular value decomposition of $\mathbf{H}(s)$ as

$$\mathbf{H}(s) = \hat{\mathbf{U}}\hat{\Sigma}\hat{\mathbf{U}}^*,$$

where $\hat{\Sigma} = \text{diag}(\hat{\sigma}_1, \dots, \hat{\sigma}_n, \hat{\sigma}_{n+1}, \dots, \hat{\sigma}_{s+1})$ with the singular values $\hat{\sigma}_j, 1 \leq j \leq s+1$, ordered in a decreasing manner. The source number n is then detected by thresholding on these singular values. In the next subsection we will provide the theoretical guarantee of the threshold.

4.3 Theoretical Guarantee

Note that when there is no noise, $\mathbf{H}(s) = \mathbf{B}\mathbf{A}\mathbf{B}^T$. We have the following estimate for the singular values of $\mathbf{B}\mathbf{A}\mathbf{B}^T$.

Lemma 4.1 *Let $n \geq 2, s \geq n, \mathbf{y}_j \in [-\frac{s\pi}{6\Omega}, \frac{s\pi}{6\Omega}]^2, 1 \leq j \leq n$, and \mathbf{v} in (4.3) be $(0, \frac{s\pi}{\Omega})^\top$. Let*

$$\sigma_1, \dots, \sigma_n, 0, \dots, 0$$

be the singular values of BAB^T in (4.5) ordered in a decreasing manner. Then the following estimate holds

$$\sigma_n \geq \frac{m_{\min}(3\theta_{\min}(\Omega, s))^{2n-2}}{n(2(1 + \sqrt{3})\pi)^{2n-2}}, \tag{4.8}$$

where $\theta_{\min}(\Omega, s) = \min_{p \neq j} \left\| \mathbf{y}_p \frac{\Omega}{2s} - \mathbf{y}_j \frac{\Omega}{2s} \right\|_1$.

Proof Recall that σ_n is the minimum nonzero singular value of BAB^T . Let $\ker(B^T)$ be the kernel space of B^T and $\ker^\perp(B^T)$ be its orthogonal complement. Then we have

$$\begin{aligned} \sigma_n &= \min_{\|x\|_2=1, x \in \ker^\perp(B^T)} \|BAB^T x\|_2 \geq \sigma_{\min}(BA)\sigma_n(B^T) \\ &\geq \sigma_{\min}(B)\sigma_{\min}(A)\sigma_{\min}(B). \end{aligned}$$

On the other hand, since by the condition of the lemma $\mathbf{x}_j = \mathbf{y}_j + \mathbf{v} \in [-\frac{s\pi}{6\Omega}, \frac{s\pi}{6\Omega}] \times [\frac{5s\pi}{6\Omega}, \frac{7s\pi}{6\Omega}]$, we have $\frac{\Omega \mathbf{x}_j}{2s} \in [-\frac{\pi}{12}, \frac{\pi}{12}] \times [\frac{5\pi}{12}, \frac{7\pi}{12}]$. Thus, by Lemma 3.2, for $r = \frac{\Omega}{2s}$,

$$\min_{p \neq q} \left| e^{i\mathbf{x}_{p,1}r} + e^{i\mathbf{x}_{p,2}r} - (e^{i\mathbf{x}_{q,1}r} + e^{i\mathbf{x}_{q,2}r}) \right| \geq \frac{3}{2\pi} \theta_{\min}(\Omega, s).$$

Note also that $|e^{i\mathbf{x}_{p,1}r} + e^{i\mathbf{x}_{p,2}r}| \leq \sqrt{3}$. Thus applying Lemma 6.3 and Corollary 6.2, we have

$$\sigma_{\min}(B) \geq \frac{1}{\sqrt{n}} \frac{\left(\frac{3}{2\pi} \theta_{\min}(\Omega, s)\right)^{n-1}}{(1 + \sqrt{3})^{n-1}}.$$

Then, it follows that

$$\sigma_n \geq \sigma_{\min}(A) \left(\frac{\left(\frac{3}{2\pi} \theta_{\min}(\Omega, s)\right)^{n-1}}{(1 + \sqrt{3})^{n-1}} \right)^2 \geq \frac{m_{\min}(3\theta_{\min}(\Omega, s))^{2n-2}}{n(2(1 + \sqrt{3})\pi)^{2n-2}}.$$

□

We now present the main result on the threshold for the singular values of the matrix $\mathbf{H}(s)$.

Theorem 4.1 Let $n \geq 2, s \geq n$ and $\mu = \sum_{j=1}^n a_j \delta_{y_j}$ with $\mathbf{y}_j \in [-\frac{s\pi}{6\Omega}, \frac{s\pi}{6\Omega}]^2, 1 \leq j \leq n$. Let \mathbf{v} in (4.3) be equal to $(0, \frac{s\pi}{\Omega})^\top$. Then for the singular values of $\mathbf{H}(s)$ in (4.4), we have

$$\hat{\sigma}_j < \frac{4^{s+1}\sigma}{3}, \quad j = n + 1, \dots, s + 1. \tag{4.9}$$

Moreover, if the following separation condition is satisfied

$$\min_{p \neq j} \|\mathbf{y}_p - \mathbf{y}_j\|_1 \geq \frac{4(1 + \sqrt{3})\pi s}{3\Omega} \left(\frac{2n4^{s+1}}{3} \frac{\sigma}{m_{\min}} \right)^{\frac{1}{2n-2}}, \tag{4.10}$$

then

$$\hat{\sigma}_n \geq \frac{4^{s+1}\sigma}{3}. \tag{4.11}$$

Proof We first estimate $\|\Delta\|_2$ for Δ in (4.7). By the definition of $\hat{\mathbf{W}}(t)$, we have $|\hat{\mathbf{W}}(t)| < 2^t \sigma$. Thus $\|\Delta\|_2 \leq \|\Delta\|_F < \frac{4^{s+1}\sigma}{3}$. By Weyl’s theorem, we have $|\hat{\sigma}_j - \sigma_j| \leq \|\Delta\|_2, j = 1, \dots, n$. Together with $\sigma_j = 0, n + 1 \leq j \leq s + 1$, we get $|\hat{\sigma}_j| \leq \|\Delta\|_2 < \frac{4^{s+1}\sigma}{3}, n + 1 \leq j \leq s + 1$. This proves (4.9).

Let $\theta_{\min}(\Omega, s) = \frac{\Omega}{2^s} \min_{p \neq q} \|\mathbf{y}_p - \mathbf{y}_q\|_1$. The separation condition (4.10) implies that

$$\theta_{\min}(\Omega, s) \geq \frac{2(1 + \sqrt{3})\pi}{3} \left(\frac{2n4^{s+1}}{3} \frac{\sigma}{m_{\min}} \right)^{\frac{1}{2n-2}}.$$

By Lemma 4.1, we have

$$\sigma_n \geq \frac{m_{\min} (3\theta_{\min}(\Omega, s))^{2n-2}}{n(2(1 + \sqrt{3})\pi)^{2n-2}} > 2 \frac{4^{s+1}\sigma}{3}. \tag{4.12}$$

Similarly, by Weyl’s theorem, $|\hat{\sigma}_n - \sigma_n| \leq \|\Delta\|_2$. Thus, $\hat{\sigma}_n \geq \sigma_n - \|\Delta\|_2 > 2 \frac{4^{s+1}\sigma}{3} - \frac{4^{s+1}\sigma}{3} = \frac{4^{s+1}\sigma}{3}$. The conclusion (4.11) then follows. \square

4.4 Coordinate-Combination-Based Sweeping Singular-Value-Thresholding Number Detection Algorithm

Based on Theorem 4.1, we can propose a simple thresholding algorithm, *Algorithm 1*, for the number detection.

Algorithm 1: Coordinate-combination-based singular-value-thresholding number detection algorithm

- Input:** Number s ; Noise level σ ;
Input: Measurement: $\mathbf{Y}(\omega)$, $\omega \in \{0, 1, \dots, \Omega\}^2$;
Input: Translation vector \mathbf{v} in \mathbb{R}^2 ;
 1: Modify the measurement and get $\mathbf{X}(\omega) = e^{i\mathbf{v}^\top \omega} \mathbf{Y}(\omega)$;
 2: Let $r = \Omega \bmod 2s$, formulate $\mathbf{D}(t) = \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} \binom{t}{t_1} \mathbf{X}(\omega_{rt_1, rt_2})$, $t = 0, \dots, 2s$;
 3: Assemble the $(s + 1) \times (s + 1)$ Hankel matrix $\mathbf{H}(s)$ like (4.4) from $\mathbf{D}(t)$'s, and compute the singular value of $\mathbf{H}(s)$ as $\hat{\sigma}_1, \dots, \hat{\sigma}_{s+1}$ distributed in a decreasing manner;
 4: Determine n by $\hat{\sigma}_n \geq \frac{4^{s+1}\sigma}{3}$ and $\hat{\sigma}_j < \frac{4^{s+1}\sigma}{3}$, $j = n + 1, \dots, s + 1$;
Return: n
-

Note that if the minimum separation distance d_{\min} in (4.10) increases to cd_{\min} for some $c > 1$, then we have

$$\hat{\sigma}_n > c^{2n-2} \frac{4^{s+1}\sigma}{3}.$$

Thus even if we increase a little the separation distance, then we will get a large gap between $\hat{\sigma}_n$ and $\hat{\sigma}_{n+1}$. This indicates that our algorithm is very stable and not sensitive to the noise level estimate. In practical applications, the noise level is usually not precisely known, but the robustness of our algorithm means that one can estimate a noise level although not tight and utilize our algorithm to detect the source number with high resolution.

On the other hand, if $s < n$ in the applications, by (4.11) we have

$$\hat{\sigma}_j \geq \frac{4^{s+1}\sigma}{3}, \quad j = 1, \dots, s + 1,$$

when separation condition (4.10) holds, and the source number is deduced to be $s + 1$ by *Algorithm 1*. When $s \geq n$ and (4.10) holds, *Algorithm 1* gives the exact n . Thus in applications, one is suggested to choose a suitable and large enough s . However, a suitable s is not easy to estimate and large s may lead to a deterioration of the resolution as indicated by (4.10). To remedy this issue, we propose a sweeping singular-value-thresholding number detection algorithm (*Algorithm 2*) below. In short, we detect the number n_{recover} by *Algorithm 1* for all s from 2 to $\lfloor \frac{\Omega-1}{2} \rfloor$, and choose the greatest one n_{max} as the number of point sources. When the detected n_{recover} becomes smaller than n_{max} for a large number of iterations, we will stop the loop. The details are summarized in *Algorithm 2* below.

We remark that when $s = n$ and the point sources satisfy

$$\min_{p \neq q} \|\mathbf{y}_p - \mathbf{y}_q\|_1 \geq \frac{Cn}{\Omega} \left(\frac{\sigma}{m_{\min}} \right)^{\frac{1}{2n-2}}, \tag{4.13}$$

for some constant C , then (4.10) is satisfied. Thus by Theorem 4.1, for a suitable choice of \mathbf{v} , Algorithm 1 can exactly detect the number n when $s = n$. As s increases to values greater than n , (4.9) implies that the number detected by Algorithm 1 will not exceed n . Therefore, the sweeping singular-value-thresholding algorithm (Algorithm 2) can detect the exact number n when Ω is greater than $2n + 1$ and the point sources are separated by the minimal separation distance we derived in Theorem 2.1. This demonstrates the optimal performance of Algorithm 2. We also remark that the theoretical threshold derived in Theorem 4.1 seems to be larger than the one that is needed. One can improve the algorithm by choosing smaller threshold. Deriving new estimates for the thresholds in different cases is another interesting problem.

Algorithm 2: Coordinate-combination-based sweeping singular-value-thresholding number detection algorithm

Input: Noise level σ ; Measurement: \mathbf{Y} ; Translation vector \mathbf{v} ;
Input: $n_{max} = 0, s_{max_index} = 2$
for $s = 2 : \lfloor \frac{\Omega-1}{2} \rfloor$ **do**
 Input $s, \sigma, \mathbf{Y}, \mathbf{v}$ to Algorithm 1, save the output of Algorithm 1 as $n_{recover}$;
 if $n_{recover} > n_{max}$ **then**
 $n_{max} = n_{recover}$;
 $s_{max_index} = s$;
 if $s \geq s_{max_index} + 2$ **then**
 break;
Return n_{max} .

4.5 Phase Transition and Performance of Algorithm 2

In this subsection, we conduct numerical experiments to demonstrate the phase transition phenomenon regarding the super-resolution factor (SRF) and the SNR using Algorithm 2. We consider recovering the number of three and four sources. We fix $\Omega = 10$ and detect the source number from their noisy Fourier data at $\{0, 1, \dots, \Omega\}^2$. We consider sources in $[0, \frac{\pi}{2}]^2$ and the translation vector in Algorithm 1 is $\mathbf{v} = (0, \frac{\pi}{2})^T$. The noise level is σ and the minimum separation distance between sources is D_{\min} . We perform 10,000 random experiments (the randomness is in the choice of $(D_{\min}, \sigma, \mathbf{y}_j, a_j)$) and detect the source number by Algorithm 2. We record the number of each successful detection (source number is detected exactly) and failed detection. Figure 2 shows the result for the successful and unsuccessfully recovery in the parameter space $\log(\text{SNR})$ versus $\log(\text{SRF})$. It is observed that there is a line with slope $(2n - 2)$ in the parameter space of $\log(\text{SRF})$ - $\log(\text{SNR})$ above which the number detection is always successful. This phase transition phenomenon is exactly the one predicted by our theoretical results in Theorems 2.1 and 4.1. It also illustrates the efficiency of Algorithm 2

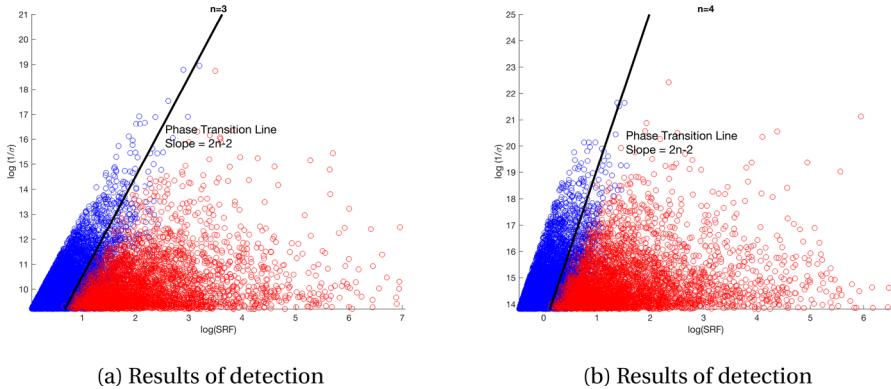


Fig. 2 Plots of the successful and the unsuccessful number detection by *Algorithm 2* depending on the relation between $\log(\text{SRF})$ and $\log(\frac{1}{\sigma})$. **a** illustrates that three sources can be exactly detected if $\log(\frac{1}{\sigma})$ is above a line of slope 4 in the parameter space. **b** illustrates that four sources can be exactly detected if $\log(\frac{1}{\sigma})$ is above a line of slope 6 in the parameter space

as it can resolve the source number correctly in the regime where the source separation distance is of the order of the computational resolution limit.

5 An Algorithm for the Source Reconstruction in Two-Dimensional DOA Problems

In this section, based on the idea of coordinate-combination, we propose a new MUSIC algorithm for resolving the sources in the two-dimensional DOA estimation. Our algorithm is named as coordinate-combination-based MUSIC algorithm; see *Algorithm 4*.

5.1 Hankel Matrix Construction

Similarly to the number detection algorithm in the above section, the MUSIC algorithm also relies on a singular value decomposition of certain Hankel matrix. Compared to conventional MUSIC-based DOA algorithms, the main novelty of our algorithm lies in a different way of assembling Hankel matrices. Similarly to Sect. 4.2, the Hankel matrix construction here is also based on observations in Sect. 3.1 and the details are presented below.

Measurement modification by source translation We consider the same model setting as (4.2) for the available measurement. We also perform the source translation and modify the measurement to get

$$\begin{aligned}
 \mathbf{X}(\omega) &= e^{i\mathbf{v}^T \omega} \mathbf{Y}(\omega) = \sum_{j=1}^n a_j e^{i(\mathbf{y}_j + \mathbf{v})^T \omega} + e^{i\mathbf{v}^T \omega} \mathbf{W}(\omega) \\
 &= \sum_{j=1}^n a_j e^{i\mathbf{x}_j^T \omega} + \tilde{\mathbf{W}}(\omega), \quad \omega \in \{0, 1, \dots, \Omega\}^2,
 \end{aligned}
 \tag{5.1}$$

where $\mathbf{x}_j = \mathbf{y}_j + \mathbf{v}$ for a suitable $\mathbf{v} \in \mathbb{R}^2$ and $|\tilde{\mathbf{W}}(\boldsymbol{\omega})| < \sigma$.

Measurement modification by the coordinate-combination technique

Let $s = \lfloor \frac{\Omega}{2} \rfloor$. From the modified measurement $\mathbf{X}(\boldsymbol{\omega})$, we construct the following two lists of data:

$$\mathbf{D}(t) = \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} \binom{t}{t_1} \mathbf{X}(\boldsymbol{\omega}_{t_1, t_2}), \quad t = 0, \dots, 2s,$$

$$\mathbf{G}(t) = \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} (-1)^{t_2} \binom{t}{t_1} \mathbf{X}(\boldsymbol{\omega}_{t_1, t_2}), \quad t = 0, \dots, 2s,$$

where $\boldsymbol{\omega}_{t_1, t_2} = (t_1, t_2)^\top$. Note that

$$\begin{aligned} \mathbf{D}(t) &= \sum_{j=1}^n a_j (e^{i\mathbf{x}_{j,1}} + e^{i\mathbf{x}_{j,2}})^t + \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} \binom{t}{t_1} \tilde{\mathbf{W}}(\boldsymbol{\omega}_{t_1, t_2}) \\ &= \sum_{j=1}^n a_j (e^{i\mathbf{x}_{j,1}} + e^{i\mathbf{x}_{j,2}})^t + \hat{\mathbf{W}}_d(t), \\ \mathbf{G}(t) &= \sum_{j=1}^n a_j (e^{i\mathbf{x}_{j,1}} - e^{i\mathbf{x}_{j,2}})^t + \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} (-1)^{t_2} \binom{t}{t_1} \tilde{\mathbf{W}}(\boldsymbol{\omega}_{t_1, t_2}) \\ &= \sum_{j=1}^n a_j (e^{i\mathbf{x}_{j,1}} - e^{i\mathbf{x}_{j,2}})^t + \hat{\mathbf{W}}_g(t), \end{aligned}$$

where $\hat{\mathbf{W}}_d(t) = \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} \binom{t}{t_1} \tilde{\mathbf{W}}(\boldsymbol{\omega}_{t_1, t_2})$ and $\hat{\mathbf{W}}_g(t) = \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} (-1)^{t_2} \binom{t}{t_1} \tilde{\mathbf{W}}(\boldsymbol{\omega}_{t_1, t_2})$.

Hankel matrix construction

Finally, from these $\mathbf{D}(t)$, $\mathbf{G}(t)$'s, we assemble the following Hankel matrices:

$$\mathbf{H}_d(s) = \begin{pmatrix} \mathbf{D}(0) & \mathbf{D}(1) & \dots & \mathbf{D}(s) \\ \mathbf{D}(1) & \mathbf{D}(2) & \dots & \mathbf{D}(s+1) \\ \dots & \dots & \ddots & \dots \\ \mathbf{D}(s) & \mathbf{D}(s+1) & \dots & \mathbf{D}(2s) \end{pmatrix},$$

$$\mathbf{H}_g(s) = \begin{pmatrix} \mathbf{G}(0) & \mathbf{G}(1) & \dots & \mathbf{G}(s) \\ \mathbf{G}(1) & \mathbf{G}(2) & \dots & \mathbf{G}(s+1) \\ \dots & \dots & \ddots & \dots \\ \mathbf{G}(s) & \mathbf{G}(s+1) & \dots & \mathbf{G}(2s) \end{pmatrix}. \tag{5.2}$$

5.2 Standard MUSIC Algorithm

In this subsection, we perform the standard MUSIC algorithm [40, 47, 60, 67] for the Hankel matrix $\mathbf{H}_d(s)$, $\mathbf{H}_g(s)$ in (5.2). For ease of presentation, we only introduce the

MUSIC algorithm for $\mathbf{H}_d(s)$. The one for $\mathbf{H}_g(s)$ can be developed in the same manner. Our algorithm first performs the singular value decomposition of $\mathbf{H}_d(s)$,

$$\mathbf{H}_d(s) = \hat{U} \hat{\Sigma} \hat{U}^* = [\hat{U}_1 \quad \hat{U}_2] \text{diag}(\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_n, \hat{\sigma}_{n+1}, \dots, \hat{\sigma}_{s+1}) [\hat{U}_1 \quad \hat{U}_2]^*,$$

where $\hat{U}_1 = (\hat{U}(1), \dots, \hat{U}(n))$, $\hat{U}_2 = (\hat{U}(n+1), \dots, \hat{U}(s+1))$ with n being the estimated source number (model order). The source number n can be detected by *Algorithm 2* and many other algorithms such as those in [2, 13, 27, 28, 44, 45, 62, 75, 76]. Denote the orthogonal projection onto the space \hat{U}_2 by $\hat{P}_2 x = \hat{U}_2 (\hat{U}_2^* x)$. For a test vector $\Phi(d) = (1, d, \dots, d^s)^\top$, one defines the MUSIC imaging functional

$$\hat{J}(d) = \frac{\|\Phi(d)\|_2}{\|\hat{P}_2 \Phi(d)\|_2} = \frac{\|\Phi(d)\|_2}{\|\hat{U}_2^* \Phi(d)\|_2}.$$

The local maximizer of $\hat{J}(d)$ indicates the supports of the sources. In practice, one can test evenly spaced points in a specified region and plot the discrete imaging functional and then determine the sources by detecting the peaks. In our case, we only need to test some discrete points $d \in \mathbb{C}$ with $|d| \leq 2$ and select the peak by certain algorithms (such as the one in [47] or its two-dimensional analog). Finally, we summarize the standard MUSIC algorithm in *Algorithm 3* below.

Algorithm 3: Standard MUSIC algorithm

- Input:** Source number n ;
 - Input:** Modified measurements: $\mathbf{D}(t)$ (or $\mathbf{G}(t)$), $t = 0, \dots, s$ with $s \geq n$;
 - Input:** Test points d 's;
 - 1: Formulate the $(s+1) \times (s+1)$ Hankel matrix $\mathbf{H}_d(s)$ from $\mathbf{D}(t)$'s as (5.2);
 - 2: Compute the singular vectors of $\mathbf{H}_d(s)$ as $\hat{U}(1), \hat{U}(2), \dots, \hat{U}(s+1)$ and form the noise space $\hat{U}_2 = (\hat{U}(n+1), \dots, \hat{U}(s+1))$;
 - 3: For test points d 's, construct the test vector $\Phi(d) = (1, d, \dots, d^s)^\top$;
 - 4: Plot the MUSIC imaging functional $\hat{J}(d) = \frac{\|\Phi(d)\|_2}{\|\hat{U}_2^* \Phi(d)\|_2}$;
 - 5: Select the peak locations \hat{d}_j 's in the plot of $\hat{J}(d)$.
-

5.3 Coordinate-Combination-Based MUSIC Algorithm

After applying the MUSIC algorithm to both $\mathbf{H}_d(s)$, $\mathbf{H}_g(s)$, we expect to reconstruct n \hat{d}_j 's which is close to $d_j = e^{ix_{j,1}} + e^{ix_{j,2}}$, and n \hat{g}_j 's which is close to $g_j = e^{ix_{j,1}} - e^{ix_{j,2}}$. The next question is how to link the pair \hat{d}_j, \hat{g}_j that correspond to the same source. This is an inevitable pair matching issue in most of the two-dimensional DOA algorithms [44], where ad hoc schemes [14, 34, 79, 81] were derived to associate the estimated azimuth and elevation angles. Here, in contrast with conventional DOA algorithms, we do not need to link the azimuth and elevation angles but to link \hat{d}_j and \hat{g}_j .

Observe that $|d_j + g_j| = |2e^{i\mathbf{x}_{j,1}}| = 2$ and $|d_j - g_j| = |2e^{i\mathbf{x}_{j,2}}| = 2$. We can use this criterion to match the pair \hat{d}_j, \hat{g}_j that they should satisfy

$$|\hat{d}_j + \hat{g}_j| \approx 2, \quad |\hat{d}_j - \hat{g}_j| \approx 2. \quad (5.3)$$

For example, we could consider the following minimization problem:

$$\min_{\pi \in \zeta(n)} \sum_{j=1}^n \left| |\hat{d}_j + \hat{g}_{\pi_j}| - 2 \right| + \left| |\hat{d}_j - \hat{g}_{\pi_j}| - 2 \right|, \quad (5.4)$$

where $\zeta(n)$ is the set of all permutations of $\{1, \dots, n\}$. This can be viewed as a balanced assignment problem [54], which can be solved efficiently by many algorithms such as the Hungarian algorithm.

We remark that our pair matching algorithm is not the one usually required in other one-dimensional based DOA algorithms. Unlike our case, the other pair matching problem is not an assignment problem, wherefore the pair matching is usually time consuming or complex processing is conducted to reduce the computational cost.

Algorithm 4: Coordinate-combination-based MUSIC algorithm for two-dimensional DOA

Input: Source number n ; noise level σ ;

Input: Measurement: $\mathbf{Y}(\omega)$, $\omega \in \{0, 1, \dots, \Omega\}^2$;

Input: Translation vector \mathbf{v} in \mathbb{R}^2 ;

Input: Evenly spaced test points $d \in \mathbb{C}$ with $|d| \leq 2$;

1: Modify the measurement and get $\mathbf{X}(\omega) = e^{i\mathbf{v}^\top \omega} \mathbf{Y}(\omega)$;

2: Let $s = \lfloor \frac{\Omega}{2} \rfloor$, formulate $\mathbf{D}(t) = \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} \binom{t}{t_1} \mathbf{X}(\omega_{t_1, t_2})$, $\mathbf{G}(t) = \sum_{t_1+t_2=t, 0 \leq t_1, t_2 \leq t} (-1)^{t_2} \binom{t}{t_1} \mathbf{X}(\omega_{t_1, t_2})$, $t = 0, \dots, 2s$;

3: Input \mathbf{D} , n and test points d 's into **Algorithm 3** and get the output $\hat{d}_1, \dots, \hat{d}_n$;

4: Input \mathbf{G} , n and test points d 's into **Algorithm 3** and get the output $\hat{g}_1, \dots, \hat{g}_n$;

5: Matching the \hat{d}_j, \hat{g}_j 's by applying an assignment algorithm (match pairs in matlab) to solve (5.4) and get the pair list $\{(\hat{d}_j, \hat{g}_j)\}_{j=1}^n$;

6: Get $\frac{\hat{d}_j + \hat{g}_j}{2}$ and $\frac{\hat{d}_j - \hat{g}_j}{2}$, $j = 1, \dots, n$. Get $e^{i\hat{\mathbf{x}}_{j,1r}}$ by considering the closest point to $\frac{\hat{d}_j + \hat{g}_j}{2}$ on the unit circle. Get $e^{i\hat{\mathbf{x}}_{j,2}}$ by considering the closest point to $\frac{\hat{d}_j - \hat{g}_j}{2}$ on the unit circle;

7: The recovered $\hat{\mathbf{x}}_j = (\hat{\mathbf{x}}_{j,1}, \hat{\mathbf{x}}_{j,2})^\top$. Reconstruct $\hat{\mathbf{y}}_j = \hat{\mathbf{x}}_j - \mathbf{v}$, $j = 1, \dots, n$;

Return: $\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_n$.

5.4 Superiority of the Algorithm

5.4.1 Overcome the Issue of Separation Distance Loss in Conventional Two-Dimensional DOA Algorithms

Despite the fact that different recovering methods are proposed for DOA estimation in two dimensions, the conventional way for tackling the problem has hardly exceeded the scope of recovering the two direction (x- and y-direction) components of sources individually. Thus, as illustrated in Fig. 3a, severe loss of the source separation distance in one dimension is always an inevitable issue that causes unstable recovery of the direction components. Most of the researches ignored this issue and some papers [73, 74] proposed ad hoc schemes to enhance the reconstruction but in a complex manner. We also present an example in Fig. 3b to compare the recoveries of our *Algorithm 4* and a projection-based MUSIC algorithm (similar to the one in [44]). It clearly illustrates the recovery instability that is due to the projection.

Our method is a new one-dimensional-based algorithm where the issue of severe source separation distance loss is avoided in a simple way. In our algorithm, the separation distance between direction components of sources are still preserved. This has been demonstrated by Lemma 3.2 for $\theta_j \in [0, \frac{2}{3}\pi]^2, j = 1, 2$, with $\frac{\pi}{3} \leq \theta_{j,2} - \theta_{j,1} \leq \frac{2}{3}\pi, j = 1, 2$. Furthermore, Theorem 5.1 shows that, for $\mathbf{y}_j \in [0, \frac{\pi}{2}]^2$ and $\mathbf{v} = (0, \frac{\pi}{2})^\top$, the separation distance between $\mathbf{x}_j = \mathbf{y}_j + \mathbf{v}$'s can be preserved after the coordinate-combination. By Theorem 5.1, if the distance between the \mathbf{x}_j 's is a certain constant C , then the distance between $e^{i\mathbf{x}_{j,1}} + e^{i\mathbf{x}_{j,2}}$ is larger than $\frac{2C}{\pi^2}$ times the original distance. For better results of preservation of the distance, as indicated by Theorems 2.1 and 2.2, we could consider sources in a smaller region with a specified translation. In the numerical experiments presented in this paper, for ease of discussion and presentation, we will consider sources in $[0, \frac{\pi}{2}]^2$ and the translation vector $\mathbf{v} = (0, \frac{\pi}{2})^\top$. We leave

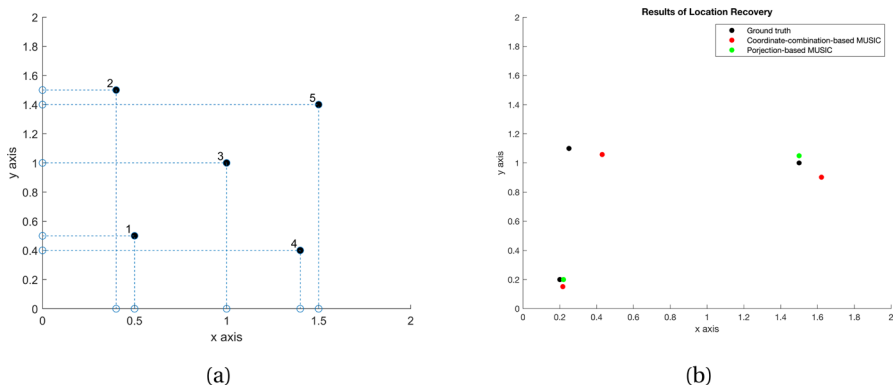


Fig. 3 **a** shows the severe loss of distance when considering the direction components of the sources. **b** illustrates an example of location recovery. The black points are the underlying sources and the red points are locations recovered by coordinate-combination-based MUSIC algorithm. The green points are locations recovered by one projection-based MUSIC algorithm. Although three sources are presented, the projection-based MUSIC can only recover two sources (Color figure online)

the recovering strategies of the whole region $[0, 2\pi]^2$ and other enhancement for future work.

Theorem 5.1 For two different vectors $\mathbf{x}_j \in [0, \frac{\pi}{2}] \times [\frac{\pi}{2}, \pi]$, $j = 1, 2$, if $\|\mathbf{x}_1 - \mathbf{x}_2\|_1 \geq C$ for a constant C , then

$$\left| e^{i\mathbf{x}_{1,1}} + e^{i\mathbf{x}_{1,2}} - (e^{i\mathbf{x}_{2,1}} + e^{i\mathbf{x}_{2,2}}) \right| \geq \frac{2C}{\pi^2} C.$$

Proof We prove the lemma by considering the following two cases.

Case 1 $0 \leq \mathbf{x}_{1,1} \leq \mathbf{x}_{2,1} \leq \mathbf{x}_{2,2} \leq \mathbf{x}_{1,2} \leq \pi$.

In this case,

$$\begin{aligned} \left| e^{i\mathbf{x}_{1,1}} + e^{i\mathbf{x}_{1,2}} - (e^{i\mathbf{x}_{2,1}} + e^{i\mathbf{x}_{2,2}}) \right| &\geq \left| e^{i\mathbf{x}_{2,1}} + e^{i\mathbf{x}_{2,2}} \right| - \left| e^{i\mathbf{x}_{1,1}} + e^{i\mathbf{x}_{1,2}} \right| \\ &\geq 2 \left(\cos \left(\frac{\phi_2}{2} \right) - \cos \left(\frac{\phi_1}{2} \right) \right), \end{aligned}$$

where $\phi_j = \mathbf{x}_{j,2} - \mathbf{x}_{j,1}$, $j = 1, 2$. By the assumption of the theorem, we have $C \leq \phi_1 - \phi_2 \leq \pi$ and $C \leq \phi_1 + \phi_2 \leq 2\pi$. Thus

$$\begin{aligned} 2 \left(\cos \left(\frac{\phi_2}{2} \right) - \cos \left(\frac{\phi_1}{2} \right) \right) &= 4 \sin \left(\frac{\phi_1 + \phi_2}{4} \right) \sin \left(\frac{\phi_1 - \phi_2}{4} \right) \\ &\geq 4 \sin \left(\frac{C}{4} \right) \sin \left(\frac{C}{4} \right) \geq \frac{2C^2}{\pi^2}. \end{aligned}$$

where the last inequality uses $\sin(\frac{C}{4}) \geq \frac{2\sqrt{2}}{\pi} \frac{C}{4}$ for $0 \leq \frac{C}{4} \leq \frac{\pi}{4}$.

Case 2 $0 \leq \mathbf{x}_{1,1} \leq \mathbf{x}_{2,1} \leq \mathbf{x}_{1,2} \leq \mathbf{x}_{2,2} \leq \pi$.

Again, the idea is to calculate the angle between $e^{i\mathbf{x}_{1,1}} + e^{i\mathbf{x}_{1,2}}$ and $e^{i\mathbf{x}_{2,1}} + e^{i\mathbf{x}_{2,2}}$. By a simple analysis of the angle relations between $e^{i\mathbf{x}_{1,1}}, e^{i\mathbf{x}_{1,2}}, e^{i\mathbf{x}_{2,1}}, e^{i\mathbf{x}_{2,2}}$, we obtain that the angle between $e^{i\mathbf{x}_{1,1}} + e^{i\mathbf{x}_{1,2}}$ and $e^{i\mathbf{x}_{2,1}} + e^{i\mathbf{x}_{2,2}}$ is $\frac{\mathbf{x}_{2,1} - \mathbf{x}_{1,1} + \mathbf{x}_{2,2} - \mathbf{x}_{1,2}}{2}$ which is larger than $\frac{C}{2}$. Thus

$$\begin{aligned} &\left| e^{i\mathbf{x}_{1,1}} + e^{i\mathbf{x}_{1,2}} - (e^{i\mathbf{x}_{2,1}} + e^{i\mathbf{x}_{2,2}}) \right| \\ &\geq \max \left(\left| e^{i\mathbf{x}_{1,1}} + e^{i\mathbf{x}_{1,2}} \right|, \left| e^{i\mathbf{x}_{2,1}} + e^{i\mathbf{x}_{2,2}} \right| \right) \sin \left(\frac{C}{2} \right). \end{aligned} \tag{5.5}$$

We next claim that

$$\max \left(\left| e^{i\mathbf{x}_{1,1}} + e^{i\mathbf{x}_{1,2}} \right|, \left| e^{i\mathbf{x}_{2,1}} + e^{i\mathbf{x}_{2,2}} \right| \right) \geq 2 \cos \left(\frac{\pi - C/2}{2} \right).$$

Otherwise, $\pi - \frac{C}{2} < \mathbf{x}_{1,2} - \mathbf{x}_{1,1} \leq \pi$ and $\pi - \frac{C}{2} < \mathbf{x}_{2,2} - \mathbf{x}_{2,1} \leq \pi$, which is impossible when $\|\mathbf{x}_1 - \mathbf{x}_2\|_1 \geq C$. Thus the claim is proved. Together with (5.5), we

arrive at

$$\left| e^{ix_{1,1}} + e^{ix_{1,2}} - (e^{ix_{2,1}} + e^{ix_{2,2}}) \right| \geq 2 \sin\left(\frac{C}{4}\right) \sin\left(\frac{C}{2}\right) \geq \frac{2C^2}{\pi^2}.$$

This completes the proof. □

5.4.2 Phase Transition and Performance of Algorithm 4

Most of the conventional two-dimensional DOA algorithms consider multiple snapshots of measurements from coherent or incoherent signals. Also, the noise is usually assumed to be white Gaussian noise such that the expectation of the covariance matrix of the measurement vector is a sum of two terms, where the first term is from the correlation of the signals and the second one is the noise correlation matrix. Based on this crucial observation, many algorithms were derived to tackle the problem. Differently to the above model, we consider recovering the source from a single measurement with deterministic noise. Thus we do not compare the performance of our algorithm with those algorithms with statistical model. We demonstrate the super-resolution capacity of our algorithm for the single snapshot case by showing the phase transition of the algorithm. We will derive a coordinate-combination-based MUSIC algorithm for multiple snapshots case in a forthcoming work.

We now describe the numerical experiments for demonstrating the phase transition phenomenon of our algorithm in terms of the SNR versus the super-resolution factor. We fix $\Omega = 10$ and consider three and four sources separated by the minimum separation distance D_{\min} , i.e., $\min_{p \neq q} \|\mathbf{y}_p - \mathbf{y}_q\|_1 \geq D_{\min}$. We perform 10,000 random experiments (the randomness is in the choice of $(D_{\min}, \sigma, \mathbf{y}_j, a_j)$) to recover the sources using *Algorithm 4*. The reconstruction is viewed and recorded as successful if

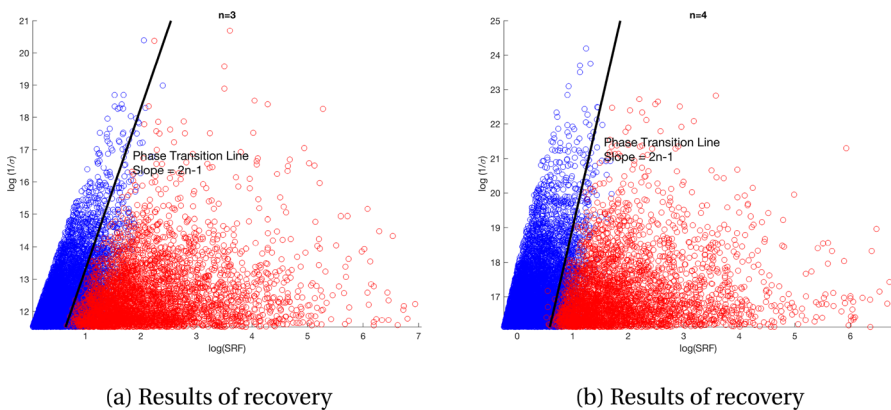


Fig. 4 Plots of the successful and the unsuccessful location recoveries by *Algorithm 4* in terms of $\log(\frac{1}{\sigma})$ versus $\log(\text{SRF})$. **a** illustrates that locations of three point sources can be stably recovered if $\log(\frac{1}{\sigma})$ is above a line of slope 5 in the parameter space. Conversely, for the same case, **b** shows that locations of four point sources can be stably recovered if $\log(\frac{1}{\sigma})$ is above a line of slope 7 in the parameter space

the recovered source is in a $\frac{D_{\min}}{3}$ -neighborhood of the underlying source, otherwise it is unsuccessful; See *Algorithm 5* for the details of a single experiment. The results of the experiments are summarized in Fig. 4 which shows each successful and unsuccessfully recovery with respect to the $\log(\text{SRF})$ and $\log(\text{SNR})$. It is observed that there is a line with slope $(2n - 1)$ in the parameter space $\log(\text{SRF})$ versus $\log(\text{SNR})$ above which the source is stably reconstructed for every realization. This phase transition phenomenon is exactly the one predicted by our theoretical result in Theorem 2.2. It also manifests the efficiency of *Algorithm 4* as it can resolve the source in the regime where the source separation distance is of the order of the computational resolution limit.

Algorithm 5: A single experiment

Input: Sources $\mu = \sum_{j=1}^n a_j \delta_{\mathbf{y}_j}$; Noise level σ ;
Input: Measurements: $\mathbf{Y}(\omega)$, $\omega = \{0, 1, \dots, \Omega\}^2$;
 1: Successnumber = 0;
 2: Input source number n and measurement \mathbf{Y} to *Algorithm 4* and save the output as $\mathbf{y}_1, \dots, \mathbf{y}_n$;
for each $1 \leq j \leq n$ **do**
 Compute the error for the source location \mathbf{y}_j : $e_j := \min_{\hat{\mathbf{y}}_l, l=1, \dots, n} \|\hat{\mathbf{y}}_l - \mathbf{y}_j\|_2$;
 The source location \mathbf{y}_j is recovered successfully if

$$e_j < \frac{\min_{p \neq j} \|\mathbf{y}_p - \mathbf{y}_j\|_2}{3};$$

 and

$$\text{Successnumber} = \text{Successnumber} + 1;$$

if $\text{Successnumber} == n$ **then**
 Return Success
else
 Return Fail

6 A Nonlinear Approximation Theory in Vandermonde Space

Recall that in the proofs of Theorems 2.1 and 2.2, the three observations in Sect. 3.1 are used to transform the two-dimensional super-resolution problems into the following approximation

$$\left| \sum_{j=1}^q \hat{a}_j \hat{d}_j^t - \sum_{j=1}^n a_j d_j^t \right| < C\sigma, \quad t = 0, 1, \dots, s(n), \tag{6.1}$$

where $\hat{a}_j \in \mathbb{C}$, $\hat{d}_j \in \mathbb{C}$, $a_j \in \mathbb{C}$, d_j 's are complex numbers with a minimum separation, $q \leq n$, and $s(n) = 2n - 2$ or $2n - 1$. We are interested in the stability of recovering d_j 's by \hat{d}_j 's from above approximation and call these problems the approximation problems in Vandermonde space. Note that, as already shown in the proofs, they play

a crucial role in proving Theorems 2.1 and 2.2. The main motivation for transforming the two-dimensional super-resolution into the above problems is that we already have enough tools to analyze them. To be specific, in [45, 46], we have introduced a nonlinear approximation theory in Vandermonde space for certain cases to deal with one-dimensional super-resolution problems. In [46], we have derived the theory for real numbers and in [45] for complex numbers on the unit circle. Here, we derive a different theory for arbitrary bounded complex numbers ($d_j \in \mathbb{C}, 1 \leq j \leq n$), which is used to prove Theorems 2.1 and 2.2.

Especially, for a given positive integer s and $\omega \in \mathbb{C}$, we denote by

$$\phi_s(\omega) = (1, \omega, \dots, \omega^s)^\top \tag{6.2}$$

and call ϕ_s a Vandermonde vector. We first study the following nonlinear approximation problem in the Vandermonde space,

$$\min_{\hat{a}_j, \hat{d}_j \in \mathbb{R}, |\hat{d}_j| \leq d, j=1, \dots, k} \left\| \sum_{j=1}^k \hat{a}_j \phi_{2k}(\hat{d}_j) - \sum_{j=1}^{k+1} a_j \phi_{2k}(d_j) \right\|_2, \tag{6.3}$$

where d_j 's are bounded complex numbers with a minimum separation distance. We derive a sharp lower bound on the above approximation in Theorem 6.2. In addition, we also investigate the stability of recovering d_j 's by \hat{d}_j 's from the condition

$$\left\| \sum_{j=1}^k \hat{a}_j \phi_{2k-1}(\hat{d}_j) - \sum_{j=1}^k a_j \phi_{2k-1}(d_j) \right\|_2 < \sigma, \tag{6.4}$$

and establish Theorem 6.3. These results are closely related to problem (6.1) and hence are also closely related to the proofs of the main results of this paper.

6.1 Notation and Preliminaries

We introduce some notation and preliminaries. We denote the Vandermonde matrix by

$$V_s(k) = \begin{pmatrix} 1 & \dots & 1 \\ d_1 & \dots & d_k \\ \vdots & \ddots & \vdots \\ d_1^s & \dots & d_k^s \end{pmatrix} = \left(\phi_s(d_1) \ \phi_s(d_2) \ \dots \ \phi_s(d_k) \right). \tag{6.5}$$

For a real matrix or a vector A , we denote by A^\top its transpose and by A^* its conjugate transpose.

We first present some basic properties of Vandermonde matrices.

Lemma 6.1 For k distinct complex numbers d_j 's, we have

$$\|V_{k-1}(k)^{-1}\|_\infty \leq \max_{1 \leq i \leq k} \prod_{1 \leq p \leq k, p \neq i} \frac{1 + |d_p|}{|d_i - d_p|},$$

where $V_{k-1}(k)$ is the Vandermonde matrix $V_{k-1}(k)$ defined as in (6.5).

Proof See Theorem 1 in [26]. □

As a consequence, we directly have the following corollary.

Corollary 6.2 Let $d_{\min} = \min_{i \neq j} |d_i - d_j|$ and assume that $\max_{i=1, \dots, k} |d_i| \leq d$. Then

$$\|V_{k-1}(k)^{-1}\|_\infty \leq \frac{(1 + d)^{k-1}}{(d_{\min})^{k-1}}.$$

Lemma 6.3 For distinct $d_1, \dots, d_k \in \mathbb{C}$, define the Vandermonde matrices $V_{k-1}(k)$, $V_s(k)$ as in (6.5) with $s \geq k - 1$. Then the following estimate on their singular values holds:

$$\frac{1}{\sqrt{k} \|V_{k-1}(k)^{-1}\|_\infty} \leq \frac{1}{\|V_{k-1}(k)^{-1}\|_2} \leq \sigma_{\min}(V_{k-1}(k)) \leq \sigma_{\min}(V_s(k)).$$

Proof The result holds by using properties of matrix norms. □

Denote by

$$S_{1k}^j := \left\{ \{\tau_1, \dots, \tau_j\} : \tau_p \in \{1, \dots, k\}, p = 1, \dots, j \text{ and } \tau_p \neq \tau_q, \text{ for } p \neq q \right\}.$$

Note that there is no order in $\{\tau_1, \dots, \tau_j\}$, i.e., $\{1, 2\}$ and $\{2, 1\}$ are the same sets. We then have the following decomposition of the Vandermonde matrix.

Proposition 6.4 The Vandermonde matrix $V_k(k)$ defined as in (6.5) can be reduced to the following form by using elementary column-addition operations, i.e.,

$$V_k(k)G(1) \cdots G(k-1)DQ(1) \cdots Q(k-1) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ v_{(k+1)1} & v_{(k+1)2} & \cdots & v_{(k+1)k} \end{pmatrix}, \tag{6.6}$$

where $G(1), \dots, G(k-1), Q(1), \dots, Q(k-1)$ are elementary column-addition matrices,

$$D = \text{diag} \left(1, \frac{1}{(d_2 - d_1)}, \dots, \frac{1}{\prod_{p=1}^{k-1} (d_k - d_p)} \right)$$

and

$$v_{(k+1)j} = (-1)^{k-j} \sum_{\{\tau_1, \dots, \tau_{k+1-j}\} \in S_{1k}^{k+1-j}} d_{\tau_1} \cdots d_{\tau_{k+1-j}}. \tag{6.7}$$

Proof See ‘‘Appendix B’’ in [46]. □

Lemma 6.5 For an $s \times k$ complex matrix A of rank k with $s > k$, let V be the space spanned by columns of A and V^\perp be the orthogonal complement of V . Denote by P_{V^\perp} the orthogonal projection to V^\perp , and set $D = (A, v)$. We have

$$\min_{a \in \mathbb{C}^k} \|Aa - v\|_2 = \|P_{V^\perp}(v)\|_2 = \sqrt{\frac{\det(D^*D)}{\det(A^*A)}}.$$

Proof See Lemma 1 in [45]. □

Lemma 6.6 We have

$$\sqrt{\frac{\det(V_k(k)^* V_k(k))}{\det(V_{k-1}(k)^* V_{k-1}(k))}} = \sqrt{\sum_{j=0}^k |v_j|^2}, \tag{6.8}$$

where $V_s(k)$ is defined as in (6.5) and $v_j = \sum_{\{\tau_1, \dots, \tau_j\} \in S_{1k}^j} d_{\tau_1} \cdots d_{\tau_j}$. Especially, if $|d_j| < d, j = 1, \dots, k$, then

$$\sqrt{\frac{\det(V_k(k)^* V_k(k))}{\det(V_{k-1}(k)^* V_{k-1}(k))}} \leq (1 + d)^k. \tag{6.9}$$

Proof Note that in Proposition 6.4, all the elementary column-addition matrices have unit determinant. As a result, $\det(V_k(k)^* V_k(k)) = \frac{\det(F^*F)}{\det(D^*D)}$, where F is the matrix in the right-hand side of (6.6), and D is the diagonal matrix in Proposition 6.4. A direct calculation shows that $\det(F^*F) = \sum_{j=0}^k |v_j|^2$, where we use (6.7). On the other hand, $V_{k-1}(k)$ is a standard Vandermonde matrix and we have $\det(V_{k-1}(k)^* V_{k-1}(k)) = \frac{1}{\det(D^*D)}$. Combining these results, (6.8) follows. The last statement can be derived from (6.8) and the estimate that

$$\sqrt{\sum_{j=0}^k |v_j|^2} \leq \sum_{j=0}^k |v_j| \leq \sum_{j=0}^k \binom{k}{j} d^j = (1 + d)^k.$$

□

For reader's convenience, we finally present two auxiliary lemmas. For positive integers p, q and complex numbers $z_1, \dots, z_p, \hat{z}_1, \dots, \hat{z}_q$, we define the vector

$$\eta_{p,q}(z_1, \dots, z_p, \hat{z}_1, \dots, \hat{z}_q) = \begin{pmatrix} |z_1 - \hat{z}_1| \cdots |z_1 - \hat{z}_q| \\ |z_2 - \hat{z}_1| \cdots |z_2 - \hat{z}_q| \\ \vdots \\ |z_p - \hat{z}_1| \cdots |z_p - \hat{z}_q| \end{pmatrix}. \tag{6.10}$$

The following two properties of $\eta_{p,q}$ hold.

Lemma 6.7 For complex numbers d_j, \hat{d}_j 's, we have the following estimate

$$\left\| \eta_{k+1,k}(d_1, \dots, d_{k+1}, \hat{d}_1, \dots, \hat{d}_k) \right\|_{\infty} \geq \left(\frac{d_{\min}}{2} \right)^k,$$

where $d_{\min} = \min_{j \neq p} |d_j - d_p|$ and $\eta_{k+1,k}(d_1, \dots, d_{k+1}, \hat{d}_1, \dots, \hat{d}_k)$ is defined as in (6.10).

Proof Because we have $k + 1$ d_j 's and only k \hat{d}_j 's, there must exist one d_{j_0} so that

$$|d_{j_0} - \hat{d}_j| \geq \frac{d_{\min}}{2}, \quad j = 1, \dots, k.$$

Then the estimate in the lemma follows. □

Lemma 6.8 Let $d_j, \hat{d}_j \in \mathbb{C}, j = 1, \dots, k$ satisfy $|d_j|, |\hat{d}_j| \leq d$. Assume that

$$\left\| \eta_{k,k}(d_1, \dots, d_k, \hat{d}_1, \dots, \hat{d}_k) \right\|_{\infty} < \epsilon, \tag{6.11}$$

where $\eta_{k,k}(\dots)$ is defined as in (6.10), and that

$$d_{\min} = \min_{p \neq q} |d_p - d_q| \geq 2\epsilon^{\frac{1}{k}}. \tag{6.12}$$

Then after reordering d_j 's, we have

$$\left| \hat{d}_j - d_j \right| < \frac{d_{\min}}{2}, \quad j = 1, \dots, k, \tag{6.13}$$

and moreover

$$\left| \hat{d}_j - d_j \right| \leq \left(\frac{2}{d_{\min}} \right)^{k-1} \epsilon, \quad j = 1, \dots, k. \tag{6.14}$$

Proof See ‘‘Appendix A’’. □

6.2 Lower-Bound for the Approximation Problem (6.3)

In this section, we derive a lower-bound for the nonlinear approximation problem (6.3). We first consider a special case.

Theorem 6.1 *Let $k \geq 1$ and $\hat{d}_1, \dots, \hat{d}_k$ be k distinct complex numbers with $|\hat{d}_j| \leq \hat{d}, 1 \leq j \leq k$. Define $A := (\phi_k(\hat{d}_1), \dots, \phi_k(\hat{d}_k))$, where $\phi_k(\hat{d}_j)$'s are defined as in (6.2). Let V be the k -dimensional space spanned by the column vectors of A , and let V^\perp be the one-dimensional orthogonal complement of V in \mathbb{C}^{k+1} . Let P_{V^\perp} be the orthogonal projection onto V^\perp in \mathbb{C}^{k+1} . Then we have*

$$\min_{a \in \mathbb{C}^k} \|Aa - \phi_k(x)\|_2 = \|P_{V^\perp}(\phi_k(x))\|_2 = |v^* \phi_k(x)| \geq \frac{1}{(1 + \hat{d})^k} \left| \prod_{j=1}^k (x - \hat{d}_j) \right|,$$

where v is a unit vector in V^\perp and v^* is its conjugate transpose.

Proof By Lemma 6.5, it follows that

$$\min_{a \in \mathbb{C}^k} \|Aa - \phi_k(x)\|_2 = \sqrt{\frac{\det(D^*D)}{\det(A^*A)}},$$

where $D = (\phi_k(\hat{d}_1), \dots, \phi_k(\hat{d}_k), \phi_k(x))$. Denote $\tilde{A} = (\phi_{k-1}(\hat{d}_1), \dots, \phi_{k-1}(\hat{d}_k))$. By (6.9), we have

$$\sqrt{\frac{\det(A^*A)}{\det(\tilde{A}^*\tilde{A})}} \leq (1 + \hat{d})^k.$$

Therefore,

$$\min_{a \in \mathbb{C}^k} \|Aa - \phi_k(x)\|_2 \geq \frac{1}{(1 + \hat{d})^k} \sqrt{\frac{\det(D^*D)}{\det(\tilde{A}^*\tilde{A})}}.$$

Note that D and \tilde{A} are square Vandermonde matrices. We can use the determinant formula to derive that

$$\begin{aligned} \min_{a \in \mathbb{C}^k} \|Aa - \phi_k(x)\|_2 &\geq \frac{1}{(1 + \hat{d})^k} \frac{|\prod_{1 \leq t < p \leq k} (\hat{d}_t - \hat{d}_p) \prod_{q=1}^k (x - \hat{d}_q)|}{|\prod_{1 \leq t < p \leq k} (\hat{d}_t - \hat{d}_p)|} \\ &= \frac{1}{(1 + \hat{d})^k} |\prod_{j=1}^k (x - \hat{d}_j)|. \end{aligned}$$

This completes the proof of the theorem. □

We now consider the approximation problem (6.3).

Theorem 6.2 Let $k \geq 1$. Assume $(k + 1)$ different complex numbers $d_j \in \mathbb{C}$, $j = 1, \dots, k + 1$ with $|d_j| \leq d$ and $(k + 1)$ $a_j \in \mathbb{C}$ with $|a_j| \geq m_{\min}$. Let $d_{\min} := \min_{j \neq p} |d_j - d_p|$. For $q \leq k$, let $\hat{a}(q) = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_q)^\top$, $a = (a_1, a_2, \dots, a_{k+1})^\top$, and

$$\hat{A}(q) = (\phi_{2k}(\hat{d}_1), \dots, \phi_{2k}(\hat{d}_q)), \quad A = (\phi_{2k}(d_1), \dots, \phi_{2k}(d_{k+1})),$$

where $\phi_{2k}(z)$ is defined as in (6.2). Then

$$\min_{\hat{a}_p, \hat{d}_p \in \mathbb{C}, |\hat{d}_p| \leq \hat{d}, p=1, \dots, q} \|\hat{A}(q)\hat{a}(q) - Aa\|_2 \geq \frac{m_{\min}(d_{\min})^{2k}}{2^k(1+d)^k(1+\hat{d})^k}.$$

Proof Step 1 Note that for $q < k$, we have

$$\begin{aligned} & \min_{\hat{a}_p, \hat{d}_p \in \mathbb{C}, |\hat{d}_p| \leq d, p=1, \dots, q} \|\hat{A}(q)\hat{a}(q) - Aa\|_2 \\ & \geq \min_{\hat{a}_p, \hat{d}_p \in \mathbb{C}, |\hat{d}_p| \leq d, p=1, \dots, k} \|\hat{A}(k)\hat{a}(k) - Aa\|_2. \end{aligned}$$

Hence we need only to consider the case when $q = k$. It then suffices to show that for any given $\hat{d}_j \in \mathbb{C}$, $|\hat{d}_j| \leq \hat{d}$, $j = 1, \dots, k$, the following holds

$$\min_{\hat{a}_p \in \mathbb{C}, p=1, \dots, k} \|\hat{A}(k)\hat{a}(k) - Aa\|_2 \geq \frac{m_{\min}(d_{\min})^{2k}}{2^k(1+d)^k(1+\hat{d})^k}. \tag{6.15}$$

So we fix $\hat{d}_1, \dots, \hat{d}_k$ in our subsequent argument.

Step 2 For $l = 0, \dots, k$, we define the following partial matrices

$$\hat{A}_l = \begin{pmatrix} \hat{d}_1^l & \dots & \hat{d}_k^l \\ \hat{d}_1^{l+1} & \dots & \hat{d}_k^{l+1} \\ \vdots & \vdots & \vdots \\ \hat{d}_1^{l+k} & \dots & \hat{d}_k^{l+k} \end{pmatrix}, \quad A_l = \begin{pmatrix} (d_1)^l & \dots & (d_{k+1})^l \\ (d_1)^{l+1} & \dots & (d_{k+1})^{l+1} \\ \vdots & \vdots & \vdots \\ (d_1)^{l+k} & \dots & (d_{k+1})^{l+k} \end{pmatrix}.$$

It is clear that for all l ,

$$\min_{\hat{a}(k) \in \mathbb{C}^k} \|\hat{A}(k)\hat{a}(k) - Aa\|_2 \geq \min_{\hat{a} \in \mathbb{C}^k} \|\hat{A}_l\hat{a} - A_l a\|_2. \tag{6.16}$$

Step 3 For each l , observe that $\hat{A}_l = \hat{A}_0 \text{diag}(\hat{d}_1^l, \dots, \hat{d}_k^l)$, $A_l = A_0 \text{diag}(d_1^l, \dots, d_{k+1}^l)$, and thus

$$\min_{\hat{a} \in \mathbb{C}^k} \|\hat{A}_l\hat{a} - A_l a\|_2 \geq \min_{\hat{\alpha}_l \in \mathbb{C}^k} \|A_0 \hat{\alpha}_l - A_0 \alpha_l\|_2, \tag{6.17}$$

where $\alpha_l = (a_1(d_1)^l, \dots, a_{k+1}(d_{k+1})^l)^\top$. Let V be the space spanned by the column vectors of A_0 . Then the dimension of V is k , and the dimension of V^\perp , the orthogonal complement of V in \mathbb{C}^{k+1} , is one. Let P_{V^\perp} be the orthogonal projection onto V^\perp . Note that $\|P_{V^\perp}u\|_2 = |v^*u|$ for $u \in \mathbb{C}^{k+1}$, where v is a unit vector in V^\perp and v^* is its conjugate transpose. We have

$$\begin{aligned} \min_{\hat{\alpha}_l \in \mathbb{C}^k} \|\hat{A}_0 \hat{\alpha}_l - A_0 \alpha_l\|_2 &= \|P_{V^\perp}(A_0 \alpha_l)\|_2 = |v^* A_0 \alpha_l| \\ &= \left| \sum_{j=1}^{k+1} a_j (d_j)^l v^* \phi_k(d_j) \right| = |\beta_l|, \end{aligned} \tag{6.18}$$

where

$$\beta_l = \sum_{j=1}^{k+1} a_j (d_j)^l v^* \phi_k(d_j), \quad \text{for } l = 0, 1, \dots, k.$$

Step 4 Denote $\beta = (\beta_0, \dots, \beta_k)^\top$. We have $B\hat{\eta} = \beta$, where

$$B = \begin{pmatrix} a_1 & a_2 & \cdots & a_{k+1} \\ a_1 d_1 & a_2 d_2 & \cdots & a_{k+1} d_{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ a_1 (d_1)^k & a_2 (d_2)^k & \cdots & a_{k+1} (d_{k+1})^k \end{pmatrix}, \quad \hat{\eta} = \begin{pmatrix} v^* \phi_k(d_1) \\ v^* \phi_k(d_2) \\ \vdots \\ v^* \phi_k(d_{k+1}) \end{pmatrix}.$$

Corollary 6.2 yields

$$\|\hat{\eta}\|_\infty = \|B^{-1}\beta\|_\infty \leq \|B^{-1}\|_\infty \|\beta\|_\infty \leq \frac{(1+d)^k}{m_{\min}(d_{\min})^k} \|\beta\|_\infty.$$

On the other hand, applying Theorem 6.1 to each term $|v^* \phi_k(d_j)|$, $j = 1, 2, \dots, k+1$, we have

$$\|\hat{\eta}\|_\infty \geq \frac{1}{(1+\hat{d})^k} \|\eta_{k+1,k}(d_1, \dots, d_{k+1}, \hat{d}_1, \dots, \hat{d}_k)\|_\infty,$$

where $\eta_{k+1,k}(\dots)$ is defined as in (6.10). Combining this inequality with Lemma 6.7, we get

$$\|\hat{\eta}\|_\infty \geq \frac{(d_{\min})^k}{2^k (1+\hat{d})^k}.$$

Then it follows that

$$\|\beta\|_\infty \geq \frac{m_{\min}(d_{\min})^{2k}}{2^k (1+d)^k (1+\hat{d})^k}.$$

Therefore, recalling (6.16)–(6.18), we arrive at

$$\begin{aligned} \min_{\hat{a}(k) \in \mathbb{C}^k} \|\hat{A}(k)\hat{a}(k) - Aa\|_2 &\geq \max_{0 \leq l \leq k} \min_{\hat{a} \in \mathbb{C}^k} \|\hat{A}_l \hat{a} - A_l a\|_2 \\ &= \max_{0 \leq l \leq k} |\beta_l| = \|\beta\|_\infty \geq \frac{m_{\min}(d_{\min})^{2k}}{2^k(1+d)^k(1+\hat{d})^k}. \end{aligned}$$

This proves (6.15) and hence the theorem. \square

6.3 Stability of the Approximation Problem (6.4)

In the section, we present a stability result for the approximation problem (6.4).

Theorem 6.3 *Let $k \geq 1$. Assume k different complex numbers $d_j \in \mathbb{C}$, $j = 1, \dots, k$ with $|d_j| \leq d$ and k $a_j \in \mathbb{C}$ with $|a_j| \geq m_{\min}$. Let $d_{\min} := \min_{p \neq q} |d_p - d_q|$. Assume that $\hat{d}_j \in \mathbb{C}$, $j = 1, \dots, k$ with $|\hat{d}_j| \leq d$ satisfy*

$$\|\hat{A}\hat{a} - Aa\|_2 < \sigma,$$

where $\hat{a} = (\hat{a}_1, \dots, \hat{a}_k)^\top$, $a = (a_1, \dots, a_k)^\top$, and

$$\hat{A} = (\phi_{2k-1}(\hat{d}_1), \dots, \phi_{2k-1}(\hat{d}_k)), \quad A = (\phi_{2k-1}(d_1), \dots, \phi_{2k-1}(d_k)).$$

Then

$$\left\| \eta_{k,k}(d_1, \dots, d_k, \hat{d}_1, \dots, \hat{d}_k) \right\|_\infty < \frac{(1+d)^{2k-1} \sigma}{d_{\min}^{k-1} m_{\min}}.$$

Proof Since $\|\hat{A}\hat{a} - Aa\|_2 < \sigma$, we have

$$\min_{\hat{\alpha} \in \mathbb{C}^k} \|\hat{A}\hat{\alpha} - Aa\|_2 < \sigma,$$

and hence

$$\max_{0 \leq l \leq k-1} \min_{\hat{\alpha} \in \mathbb{C}^k} \|\hat{A}_l \hat{\alpha} - A_l a\|_2 \leq \min_{\hat{\alpha} \in \mathbb{C}^k} \|\hat{A}\hat{\alpha} - Aa\|_2 < \sigma, \quad (6.19)$$

where

$$\hat{A}_l = \begin{pmatrix} \hat{d}_1^l & \cdots & \hat{d}_k^l \\ \hat{d}_1^{l+1} & \cdots & \hat{d}_k^{l+1} \\ \vdots & \vdots & \vdots \\ \hat{d}_1^{l+k} & \cdots & \hat{d}_k^{l+k} \end{pmatrix}, \quad A_l = \begin{pmatrix} d_1^l & \cdots & d_k^l \\ d_1^{l+1} & \cdots & d_k^{l+1} \\ \vdots & \vdots & \vdots \\ d_1^{l+k} & \cdots & d_k^{l+k} \end{pmatrix}.$$

For each l , from the decomposition $\hat{A}_l = \hat{A}_0 \text{diag}(\hat{d}_1^l, \dots, \hat{d}_k^l)$, $A_l = A_0 \text{diag}((d_1)^l, \dots, (d_k)^l)$, we get

$$\min_{\hat{\alpha} \in \mathbb{C}^k} \|\hat{A}_l \hat{\alpha} - A_l a\|_2 \geq \min_{\hat{\alpha}_l \in \mathbb{C}^k} \|\hat{A}_0 \hat{\alpha}_l - A_0 \alpha_l\|_2, \tag{6.20}$$

where $\alpha_l = (a_1(d_1)^l, \dots, a_k(d_k)^l)^\top$. Let V be the space spanned by the column vectors of \hat{A}_0 . Then the dimension of V is k , and V^\perp , the orthogonal complement of V in \mathbb{C}^{k+1} is of dimension one. We let v be a unit vector in V^\perp and let P_{V^\perp} be the orthogonal projection onto V^\perp . Similarly to (6.18), we have

$$\min_{\hat{\alpha}_l \in \mathbb{C}^k} \|\hat{A}_0 \hat{\alpha}_l - A_0 \alpha_l\|_2 = \|P_{V^\perp}(A_0 \alpha_l)\|_2 = |v^* A_0 \alpha_l| = \left| \sum_{j=1}^k a_j (d_j)^l v^* \phi_k(d_j) \right| = |\beta_l|, \tag{6.21}$$

where $\beta_l = \sum_{j=1}^k a_j (d_j)^l v^* \phi_k(d_j)$. Let $\beta = (\beta_0, \dots, \beta_{k-1})^\top$. Moreover, similarly to Step 4 in the proof of Theorem 6.2, we have

$$\left\| \eta_{k,k}(d_1, \dots, d_k, \hat{d}_1, \dots, \hat{d}_k) \right\|_\infty \leq \frac{(1+d)^{2k-1}}{m_{\min}(d_{\min})^{k-1}} \|\beta\|_\infty.$$

On the other hand, (6.19)–(6.21) indicate that $\|\beta\|_\infty < \sigma$. Hence, we obtain that

$$\left\| \eta_{k,k}(d_1, \dots, d_k, \hat{d}_1, \dots, \hat{d}_k) \right\|_\infty \leq \frac{(1+d)^{2k-1}}{(d_{\min})^{k-1}} \frac{\sigma}{m_{\min}}.$$

This completes the proof. □

7 Conclusions and Future Work

In this paper, we have improved the estimates of resolution limits in two-dimensional super-resolution problems. We also theoretically demonstrate the optimal performance of a sparsity-promoting algorithm. Leveraging the new techniques in the proof, we have proposed a coordinate-combination-based model order detection algorithm and a coordinate-combination-based MUSIC algorithm for DOA estimation in two dimensions. The superiority of the introduced algorithms was demonstrated both theoretically or numerically.

Our work is also a start of many new topics. Firstly, one could extend the techniques to three- and k -dimensional spaces to improve the resolution estimates in higher dimensional super-resolution problems. Secondly, the idea of coordinate-combination could inspire new algorithms for two-dimensional DOA estimations in the case of multiple snapshots. These works will be presented in a near future.

Funding Open access funding provided by Swiss Federal Institute of Technology Zurich

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Proof of Lemma 6.8

Proof Step 1 We claim that for each $\hat{d}_p, 1 \leq p \leq k$, there exists one d_j such that $|\hat{d}_p - d_j| < \frac{d_{\min}}{2}$. By contradiction, suppose that there exists p_0 such that $|d_j - \hat{d}_{p_0}| \geq \frac{d_{\min}}{2}$ for all $1 \leq j \leq k$. Observe that

$$\begin{aligned} &\eta_{k,k}(d_1, \dots, d_k, \hat{d}_1, \dots, \hat{d}_k) \\ &= \text{diag} \left(|d_1 - \hat{d}_{p_0}|, \dots, |d_k - \hat{d}_{p_0}| \right) \\ &\eta_{k,k-1}(d_1, \dots, d_k, \hat{d}_1, \dots, \hat{d}_{p_0-1}, \hat{d}_{p_0+1}, \dots, \hat{d}_k). \end{aligned}$$

We write

$$\begin{aligned} \eta_{k,k} &= \eta_{k,k}(d_1, \dots, d_k, \hat{d}_1, \dots, \hat{d}_k) \quad \text{and} \\ \eta_{k,k-1} &= \eta_{k,k-1}(d_1, \dots, d_k, \hat{d}_1, \dots, \hat{d}_{p_0-1}, \hat{d}_{p_0+1}, \dots, \hat{d}_k). \end{aligned}$$

Using Lemma 6.7, we have

$$\|\eta_{k,k}\|_\infty \geq \frac{d_{\min}}{2} \|\eta_{k,k-1}\|_\infty \geq \left(\frac{d_{\min}}{2}\right)^k \geq \epsilon,$$

where we have used (6.12) in the last inequality above. This contradicts (6.11) and hence proves our claim.

Step 2 We claim that for each $d_j, 1 \leq j \leq k$, there exists one and only one \hat{d}_p such that

$$|d_j - \hat{d}_p| < \frac{d_{\min}}{2}.$$

It suffices to show that for each $d_j, 1 \leq j \leq k$, there is only one \hat{d}_p such that $|d_j - \hat{d}_p| < \frac{d_{\min}}{2}$. By contradiction, suppose that there exist p_1, p_2 , and j_0 such that $|d_{j_0} - \hat{d}_{p_1}| < \frac{d_{\min}}{2}, |d_{j_0} - \hat{d}_{p_2}| < \frac{d_{\min}}{2}$. Then for all $j \neq j_0$, we have

$$\left| (d_j - \hat{d}_{p_1})(d_j - \hat{d}_{p_2}) \right| \geq \frac{(d_{\min})^2}{4}. \tag{A.1}$$

Similarly to the argument in Step 1, we separate the factors involving $\hat{d}_{p_1}, \hat{d}_{p_2}, d_{j_0}$ from $\eta_{k,k}$ and consider

$$\eta_{k-1,k-2} = \eta_{k-1,k-2}(d_1, \dots, d_{j_0-1}, d_{j_0+1}, \dots, d_k, \hat{d}_1, \dots, \hat{d}_{p_1-1}, \hat{d}_{p_1+1}, \dots, \hat{d}_{p_2-1}, \hat{d}_{p_2+1}, \dots, \hat{d}_k).$$

Note that the components of $\eta_{k-1,k-2}$ differ from those of $\eta_{k,k}$ only by the factors $|(d_j - \hat{d}_{p_1})(d_j - \hat{d}_{p_2})|$ for $j = 1, \dots, j_0 - 1, j_0 + 1, \dots, k$. We can show that

$$\|\eta_{k,k}\|_\infty \geq \frac{(d_{\min})^2}{4} \|\eta_{k-1,k-2}\|_\infty \geq \epsilon,$$

where we have used Lemma 6.7 and (6.12) for establishing the last inequality above. This contradicts (6.11) and hence proves our claim.

Step 3 By the result in Step 2, we can reorder \hat{d}_j 's to get

$$|\hat{d}_j - d_j| < \frac{d_{\min}}{2}, \quad j = 1, \dots, k.$$

We now prove (6.14). It is clear that $|\hat{d}_p - d_j| > \frac{d_{\min}}{2}, p \neq j$. Thus

$$|(d_j - \hat{d}_1) \cdots (d_j - \hat{d}_k)| > |d_j - d_j| \left(\frac{d_{\min}}{2} \right)^{k-1}, \quad j = 1, 2, \dots, k. \quad (\text{A.2})$$

Further, we get

$$|d_j - \hat{d}_j| < \left(\frac{2}{d_{\min}} \right)^{k-1} \|\eta_{k,k}\|_\infty \leq \left(\frac{2}{d_{\min}} \right)^{k-1} \epsilon, \quad j = 1, 2, \dots, k.$$

This completes the proof of the lemma. \square

References

1. Ernst Abbe. Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung. *Archiv für mikroskopische Anatomie*, 9(1):413–468, 1873.
2. Hirotugu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pp. 199–213. Springer, 1998.
3. Hirotugu Akaike. A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike*, pp. 215–222. Springer, 1974.
4. Andrey Akinshin, Dmitry Batenkov, and Yosef Yomdin. Accuracy of spike-train fourier reconstruction for colliding nodes. In *2015 International Conference on Sampling Theory and Applications (SampTA)*, pp. 617–621. IEEE, 2015.
5. Habib Ammari. *An introduction to mathematics of emerging biomedical imaging*. Mathematics & Applications.
6. Habib Ammari, Josselin Garnier, Wenjia Jing, Hyeonbae Kang, Mikyoung Lim, Knut Sølna, and Han Wang. *Mathematical and statistical methods for multistatic imaging*. Lecture Notes in Mathematics.
7. Jean-Marc Azais, Yohann De Castro, and Fabrice Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.

8. Dmitry Batenkov, Laurent Demanet, Gil Goldman, and Yosef Yomdin. Conditioning of partial nonuniform Fourier matrices with clustered nodes. *SIAM Journal on Matrix Analysis and Applications*, 41(1):199–220, 2020.
9. Dmitry Batenkov, Gil Goldman, and Yosef Yomdin. Super-resolution of near-colliding point sources. *Information and Inference: A Journal of the IMA*, 2020.
10. Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych, Juan S Bonifacio, Michael W Davidson, Jennifer Lippincott-Schwartz, and Harald F Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *science*, 313(5793):1642–1645, 2006.
11. Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Commun. Pure Appl. Math.*, 67(6):906–956, 2014.
12. Sitan Chen and Ankur Moitra. Algorithmic foundations for the diffraction limit. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 490–503, 2021.
13. Weiguo Chen, Kon Max Wong, and James P Reilly. Detection of the number of signals: A predicted eigen-threshold approach. *IEEE Transactions on Signal Processing*, 39(5):1088–1098, 1991.
14. Y-H Chen and C-H Chen. Direction-of-arrival and frequency estimations for narrowband sources using two single rotation invariance algorithms with the marked subspace. In *IEE Proceedings F (Radar and Signal Processing)*, volume 139, pp. 297–300. IET, 1992.
15. Edward AK Cohen, Anish V Abraham, Sreevidhya Ramakrishnan, and Raimund J Ober. Resolution limit of image analysis algorithms. *Nature Communications*, 10(1):1–12, 2019.
16. Maxime Ferreira Da Costa and Yuejie Chi. On the stable resolution limit of total variation regularization for spike deconvolution. *IEEE Transactions on Information Theory*, 66(11):7237–7252, 2020.
17. Geoffrey De Villiers and E Roy Pike. *The Limits of Resolution*. CRC Press, Boca Raton, 2016.
18. JE Fernandez Del Rio and MF Catedra-Perez. The matrix pencil method for two-dimensional direction of arrival estimation employing an l-shaped array. *IEEE Transactions on Antennas and Propagation*, 45(11):1693–1694, 1997.
19. Laurent Demanet and Nam Nguyen. The recoverability limit for superresolution via sparsity. [arXiv:1502.01385](https://arxiv.org/abs/1502.01385), 2015.
20. Justin Demmerle, Eva Wegel, Lothar Schermelleh, and Ian M Dobbie. Assessing resolution in super-resolution imaging. *Methods*, 88:3–10, 2015.
21. Arnold Jan Den Dekker and A Van den Bos. Resolution: a survey. *JOSA A*, 14(3):547–557, 1997.
22. Quentin Denoyelle, Vincent Duval, and Gabriel Peyré. Support recovery for sparse super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, 23(5):1153–1194, 2017.
23. G. Toraldo Di Francia. Resolving power and information. *JOSA*, 45(7):497–501, 1955.
24. David L. Donoho. Superresolution via sparsity constraints. *SIAM Journal on Mathematical Analysis*, 23(5):1309–1331, 1992.
25. Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.
26. Walter Gautschi. On inverses of vandermonde and confluent vandermonde matrices. *Numerische Mathematik*, 4(1):117–123, 1962.
27. Keyong Han and Arye Nehorai. Improved source number detection and direction estimation with nested arrays and ULAs using jackknifing. *IEEE Transactions on Signal Processing*, 61(23):6118–6128, 2013.
28. Zhaoshui He, Andrzej Cichocki, Shengli Xie, and Kyuwan Choi. Detecting the number of clusters in n-way probabilistic clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):2006–2021, 2010.
29. Stefan W Hell and Jan Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters*, 19(11):780–782, 1994.
30. C Helstrom. The detection and resolution of optical signals. *IEEE Transactions on Information Theory*, 10(4):275–287, 1964.
31. Carl W Helstrom. Detection and resolution of incoherent objects by a background-limited optical system. *JOSA*, 59(2):164–175, 1969.
32. Samuel T Hess, Thanu PK Girirajan, and Michael D Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical Journal*, 91(11):4258–4272, 2006.
33. William V Houston. A compound interferometer for fine structure work. *Physical Review*, 29(3):478, 1927.
34. Richard L Johnson and Gina E Miner. An operational system implementation of the ESPIRIT DF algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 27(1):159–166, 1991.

35. Shohei Kikuchi, Hiroyuki Tsuji, and Akira Sano. Pair-matching method for estimating 2-d angle of arrival with a cross-correlation matrix. *IEEE Antennas and Wireless Propagation Letters*, 5:35–40, 2006.
36. DN Lawley. Tests of significance for the latent roots of covariance and correlation matrices. *biometrika*, 43(1/2):128–136, 1956.
37. Weilin Li and Wenjing Liao. Stable super-resolution limit and smallest singular value of restricted Fourier matrices. *Applied and Computational Harmonic Analysis*, 51:118–156, 2021.
38. Weilin Li, Wenjing Liao, and Albert Fannjiang. Super-resolution limit of the ESPRIT algorithm. *IEEE Transactions on Information Theory*, 66(7):4593–4608, 2020.
39. Wenjing Liao. MUSIC for multidimensional spectral estimation: stability and super-resolution. *IEEE Transactions on Signal Processing*, 63(23):6395–6406, 2015.
40. Wenjing Liao and Albert C. Fannjiang. MUSIC for single-snapshot spectral estimation: Stability and super-resolution. *Applied and Computational Harmonic Analysis*, 40(1):33–67, 2016.
41. Ping Liu. *Mathematical Theory of Computational Resolution Limit and Efficient Fast Algorithms for Super-Resolution*. Hong Kong University of Science and Technology (Hong Kong), 2021.
42. Ping Liu and Habib Ammari. A mathematical theory of super-resolution and diffraction limit. [arXiv:2211.15208](https://arxiv.org/abs/2211.15208), 2022.
43. Ping Liu, Yanchen He, and Habib Ammari. A mathematical theory of resolution limits for super-resolution of positive sources. [arXiv:2211.13541](https://arxiv.org/abs/2211.13541), 2022.
44. Ping Liu and Hai Zhang. A mathematical theory of computational resolution limit in multi-dimensional spaces. *Inverse Problems*, 37(10):104001, 2021.
45. Ping Liu and Hai Zhang. A theory of computational resolution limit for line spectral estimation. *IEEE Transactions on Information Theory*, 67(7):4812–4827, 2021.
46. Ping Liu and Hai Zhang. A mathematical theory of computational resolution limit in one dimension. *Applied and Computational Harmonic Analysis*, 56:402–446, 2022.
47. Ping Liu and Hai Zhang. A measurement decoupling based fast algorithm for super-resolving point sources with multi-cluster structure. [arXiv:2204.00469](https://arxiv.org/abs/2204.00469), 2022.
48. Tsung-Hsien Liu and Jerry M Mendel. Azimuth and elevation direction finding using arbitrary array geometries. *IEEE Transactions on Signal Processing*, 46(7):2061–2065, 1998.
49. Leon B Lucy. Resolution limits for deconvolved images. *The Astronomical Journal*, 104:1260–1265, 1992.
50. Leon B Lucy. Statistical limits to super resolution. *Astronomy and Astrophysics*, 261:706, 1992.
51. Ankur Moitra. Super-resolution, extremal functions and the condition number of vandermonde matrices. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pp. 821–830. ACM, New York, 2015.
52. Veniamin I Morgenshtern. Super-resolution of positive sources on an arbitrarily fine grid. [arXiv:2005.06756](https://arxiv.org/abs/2005.06756), 2020.
53. Veniamin I. Morgenshtern and Emmanuel J. Candès. Super-resolution of positive sources: The discrete setup. *SIAM Journal on Imaging Sciences*, 9(1):412–444, 2016.
54. David W Pentico. Assignment problems: A golden anniversary survey. *European Journal of Operational Research*, 176(2):774–793, 2007.
55. Clarice Poon and Gabriel Peyré. Multidimensional sparse super-resolution. *SIAM Journal on Mathematical Analysis*, 51(1):1–44, 2019.
56. Sripad Ram, E Sally Ward, and Raimund J Ober. Beyond Rayleigh’s criterion: a resolution measure with application to single-molecule microscopy. *Proceedings of the National Academy of Sciences*, 103(12):4457–4462, 2006.
57. Lord Rayleigh. Xxxi. investigations in optics, with special reference to the spectroscope. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 8(49):261–274, 1879.
58. Jorma Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.
59. Michael J Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nature Methods*, 3(10):793–796, 2006.
60. Ralph Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986.
61. Arthur Schuster. *An introduction to the theory of optics*. E. Arnold, 1904.
62. Gideon Schwarz et al. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

63. Morteza Shahram and Peyman Milanfar. Imaging below the diffraction limit: a statistical analysis. *IEEE Transactions on Image Processing*, 13(5):677–689, 2004.
64. Morteza Shahram and Peyman Milanfar. Statistical analysis of achievable resolution in incoherent imaging. In *Signal and Data Processing of Small Targets 2003*, volume 5204, pp. 1–9. International Society for Optics and Photonics, 2004.
65. Morteza Shahram and Peyman Milanfar. On the resolvability of sinusoids with nearby frequencies in the presence of noise. *IEEE Transactions on Signal Processing*, 53(7):2579–2588, 2005.
66. Carroll Mason Sparrow. On spectroscopic resolving power. *The Astrophysical Journal*, 44:76, 1916.
67. Petre Stoica and Arye Nehorai. MUSIC, maximum likelihood, and Cramer-Rao bound. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(5):720–741, 1989.
68. AL Swindlehurst and THOMAS Kailath. Azimuth/elevation direction finding using regular array geometries. *IEEE Transactions on Aerospace and Electronic Systems*, 29(1):145–156, 1993.
69. Gongguo Tang. Resolution limits for atomic decompositions via Markov–Bernstein type inequalities. In *2015 International Conference on Sampling Theory and Applications (SampTA)*, pp. 548–552. IEEE, 2015.
70. Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht. Near minimax line spectral estimation. *IEEE Transactions on Information Theory*, 61(1):499–512, 2014.
71. Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *IEEE Transactions on Information Theory*, 59(11):7465–7490, 2013.
72. Harald Volkman. Ernst abbe and his work. *Applied Optics*, 5(11):1720–1731, 1966.
73. Feng Wang, Xiaowei Cui, Mingquan Lu, and Zhenming Feng. Decoupled 2d direction-of-arrival estimation based on sparse signal reconstruction. *EURASIP Journal on Advances in Signal Processing*, 2015(1):1–16, 2015.
74. Yung-Yi Wang, Liang-Cheng Lee, Shih-Jen Yang, and Jiunn-Tsair Chen. A tree structure one-dimensional based algorithm for estimating the two-dimensional direction of arrivals and its performance analysis. *IEEE Transactions on Antennas and Propagation*, 56(1):178–188, 2008.
75. Mati Wax and Thomas Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):387–392, 1985.
76. Mati Wax and Ilan Ziskind. Detection of the number of coherent signals by the mdl principle. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(8):1190–1196, 1989.
77. Volker Westphal, Silvio O Rizzoli, Marcel A Lauterbach, Dirk Kamin, Reinhard Jahn, and Stefan W Hell. Video-rate far-field optical nanoscopy dissects synaptic vesicle movement. *Science*, 320(5873):246–249, 2008.
78. C-C Yeh, J-H Lee, and Y-M Chen. Estimating two-dimensional angles of arrival in coherent source environment. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(1):153–155, 1989.
79. Nuri Yilmazer, Raul Fernandez-Recio, and Tapan K Sarkar. Matrix pencil method for simultaneously estimating azimuth and elevation angles of arrival along with the frequency of the incoming signals. *Digital Signal Processing*, 16(6):796–816, 2006.
80. Michael D Zoltowski, Martin Haardt, and Cherian P Mathews. Closed-form 2-d angle estimation with rectangular arrays in element space or beamspace via unitary esprit. *IEEE Transactions on Signal Processing*, 44(2):316–328, 1996.
81. Michael D Zoltowski and Demosthenis Stavrinos. Sensor array signal processing via a procrustes rotations based eigenanalysis of the esprit data pencil. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(6):832–861, 1989.