Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

Mathematical foundation of sparsity-based multi-illumination super-resolution

P. Liu and S. Yu and O. Sabet and L. Pelkmans and H. Ammari

Research Report No. 2022-05
February 2022

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Funding SNF: 200021--200307

Mathematical foundation of sparsity-based multi-illumination super-resolution*

Ping Liu[†] Sanghyeon Yu[‡] Ola Sabet[§] Lucas Pelkmans[‡]
Habib Ammari[†]

Abstract

It is well-known that the resolution of traditional optical imaging system is limited by the so-called Rayleigh resolution or diffraction limit, which is of several hundreds of nanometers. By employing fluorescence techniques, modern microscopic methods can resolve point scatterers separated by a distance much lower than the Rayleigh resolution limit. Localization-based fluorescence subwavelength imaging techniques such as PALM and STORM can achieve spatial resolution of several tens of nanometers. However, these techniques have limited temporal resolution as they require tens of thousands of exposures. Employing sparsity-based models and recovery algorithms is a natural way to reduce the number of exposures, and hence obtain high temporal resolution. Nevertheless, to date fluorescence techniques suffer from the trade-off between spatial and temporal resolutions.

In [34], a newly multi-illumination imaging technique called Brownian Excitation Amplitude Modulation microscopy (BEAM) is introduced. BEAM achieves a threefold resolution improvement by applying a compressive sensing recovery algorithm over only few frames. Motivated by BEAM, our aim in this paper is to pioneer the mathematical foundation for sparsity-based multi-illumination super-resolution. More precisely, we consider several diffraction-limited images from sample exposed to different illumination patterns and recover the source by considering the sparsest solution. We estimate the minimum separation distance between point scatterers so that they could be stably recovered. By this estimation of the resolution of the sparsity recovery, we reveal the dependence of the resolution on the cut-off frequency of the imaging system, the signal-to-noise ratio, the sparsity of point scatterers, and the incoherence of illumination patterns. Our theory particularly highlights the importance of the high incoherence of illumination patterns in enhancing the resolution. It also demonstrates that super-resolution can be

*This work was supported in part by the Swiss National Science Foundation grant number 200021–200307.

[†]Department of Mathematics, ETH Zürich, Rämistrasse 101, CH-8092 Zürich, Switzerland (ping.liu@sam.math.ethz.ch, habib.ammari@math.ethz.ch).

[‡]Department of Mathematics, Korea University, Seoul 02841, S. Korea (sanghyeon_yu@korea.ac.kr).

[§]Department of Molecular Life Sciences, University of Zürich, Switzerland (ola.sabet@mls.uzh.ch, lucas.pelkmans@mls.uzh.ch)

achieved using sparsity-based multi-illumination imaging with very few frames, whereby the spatio-temporal super-resolution becomes possible. BEAM can be viewed as the first experimental realization of our theory, which is demonstrated to hold in both the one- and two-dimensional cases.

Mathematics Subject Classification: 65R32, 42A10, 62H20, 78A46

Keywords: spatio-temporal sparsity-based super-resolution, Brownian Excitation Amplitude Modulation Microscopy, multi-illumination incoherent data, compressive sensing, minimal separation distance, approximation theory in Vandermonde space

Contents

1. Introduction

Super-resolution fluorescent microscopy has transformed many domains of biology. To date there are two far-field classes of techniques that lead to fluorescence-based microscopy with a resolution far beyond the Rayleigh diffraction limit [19]. The first class is generically referred to as super-resolved ensemble fluorophore microscopy and the second as super-resolved single fluorophore microscopy.

The first class of techniques can be implemented either by stimulated emission depletion (STED) of fluorescence from all molecules in a sample except those in a small region of the imaged biological sample or by structured illumination microscopy (SIM). STED builds on the deterministic transitions that either switch fluorescence on or off to reduce the emission volume [17, 18, 20, 43]. In SIM, interference patterns used in sample illumination lead to a twofold gain in resolution [15, 16, 35].

The second class of techniques is based on the a prior knowledge that the measurements at a given time are from single fluorescent molecules that are separated from each other by distances larger than the Rayleigh diffraction limit. This information is used to super-localize single molecules in an image, which means finding the position of each molecule to a precision better than the Rayleigh diffraction limit. Super-resolved single fluorophore microscopy relies on the stochastic switching of fluorophores in a time sequence to localize single molecules. It can be implemented either by photo-activated localization microscopy (PALM) [5, 21] or by stochastic optical reconstruction microscopy (STORM) [6, 33].

A major disadvantage of these two classes of techniques is that they suffer from the trade-off between the spatial and temporal resolutions, which makes live cell imaging quite challenging. On one hand, super-resolved single fluorophore microscopy techniques require hundreds of thousands of exposures. This is because in every frame, the diffraction-limited image of each emitter must be well separated from its neighbours, to enable the identification of its exact position. This inevitably leads to a long acquisition cycle, typically on the order of several minutes. Consequently, fast dynamics cannot be captured by these techniques. On the other hand, SIM techniques require only tens of frames (thus they are with high temporal resolution). But, their spatial resolution enhancement is limited by a factor of two.

Most previous works on enhancing the temporal resolution focused on improving the localization accuracy in PALM/STORM. Some of them (such as CS-STORM [44] and SPARCOM [39]) used compressive sensing (CS) recovery algorithms to reduce the number of measurements, but any PALM/STORM-based techniques inevitably suffers from the trade-off challenge. The trade-off originates from the fact that stochastic single molecule switching activates only a small part of the solution in each frame.

To better explain this, let us describe the mathematical problem for PALM/STORM. We activate the fluorescent solution by stochastic switching with T number of frames. Denote by $\rho_1, \rho_2, \cdots, \rho_T$ the sparse distribution of the activated fluorescent molecules (point scatterers). Then we collect the corresponding measured images $Y_1, Y_2, \cdots, Y_T$. Hence we have

$$S\rho_t := h \circledast \rho_t = Y_t, \quad t = 1, 2, \cdots, T,$$

where h is a blurring kernel and $\circledast$ is the convolution product. In CS-STORM, we apply compressive sensing to the following deconvolution problem for reconstructing the unknown ρ_t:

$$\min_{\rho_t} \|\rho_t\|_1 \quad \text{subject to} \quad \rho_t \geq 0 \text{ and } \|S\rho_t - Y_t\|_2 \leq \sigma, \tag{1.1}$$

where σ is the noise level. Then the super-resolved image can be obtained by

$$\rho = \sum_{t=1}^{T} \rho_t.$$

When the density of activated molecules in each single frame is small, then in average the point scatterers are well separated. Then it is easy to localize them by deconvolution procedure. But the lower the density of molecules, the higher the number of frames T. This is the spatio-temporal resolution trade-off of PALM/STORM-based approaches.

In [34], a novel imaging modality called Brownian Excitation Amplitude Modulation microscopy (BEAM) is introduced, which is based on speckle imaging and compressive sensing. On one hand, it reduces significantly the number of exposures by exposing the most part of the solution at each frame to the illumination pattern. On the other hand, it involves multiple incoherent illuminations of the biological sample and achieves super-resolution microscopy across both space and time from a sequence of diffraction-limited images and can capture fast dynamics of biological samples. Hence, BEAM outperforms the PALM/STORM-based techniques. Their two key ingredients are spatial sparsity and temporal incoherence. BEAM combines the sparsity of the point scatterers and the incoherence between the illumination patterns in different frames.

There are some related works to BEAM. The Blind-SIM [31] and RIM [26] use random speckle modulations but compressive sensing was not exploited there and so the spatial resolution enhancement is limited by a factor of two (they also require a large number of measurements). The Joint Sparse Recovery approach in [27] uses both random speckles and compressive sensing. But their inverse problem is formulated in MMV (multiple measurement vectors) form whose sensing matrix has no incoherence, which is not optimal for CS, and hence requires a large number of measurements.

Let us now briefly describe the inverse problem in BEAM. Suppose we have multiple speckle patterns $I_1, I_2, \cdots, I_T$ illuminating the sparse fluorescent solution and then collect the corresponding measured images $Y_1, Y_2, \cdots, Y_T$. Then we have

$$A_t\rho := h \circledast (I_t\rho) = Y_t, \quad t = 1, 2, \cdots, T,$$

where, as before, h is a blurring kernel. We apply compressive sensing to reconstruct the unknown ρ with estimated speckle patterns I_t [7, 12, 14, 29, 30]:

$$\min_{\rho} \|\rho\|_1 \quad \text{subject to} \quad \rho \geq 0 \text{ and } \|A\rho - Y\|_2 \leq \sigma, \tag{1.2}$$

where σ is the noise level, $A = (A_t)_{t=1,\ldots,T}$ is the sensing matrix, and $Y = (Y_1, \ldots, Y_T)^\top$ (with $\top$ denoting the transpose). Notice that the columns of A have a high degree of incoherence coming from the Brownian motion of the speckle patterns I_t. This incoherence in the sensing matrix is an optimal feature for compressive sensing to work properly. The sparsity prior in BEAM enhances the spatial resolution (beyond SIM's two-fold enhancement), and at the same time, the required number of measurements stays small since our sensing matrix satisfies CS requirement (incoherence). To the best of our knowledge, BEAM is the first compressive imaging approach satisfying the incoherence requirement, which is the key to overcome the trade-off barrier between the spatial and temporal resolutions.

BEAM can be then seen as the first experimental realization of spatio-temporal sparsity-based super-resolved imaging, where threefold resolution enhancement can be achieved by applying compressive sensing over only few frames. Motivated by BEAM, our aim in this paper is to pioneer the mathematical foundation of spatio-temporal sparsity-based super-resolution. We consider mathematical models similar to (1.2) but tackle instead the sparsest solution (l_0 pseudo-norm minimizer) under the measurement constraints. The sparsest solution is usually the one targeted in sparsity-based imaging and also in the general compressive sensing theory (using tractable convex l_1-minimization). Moreover, we consider that the values of the illumination patterns may not be known. Our main results (Theorems 2.1 and 3.1) consist in deriving lower bounds for the resolution enhancement in both the one- and two-dimensional cases. More precisely, we estimate the minimal separation distance for stable recovery of point scatterers from multi-illumination incoherent data. Our estimations reveal the dependence of the resolution enhancement on the cut-off frequency of the imaging system, the signal-to-noise ratio, the sparsity of the point scatterers, and more importantly on the incoherence of the illumination patterns. Our theory highlights the importance of incoherence in the illumination patterns and theoretically demonstrates the possibility of achieving super-resolution for sparsity-based multi-illumination imaging using very few frames.

It is worth emphasizing that there are many mathematical theories for estimating the stability of super-resolution in the single measurement case. To our knowledge, the first work was by Donoho [13]. He considered a grid setting where a discrete measure is supported on a lattice (spacing by Δ) and regularized by a so-called "Rayleigh index" d. He demonstrated that the minimax error for the recovery of the strength of the scatterer is bounded by $SRF^{\alpha}\sigma$ ($2d-1 \leq \alpha \leq 2d+1$) with σ being the noise level and the super-resolution factor $SRF = 1/(\Omega\Delta)$. Here, Ω is the cut-off frequency. Donoho's results emphasized the importance of sparsity (encoded in the Rayleigh index) in the super-resolution problem. In [10], the authors considered n-sparse scatterers supported on a grid and obtained sharper bounds ($\alpha = 2n-1$) using an estimate of the minimum singular value for the measurement matrix. The case of multi-clumps was considered in [3, 22] and similar minimax error estimations were derived. See also other related works for the understanding of resolution limit from the perceptive of sample complexity [8, 28]. In [1, 4], the authors considered the minimax error for recovering off-the-grid point scatterers. Based on an analysis of the "prony-type system", they derived bounds for both strength and location reconstructions of the point scatterers. More precisely, they showed that for $\sigma \lessapprox (SRF)^{-2p+1}$ where p is the number of point scatterers in a cluster, the minimax error for the strength and the location recoveries scale respectively as $(SRF)^{2p-1}\sigma$, $(SRF)^{2p-2}\sigma/\Omega$. Moreover, for the isolated non-cluster point scatterer, the corresponding minimax error for the strength and the location recoveries scale respectively as σ and σ/Ω.

Due to the popularity of sparse modeling and compressive sensing, many sparsity-promoting algorithms were proposed to address the super-resolution problem. In the groundbreaking work of Candès and Fernandez-Granda [7], it was demonstrated that off-the-grid sources can be exactly recovered from their low-pass Fourier coefficients by total variation minimization under a minimum separation condition. Other sparsity promoting methods include the BLASSO algorithm [2, 14, 32] and the atomic norm minimization method [41, 42]. These two algorithms were proved to be able to stably recover the sources under a minimum separation condition or a non-degeneracy condition. The resolution of these convex algorithms are limited by a distance of the order of the Rayleigh diffraction limit [9, 40] for recovering general signed point scatterers. But for the case of positive sources [12, 29, 30], there is no such limitation on the resolution and the performance of these algorithms could be nearly optimal.

More recently, to analyze the resolution for recovering multiple point scatterers, in [23–25] the authors defined "computational resolution limits" which characterize the minimum required distance between point scatterers so that their number and locations can be stably resolved under certain noise level. By developing a non-linear approximation theory in a so-called Vandermonde space, they derived bounds for computational resolution limits for a deconvolution problem [25] and a line spectral problem [24] (equivalent to the super-resolution problem considered here). In particular, they showed in [24] that the computational resolution limit for number and location recovery should be respectively $\frac{C_{\mathrm{num}}}{\Omega}(\frac{\sigma}{m_{\min}})^{\frac{1}{2n-2}}$ and $\frac{C_{\mathrm{supp}}}{\Omega}(\frac{\sigma}{m_{\min}})^{\frac{1}{2n-1}}$ where $C_{\mathrm{num}}, C_{\mathrm{supp}}$ are constants and $m_{\min}$ is the minimum strength of the point scatterers. Their results demonstrate that when the point scatterers are separated larger than $\frac{C_{\mathrm{supp}}}{\Omega}(\frac{\sigma}{m_{\min}})^{\frac{1}{2n-1}}$, we can stably recover the scatterer locations. Conversely, when the point scatterers are separated by a distance less than $O(\frac{C_{\mathrm{supp}}}{\Omega}(\frac{\sigma}{m_{\min}})^{\frac{1}{2n-1}})$, stably recovering the scatterer locations is impossible in the worst case. This resolution limit indicates that super-resolution is possible for the single measurement case but requires very high signal-to-noise ratio (according to the exponent $\frac{1}{2n-1}$). This explains why it is so hard to achieve super-resolution by single illumination. Therefore, we have to resort to multiple illuminations in order to super-resolve point scatterers.

As we have seen, the mathematics behind resolution limit for single illumination imaging is towards to be fully understood. Nevertheless, the multiple illumination case still lacks or even is without any mathematical foundation. Thus, our paper serves as a first step towards understanding the resolution limit (or performance) of multi-illumination imaging. We consider both the one- and two-dimensional cases. Our results demonstrate that the resolution for the multiple illumination imaging problem in the one-dimensional case is less than

$$\frac{2.2e\pi}{\Omega}\Big(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n}},$$

where $\sigma_{\infty,\min}(I)$ is defined by (2.8). In two dimensions, the resolution limit is multiplied by $(n+1)(n+2)$ when the n point scatterers are assumed to be in a disk of radius $n\pi/\Omega$.

Our paper is organized in the following way. Section 2.1 formulates the minimization problem for recovering point scatterers from multi-illumination data. Sections 2 and 3 present the main results on the spatio-temporal super-resolution in respectively the one- and two-dimensional case and a detailed discussion on their significance. Section 4 introduces the main technique (namely the approximation theory in Vandermonde space) that is used to show the main results of this paper. In Section 5, Theorem 2.1 is proved. Section 6 is devoted to the proof of Theorem 3.1. Finally, the appendix provides some lemmas and inequalities that are used in the paper.

2. Resolution in the one-dimensional case

2.1. Problem setting

Let $\Omega > 0$ be the cut-off frequency. For a smooth function f supported in $[-\Omega, \Omega]$, let

$$||f||_2 = \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} |f(\omega)|^2 d\omega \quad \text{and} \quad ||f||_\infty = \max_{\omega \in [-\Omega, \Omega]} |f(\omega)|.$$

For $\Lambda > 0$, we define the warped-around distance for $x, y \in \mathbb{R}$ by

$$\Big| x - y \Big|_\Lambda = \min_{k \in \mathbb{Z}} \Big| x - y - k\Lambda \Big|. \tag{2.1}$$

Let $\mu = \sum_{j=1}^{n} a_j \delta_{y_j}$ be a discrete measure, where $y_j \in \mathbb{R}, j = 1, \cdots, n$, represent the locations of the point scatterers and $a_j \in \mathbb{C}, j = 1, \cdots, n$, their strengths. We set

$$m_{\min} = \min_{j=1,\cdots,n} |a_j|, \quad d_{\min} = \min_{p \neq j} |y_p - y_j|. \tag{2.2}$$

We assume that the point scatterers are illuminated by some illumination pattern I_t for each time step $t \in \mathbb{N}, 1 \le t \le T$, where T is the total number of frames. Then $I_t \mu$ is given by

$$I_t \mu = \sum_{j=1}^{n} I_t(y_j) a_j \delta_{y_j}, \ t = 1, \cdots, T.$$

The available measurements are the noisy Fourier data of $I_t \mu$ in a bounded interval. More precisely, they are given by

$$\mathbf{Y}_t(\omega) = \mathscr{F}[I_t \mu](\omega) + \mathbf{W}_t(\omega) = \sum_{j=1}^{n} I_t(y_j) a_j e^{i y_j \omega} + \mathbf{W}_t(\omega), \quad 1 \le t \le T, \ \omega \in [-\Omega, \Omega], \tag{2.3}$$

where $\mathscr{F}[I_t \mu]$ denotes the Fourier transform of $I_t \mu$ and $\mathbf{W}_t(\omega)$ is the noise. We assume that $||\mathbf{W}_t||_2 < \sigma$ with σ being the noise level. Recall that π/Ω is the Rayleigh resolution limit.

The inverse problem we are concerned with is to recover the sparsest measure that could generate these diffraction-limited images $\mathbf{Y}_t$'s under certain illuminations. In modern imaging techniques, there are three different cases of interest:

- The illumination patterns are exactly known, such as in SIM and STORM;
- The illumination patterns are unknown but can be approximated, such as in BEAM;
- The illumination patterns are completely unknown.

In this paper, we consider reconstructing the point scatterers as the sparsest solution under the measurement constraint for all these three cases. More specifically, when the illumination patterns are exactly known, we consider the following l_0-minimization problem:

$$\min_{\rho} ||\rho||_0 \quad \text{subject to} \quad ||\mathscr{F}[I_t\rho] - Y_t||_2 < \sigma, \quad 1 \leq t \leq T, \tag{2.4}$$

where $||\rho||_0$ is the number of Dirac masses representing the discrete measure ρ. When the illumination patterns are not exactly known but could be approximated, we consider the l_0-minimization problem:

$$\min_{\rho} ||\rho||_0 \quad \text{subject to} \quad ||\mathscr{F}[\hat{I}_t\rho] - Y_t||_2 < \sigma, \quad 1 \leq t \leq T, \tag{2.5}$$

where $\hat{I}_t$ is an approximation of each I_t so that the feasible set contains some discrete measures with n supports. When the illumination patterns are completely unknown, we consider the following l_0-minimization problem:

$$\min_{\rho} ||\rho||_0 \quad \text{subject to the existence of } \hat{I}_t\text{'s such that } ||\mathscr{F}[\hat{I}_t\rho] - Y_t||_2 < \sigma, \quad 1 \leq t \leq T. \tag{2.6}$$

Our main result in the next section gives an estimation of the resolution of these sparsity recovery problems in the one-dimensional case.

2.2. Main results for the stability of problem (2.4)

We first introduce the illumination matrix as

$$I = \begin{pmatrix} I_1(y_1) & \cdots & I_1(y_n) \\ \vdots & \vdots & \vdots \\ I_T(y_1) & \cdots & I_T(y_n) \end{pmatrix}. \tag{2.7}$$

Then we define, for a $m \times k$ matrix A, $\sigma_{\infty,\min}(A)$ by

$$\sigma_{\infty,\min}(A) = \min_{x\in\mathbb{C}^k, ||x||_\infty \geq 1} ||Ax||_\infty. \tag{2.8}$$

It is easy to see that $\sigma_{\infty,\min}(A)$ characterizes the correlation between the columns of A.

We have the following result on the stability of problems (2.4), (2.5), and (2.6). Its proof is given in Section 5.

Theorem 2.1. *Suppose that $\mu = \sum_{j=1}^n a_j\delta_{y_j}$ and the following separation condition holds:*

$$d_{\min} := \min_{p\neq j} \left|y_p - y_j\right|_{\frac{n\pi}{\Omega}} \geq \frac{2.2e\pi}{\Omega}\Big(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n}}, \tag{2.9}$$

with $\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}} \leq 1$. Here, $m_{\min}$ is defined in (2.2) and $\frac{\sigma}{m_{\min}}$ is the noise-to-signal ratio. Then any solution to (2.4), (2.5), or (2.6) contains exactly n point scatterers. Moreover, for $\rho = \sum_{j=1}^n \hat{a}_j\delta_{\hat{y}_j}$ being the corresponding solution, after reordering the $\hat{y}_j$'s, we have

$$\left|\hat{y}_j - y_j\right|_{\frac{n\pi}{\Omega}} < \frac{d_{\min}}{2}, \tag{2.10}$$

and

$$\left|\hat{y}_j - y_j\right|_{\frac{n\pi}{\Omega}} < \frac{C(n)}{\Omega} SRF^{n-1} \frac{1}{\sigma_{\infty,\min}(I)} \frac{\sigma}{m_{\min}}, \quad 1 \leq j \leq n, \tag{2.11}$$

where $C(n) = 2\sqrt{2\pi} n e^n$ *and* $SRF = \frac{\pi}{\Omega d_{\min}}$ *is the super-resolution factor.*

Remark 2.1. *In this paper, for simplicity, we assume that the measurements are for all* $\omega \in [-\Omega, \Omega]$*. Nevertheless, our results can be easily extended to the discrete sampling case, for example, when the measurements are taken at* M *evenly spaced points* $\omega_l \in [-\Omega, \Omega]$ *with* $M \geq n$*. The minimum number of sampling points at each single frame is only* n*, which shows that the sparsity recovery can reduce significantly the number of measurements. Moreover, if we consider that the point scatterers (as well as the solution of (2.4)) are supported in an interval of length of several Rayleigh resolution limits, then the warped-around distance in Theorem 2.1 can be replaced by the Euclidean distance (with only a slight modification of the results). Under this scenario, by utilizing the projection trick introduced in [23], our results can also be extended to multi-dimensional spaces.*

Remark 2.2. *For the case when* $n = 2$*, the minimal separation distance in Theorem 2.1*

$$\frac{2.2e\pi}{\Omega}\Big(\frac{1}{\sigma_{\infty,\min}(I)} \frac{\sigma}{m_{\min}}\Big)^{\frac{1}{2}},$$

applies for any k*-dimensional spaces. It means that for multi-illumination imaging in general* k*-dimensional space, the two-point resolution [8, 11, 36–38] of sparsity recoveries like (2.4), (2.5), or (2.6) is less than* $\frac{2.2e\pi}{\Omega}\Big(\frac{1}{\sigma_{\infty,\min}(I)} \frac{\sigma}{m_{\min}}\Big)^{\frac{1}{2}}$.

Remark 2.3. *Note that the stability result in Theorem 2.1 holds for any algorithm that can recover the sparsest solution (solution with* n *point scatterers). Thus it also helps to understand the performance of other sparsity-promoting algorithms, such as the* l_1*-minimization that is frequently used in the sparsity-based super-resolution. Also, our results can be generalized to the multi-clump case, where the resolution is related to the sparsity of the point scatterers in each clump rather than the total number of point scatterers. This can explain the fact that we can achieve super-resolution imaging even in the case where we have tens or hundreds of point scatterers.*

Remark 2.4. *Note also that our results can be extended to other kinds of imaging systems with different point spread functions. For example, let the point spread function be* f*. In the presence of an additive noise* $w(t)$*, the measurement in the time-domain is*

$$f \circledast \mu(t) + w(t) = \sum_{j=1}^{n} a_j f(t - y_j) + w(t).$$

By taking the Fourier transform, we obtain

$$\mathscr{F} y(\omega) = \mathscr{F} f(\omega) \mathscr{F} \mu(\omega) + \mathscr{F} w(\omega) = \mathscr{F} f(\omega)\Big(\sum_{j=1}^{n} a_j e^{i y_j \omega}\Big) + \mathscr{F} w(\omega).$$

Suppose that $|\mathscr{F} f(\omega)| > 0$ *at the sampling points. Then our results can be easily extended to the case when the point spread function is* f*.*

Theorem 2.1 demonstrates that when the point scatterers are separated by the distance $d_{\min}$ in (2.9), we can stably recover the scatterer locations. Under the minimal separation condition, each of the recovered locations is in a neighborhood of the ground truth and the deviation of them from the ground truth is also estimated. Thus the resolution of our sparsity-promoting algorithms for the multi-illumination data is less than

$$\frac{2.2e\pi}{\Omega}\Big(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n}}.$$

Based on this formula of the resolution limit, we demonstrate that the incoherence (encoded in $\sigma_{\infty,\min}(I)$) between the illumination patterns (or columns in illumination matrix (2.7)) is crucial to the sparsity-based spatio-temporal super-resolution. More precisely, applying any sparsity-promoting algorithm for images from illumination patterns with high degree of incoherence can achieve desired super-resolution, even when only a small number of frames are provided, which yields high spatio-temporal resolution. This is the most important contribution of our paper.

We remark that our result can even serve as a way to estimate explicitly the resolution for the multi-illumination imaging when we could know or estimate the incoherence of the illumination patterns and the signal-to-noise ratio. We present a simple example as follows that calculates explicitly the resolution limit of our sparsity recovery problem by the estimation (2.9). We leave the other detailed discussions on Theorem 2.1 to the following three subsections.

Example 2.1. *We consider two point scatterers that are illuminated by two illumination patterns. Suppose for instance that the illumination matrix is given by*

$$I=\begin{pmatrix}1 & 0.7\\ 0.7 & 1\end{pmatrix}.$$

Suppose also that the noise level is $\sigma=10^{-3}$ and the noise-to-signal ratio is $\frac{\sigma}{m_{\min}}=10^{-3}$. By Lemma C.3, $\sigma_{\infty,\min}(I)$, defined in (2.8), is equal to 0.3. *Hence, by Theorem 2.1, the resolution limit $d_{\min}$ in solving problem (2.4) (2.5), or (2.6) is smaller than*

$$\frac{2.2e\pi}{\Omega}\Big(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n}}\approx 0.34\frac{\pi}{\Omega},$$

where $\frac{\pi}{\Omega}$, as said before, is the classical Rayleigh resolution limit. This shows that even with only two illuminations with mild degree of incoherence, there is a threefold resolution improvement.

2.3. Discussion of $\sigma_{\infty,\min}(I)$ and the effect of multiple illumination

2.3.1. Adding the same illumination pattern will not enhance the resolution

Let

$$I=\begin{pmatrix} I_1(y_1) & \cdots & I_1(y_n)\\ \vdots & \vdots & \vdots\\ I_{T-1}(y_1) & \cdots & I_{T-1}(y_n)\\ I_T(y_1) & \cdots & I_T(y_n)\end{pmatrix},\quad \hat I=\begin{pmatrix} I_1(y_1) & \cdots & I_1(y_n)\\ \vdots & \vdots & \vdots\\ I_T(y_1) & \cdots & I_T(y_n)\\ I_{T+1}(y_1) & \cdots & I_{T+1}(y_n)\end{pmatrix}$$

with $I_{T+1} = I_T$. By the definition of $\sigma_{\infty,\min}$, it is clear that $\sigma_{\infty,\min}(\hat{I}) = \sigma_{\infty,\min}(I)$. Thus, adding the same illumination pattern can not increase the resolution in Theorem 2.1. This is consistent with our observation that multiple illuminations with different patterns are key for spatio-temporal super-resolution.

2.3.2. The incoherence between the illumination patterns is crucial

The value of $\sigma_{\infty,\min}(I)$ is related to the correlation between the columns of the illumination matrix I. In particular, we have the following rough estimation of $\sigma_{\infty,\min}(I)$:

$$\sigma_{\infty,\min}(I) \geq \frac{\sigma_{\min}(I)}{\sqrt{T}}, \tag{2.12}$$

where $\sigma_{\min}(I)$ is the minimum singular value of I. This clearly illustrates that the correlation between the columns of I is crucial to $\sigma_{\infty,\min}(I)$. The correlation between columns of I is related to the incoherence of the illumination patterns. Thus, we should employ illumination patterns with high degree of incoherence in order to increase $\sigma_{\infty,\min}(I)$, and consequently, obtain a significant resolution enhancement.

2.4. Comparison with the single illumination case

In this subsection, we compare the resolution in the single illumination case (i.e., the single measurement case) with that in the multiple illumination case, whereby we illustrate the effect of multiple illuminations in enhancing the resolution.

In [24], the authors estimate the so-called computational resolution limit for the line spectral estimation problem of the single measurement case. The line spectral estimation problem is to estimate the locations of some line spectra from the Fourier data (in a bounded domain) of one of their linear combination. So the line spectral problem is equivalent to the super-resolution problem considered here. The results in [24] show that, for the single measurement case, when the point scatterers are separated by

$$\tau = \frac{c_0}{\Omega}\left(\frac{\sigma}{m_{\min}}\right)^{\frac{1}{2n-1}},$$

for some positive constant c_0, there exists a discrete measure $\mu = \sum_{j=1}^n a_j \delta_{y_j}$ with n point scatterers located at $\{-\tau, -2\tau, -n\tau\}$ and another discrete measure $\hat{\mu} = \sum_{j=1}^n \hat{a}_j \delta_{\hat{y}_j}$ with n point scatterers located at $\{0, \tau, \cdots, (n-1)\tau\}$ such that

$$||\mathscr{F}[\hat{\mu}] - \mathscr{F}[\mu]||_\infty < \sigma,$$

and either $\min_{1\leq j\leq n} |a_j| = m_{\min}$ or $\min_{1\leq j\leq n} |\hat{a}_j| = m_{\min}$.

By the definition of $||\cdot||_\infty$ and $||\cdot||_2$, we also have

$$||\mathscr{F}[\hat{\mu}] - \mathscr{F}[\mu]||_2 < \sigma.$$

This result demonstrates that when the point scatterers are separated by $\frac{c_0}{\Omega}(\frac{\sigma}{m_{\min}})^{\frac{1}{2n-1}}$, the solution of the l_0-minimization problem in the single measurement case

$$\min_{\rho} ||\rho||_0 \quad \text{subject to} \quad ||\mathscr{F}[\rho] - Y||_2 < \sigma, \tag{2.13}$$

is not stable. In particular, the recovered point scatterers by (2.13) may be located in an interval completely disjoint from that of the ground truth.

Therefore, for the single measurement case, when the scatterers are separated by $O(\frac{(\frac{\sigma}{m_{\min}})^{\frac{1}{2n-1}}}{\Omega})$, the l_0-minimization may be unstable. However, for the multiple illumination case, when the point scatterers are separated by $O(\frac{(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}})^{\frac{1}{n}}}{\Omega})$, the l_0-minimization (2.4) is still stable.

Suppose we have illumination patterns such that $\frac{1}{\sigma_{\infty,\min}(I)}$ is of constant order, the resolution now is of order $O(\frac{(\frac{\sigma}{m_{\min}})^{\frac{1}{n}}}{\Omega})$. Compared with the resolution in the single measurement case, say of order $O(\frac{(\frac{\sigma}{m_{\min}})^{\frac{1}{2n-1}}}{\Omega})$, this clearly shows a significant enhancement and illustrates the effect of multiple illuminations in improving the resolution.

2.5. Lower bound for the resolution of multi-illumination imaging

By Theorem 2.1, when we have desired illumination patterns with high degree of incoherence so that $\sigma_{\infty,\min}(I)$ is of order one, the resolution of the sparsity recovery is expected to be less than $\frac{c_0}{\Omega}(\frac{\sigma}{m_{\min}})^{\frac{1}{n}}$ for some positive constant c_0. We next demonstrate that this resolution order is the best we can obtain if the illumination patterns are unknown. More precisely, we have the following proposition whose proof is given in Appendix B.

Proposition 2.1. *Given $n\geq 2$, $\sigma, m_{\min}$ with $\frac{\sigma}{m_{\min}}\leq 1$, and unknown illumination pattern I_t with $|I_t(y)|\leq 1, y\in\mathbb{R}, 1\leq t\leq T$, let τ be given by*

$$\tau=\frac{0.043}{\Omega}\Big(\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n}}. \tag{2.14}$$

Then there exist $\mu=\sum_{j=1}^{n}a_j\delta_{y_j}$ with n supports at $\{-\tau,-2\tau,\ldots,-n\tau\}$ and $|a_j|=m_{\min}, 1\leq j\leq n$, and $\rho=\sum_{j=1}^{n-1}\hat a_j\delta_{\hat y_j}$ with n supports at $\{0,\tau,\cdots,(n-1)\tau\}$, such that

there exist $\hat I_t$'s so that $||\mathscr{F}[\hat I_t\rho]-\mathscr{F}[I_t\mu]||_2<\sigma,\ t=1,\cdots,T$.

3. Resolution in the two-dimensional case

3.1. Problem setting

Let $\Omega>0$ be the cut-off frequency. For a smooth function $f:\mathbb{R}^2\to\mathbb{R}$ supported on $||\boldsymbol{\omega}||_2\leq\Omega$, let

$$||f||_\infty=\max_{||\boldsymbol{\omega}||_2\leq\Omega}|f(\boldsymbol{\omega})|.$$

Let $\mu=\sum_{j=1}^{n}a_j\delta_{\mathbf{y}_j}$ be a discrete measure, where $\mathbf{y}_j\in\mathbb{R}^2, j=1,\cdots,n$, represent the locations of the point scatterers and $a_j\in\mathbb{C}, j=1,\cdots,n$, their strengths. We set

$$m_{\min}=\min_{j=1,\cdots,n}|a_j|,\quad d_{\min}=\min_{p\neq j}||\mathbf{y}_p-\mathbf{y}_j||_2. \tag{3.1}$$

Again, we assume that the point scatterers are illuminated by some illumination pattern I_t for each time step $t \in \mathbb{N}, 1 \le t \le T$, where T is the total number of illumination patterns. Then $I_t\mu$ is

$$I_t\mu = \sum_{j=1}^{n} I_t(\mathbf{y}_j) a_j \delta_{\mathbf{y}_j}, \quad t = 1, \cdots, T.$$

In the time-domain, the measurements are

$$A_t\mu := h \circledast (I_t\mu), \quad t = 1, 2, \cdots, T,$$

where h is a blurring kernel in $\mathbb{R}^2$. Thus, in the Fourier-domain, the available measurements are given by

$$\mathbf{Y}_t(\boldsymbol{\omega}) = \mathscr{F}[I_t\mu](\boldsymbol{\omega}) + \mathbf{W}_t(\boldsymbol{\omega}) = \sum_{j=1}^{n} I_t(\mathbf{y}_j) a_j e^{i\mathbf{y}_j\cdot\boldsymbol{\omega}} + \mathbf{W}_t(\boldsymbol{\omega}),\ 1 \le t \le T, ||\boldsymbol{\omega}||_2 \le \Omega, \tag{3.2}$$

where $\mathscr{F}[I_t\mu]$ denotes the Fourier transform of $I_t\mu$ and $\mathbf{W}_t(\boldsymbol{\omega})$ is the noise. We assume that $||\mathbf{W}_t||_\infty < \sigma$ with σ being the noise level.

We consider reconstructing the point scatterers as the sparsest solution (solution to the l_0-minimization problem) under the measurement constraints for the three cases of illumination patterns that are discussed in Section 2.1. With a slight abuse of notation, we also denote by $\mathscr{F}[\rho]$ the function $\mathscr{F}[\rho](\boldsymbol{\omega}), ||\boldsymbol{\omega}||_2 \le \Omega$. In this section, we suppose that the point scatterers are located in a disk $\mathcal{O}$ with radius of several Rayleigh resolution limits. Then we consider the following optimization problems. When the illumination patterns are exactly known, we consider the following l_0-minimization problem:

$$\min_{\rho \text{ supported in } \mathcal{O}} ||\rho||_0 \quad \text{subject to } ||\mathscr{F}[I_t\rho] - \mathbf{Y}_t||_\infty < \sigma, \quad 1 \le t \le T, \tag{3.3}$$

where $||\rho||_0$ is the number of Dirac masses representing the discrete measure ρ. When the illumination patterns are not exactly known but could be approximated, we consider the l_0-minimization problem

$$\min_{\rho \text{ supported in } \mathcal{O}} ||\rho||_0 \quad \text{subject to } ||\mathscr{F}[\hat{I}_t\rho] - \mathbf{Y}_t||_\infty < \sigma, \quad 1 \le t \le T, \tag{3.4}$$

where $\hat{I}_t$ is an approximation of each I_t so that the feasible set contains some measures with n supports. When the illumination patterns are completely unknown, we consider the following l_0-minimization problem:

$$\min_{\rho \text{ supported in } \mathcal{O}} ||\rho||_0 \quad \text{subject to the existence of } \hat{I}_t\text{'s such that } ||\mathscr{F}[\hat{I}_t\rho] - \mathbf{Y}_t||_\infty < \sigma,\ 1 \le t \le T. \tag{3.5}$$

Our main result in the following subsection gives an estimation of the resolution of these two-dimensional sparsity recovery problems.

3.2. MAIN RESULTS FOR THE STABILITY OF SPARSITY RECOVERIES IN TWO DIMENSIONS

The illumination matrix in the two-dimensional case is

$$I=\begin{pmatrix} I_1(\mathbf{y}_1) & \cdots & I_1(\mathbf{y}_n)\\ \vdots & \vdots & \vdots \\ I_T(\mathbf{y}_1) & \cdots & I_T(\mathbf{y}_n)\end{pmatrix}. \tag{3.6}$$

We have the following theorem on the stability of problems (3.3), (3.4), and (3.5). Its proof is given in Section 6.

Theorem 3.1. *Let $n\geq 2$ and let the disk $\mathcal{O}$ be of radius $\frac{c_0 n\pi}{\Omega}$ with $c_0\geq 1$. Let $\mathbf{Y}_t$'s be the measurements that are generated by an n-sparse measure $\mu=\sum_{j=1}^{n} a_j\delta_{\mathbf{y}_j}, \mathbf{y}_j\in\mathcal{O}$ in the two-dimensional space. Assume that*

$$d_{\min}:=\min_{p\neq j}\left|\left|\mathbf{y}_p-\mathbf{y}_j\right|\right|_2\geq \frac{2.2c_0 e\pi(n+2)(n+1)}{\Omega}\Big(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n}}. \tag{3.7}$$

Here, I is the matrix in (3.6), $m_{\min}$ is defined in (3.1) and $\frac{\sigma}{m_{\min}}$ is the noise-to-signal ratio. Then any solution to (3.3), (3.4), and (3.5) contains exactly n point scatterers. Moreover, for $\rho=\sum_{j=1}^{n}\hat{a}_j\delta_{\hat{\mathbf{y}}_j}$ being the corresponding solution, after reordering the $\hat{\mathbf{y}}_j$'s, we have

$$\left|\left|\hat{\mathbf{y}}_j-\mathbf{y}_j\right|\right|_2<\frac{d_{\min}}{2}, \tag{3.8}$$

and

$$\left|\left|\hat{\mathbf{y}}_j-\mathbf{y}_j\right|\right|_2<\frac{C(n)}{\Omega}SRF^{n-1}\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}},\quad 1\leq j\leq n, \tag{3.9}$$

where $C(n)=(n+1)^n(n+2)^n\sqrt{2\pi}nc_0^{n-1}e^n$ and $SRF=\frac{\pi}{\Omega d_{\min}}$ is the super-resolution factor.

Theorem 3.1 is the two-dimensional analogue of Theorem 2.1. It reveals the dependence of the resolution of two-dimensional sparsity recoveries on the cut-off frequency of the imaging system, the signal-to-noise ratio, the sparsity of point scatters, and the incoherence of illumination patterns. It highlights the importance of multiple illumination patterns with high degree of incoherence in achieving two-dimensional spatio-temporal super-resolution.

4. NON-LINEAR APPROXIMATION THEORY IN VANDERMONDE SPACE

In this section, we present the main technique that is used in the proofs of the main results of the paper, namely the approximation theory in Vandermonde space. This theory was first introduced in [24,25]. Instead of considering the non-linear approximation problem there, we consider a different approximation problem, which is relevant to the stability analysis of (2.4). More specifically, for $s\in\mathbb{N}, s\geq 1$, and $z\in\mathbb{C}$, we define the complex Vandermonde-vector

$$\phi_s(z)=(1,z,\cdots,z^s)^{\top}. \tag{4.1}$$

Throughout this paper, for a complex matrix A, we denote $A^{\top}$ its transpose and A^* its conjugate transpose.

We consider the following non-linear problem:

$$\min_{\hat{\theta}_j\in\mathbb{R}, j=1,\cdots,k} \max_{t=1,\cdots,T} \min_{\hat{a}_{j,t}\in\mathbb{C}, j=1,\cdots,k} \Big|\Big| \sum_{j=1}^{k} \hat{a}_{j,t}\phi_s(e^{i\hat{\theta}_j}) - v_t \Big|\Big|_2, \tag{4.2}$$

where $v_t = \sum_{j=1}^{k+1} a_{j,t}\phi_s(e^{i\theta_j})$ is given with θ_j's being real numbers. We shall derive a lower bound for the optimal value of the minimization problem for the case when $s = k$. The main results are presented in Section 4.2.

4.1. Notation and Preliminaries

We first introduce some notation and preliminaries. We denote for $k \in \mathbb{N}, k \geq 1$,

$$\zeta(k) = \begin{cases} (\frac{k-1}{2}!)^2, & k \text{ is odd}, \\ (\frac{k}{2})!(\frac{k-2}{2})!, & k \text{ is even}, \end{cases} \quad \xi(k) = \begin{cases} 1/2, & k = 1, \\ \frac{(\frac{k-1}{2})!(\frac{k-3}{2})!}{4}, & k \text{ is odd}, \ k \geq 3, \\ \frac{(\frac{k-2}{2}!)^2}{4}, & k \text{ is even}. \end{cases} \tag{4.3}$$

We also define for $p, q \in \mathbb{N}, p, q \geq 1$, and $z_1, \cdots, z_p, \hat{z}_1, \cdots, \hat{z}_q \in \mathbb{C}$, the following vector in $\mathbb{R}^p$:

$$\eta_{p,q}(z_1,\cdots,z_p,\hat{z}_1,\cdots,\hat{z}_q) = \begin{pmatrix} |(z_1-\hat{z}_1)|\cdots|(z_1-\hat{z}_q)| \\ |(z_2-\hat{z}_1)|\cdots|(z_2-\hat{z}_q)| \\ \vdots \\ |(z_p-\hat{z}_1)|\cdots|(z_p-\hat{z}_q)| \end{pmatrix}. \tag{4.4}$$

We present two auxiliary lemmas that are helpful for deriving our main results. These lemmas are slightly different from the ones in [24, Section III]. Thus, we employ different techniques for proving them. Their proofs are presented in Appendix A.

Lemma 4.1. *For $\theta_j \in \mathbb{R}, j = 1,\cdots,k+1$, assume that $\min_{p\neq j} |\theta_p - \theta_j|_{2\pi} = \theta_{\min}$. Then, for any $\hat{\theta}_1,\cdots,\hat{\theta}_k \in \mathbb{R}$, we have the following estimate:*

$$||\eta_{k+1,k}(e^{i\theta_1},\cdots,e^{i\theta_{k+1}},e^{i\hat{\theta}_1},\cdots,e^{i\hat{\theta}_k})||_\infty \geq \xi(k)(\frac{2\theta_{\min}}{\pi})^k.$$

Lemma 4.2. *Let $\epsilon > 0$. For $\theta_j, \hat{\theta}_j \in \mathbb{R}, j = 1,\cdots,k$, assume that*

$$||\eta_{k,k}(e^{i\theta_1},\cdots,e^{i\theta_k},e^{i\hat{\theta}_1},\cdots,e^{i\hat{\theta}_k})||_\infty < (\frac{2}{\pi})^k\epsilon, \tag{4.5}$$

where $\eta_{k,k}$ is defined as in (4.4), and that

$$\theta_{\min} = \min_{q\neq j} |\theta_q - \theta_j|_{2\pi} \geq \Big(\frac{4\epsilon}{\lambda(k)}\Big)^{\frac{1}{k}}, \tag{4.6}$$

where

$$\lambda(k) = \begin{cases} 1, & k = 2, \\ \xi(k-2), & k \geq 3. \end{cases} \tag{4.7}$$

Then, after reordering the $\hat{\theta}_j$'s, we have

$$|\hat{\theta}_j-\theta_j|_{2\pi}<\frac{\theta_{\min}}{2},\quad j=1,\cdots,k, \tag{4.8}$$

and moreover,

$$|\hat{\theta}_j-\theta_j|_{2\pi}<\frac{2^{k-1}\epsilon}{(k-2)!(\theta_{\min})^{k-1}},\quad j=1,\cdots,k. \tag{4.9}$$

4.2. Main results on the approximation theory in Vandermonde space

Before presenting a lower bound for problem (4.2), we introduce a basic approximation result in Vandermonde space. This result was first derived in [24].

Theorem 4.1. *Let $k\geq 1$. For fixed $\hat{\theta}_1,\cdots,\hat{\theta}_k\in\mathbb{R}$, denote $\hat{A}=\big(\phi_k(e^{i\hat{\theta}_1}),\cdots,\phi_k(e^{i\hat{\theta}_k})\big)$, where the $\phi_k(e^{i\hat{\theta}_j})$'s are defined as in (4.1). Let V be the k-dimensional complex space spanned by the column vectors of $\hat{A}$ and let $V^{\perp}$ be the one-dimensional orthogonal complement of V in $\mathbb{C}^{k+1}$. Denote by $P_{V^{\perp}}$ the orthogonal projection onto $V^{\perp}$ in $\mathbb{C}^{k+1}$. Then, we have*

$$\min_{\hat{a}\in\mathbb{C}^k}||\hat{A}\hat{a}-\phi_k(e^{i\theta})||_2=||P_{V^{\perp}}\big(\phi_k(e^{i\theta})\big)||_2=|v^*\phi_k(e^{i\theta})|\geq\frac{1}{2^k}|\Pi_{j=1}^k(e^{i\theta}-e^{i\hat{\theta}_j})|,$$

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

We then have the following result for non-linear approximation (4.2) in Vandermonde space.

Theorem 4.2. *Let $k\geq 1$ and $\theta_j\in\mathbb{R}, 1\leq j\leq k+1$, be $k+1$ distinct points with $\theta_{\min}=\min_{p\neq j}|\theta_p-\theta_j|_{2\pi}>0$. For $q\leq k$, let $\hat{\alpha}_t(q)=(\hat{a}_{1,t},\cdots,\hat{a}_{q,t})^{\top}$, $\alpha_t=(a_{1,t},\cdots,a_{k+1,t})^{\top}$ and*

$$\hat{A}(q)=\big(\phi_k(e^{i\hat{\theta}_1}),\cdots,\phi_k(e^{i\hat{\theta}_q})\big),\quad A=\big(\phi_k(e^{i\theta_1}),\cdots,\phi_k(e^{i\theta_{k+1}})\big),$$

where $\phi_k(z)$ is defined as in (4.1). Then, for any $\hat{\theta}_1,\cdots,\hat{\theta}_q\in\mathbb{R}$,

$$\max_{t=1,\cdots,T}\min_{\hat{\alpha}_t(q)\in\mathbb{C}^q}||\hat{A}(q)\hat{\alpha}_t(q)-A\alpha_t||_2\geq\frac{\sigma_{\infty,\min}(B)\xi(k)(\theta_{\min})^k}{\pi^k},$$

where

$$B=\begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{k+1,1}\\ \vdots & \vdots & \vdots & \vdots\\ a_{1,T} & a_{2,T} & \cdots & a_{k+1,T}\end{pmatrix}. \tag{4.10}$$

Proof. **Step 1.** Note that, for any $\hat{\theta}_1,\cdots,\hat{\theta}_q,\cdots,\hat{\theta}_k\in\mathbb{R}$, if $q<k$, then

$$\min_{\hat{a}_t(q)\in\mathbb{C}^q}||\hat{A}(q)\hat{\alpha}_t(q)-A\alpha_t||_2\geq\min_{\hat{\alpha}_t(k)\in\mathbb{C}^k}||\hat{A}(k)\hat{\alpha}_t(k)-A\alpha_t||_2,\ 1\leq t\leq T.$$

So, we only need to consider the case when $q=k$. We shall verify that, for any k distinct points $\hat{\theta}_1,\cdots,\hat{\theta}_k\in\mathbb{R}$, we have

$$\max_{t=1,\cdots,T}\min_{\hat{\alpha}_t\in\mathbb{C}^k}||\hat{A}(k)\hat{\alpha}_t(k)-A\alpha_t||_2\geq\frac{\sigma_{\infty,\min}(B)\xi(k)(\theta_{\min})^k}{\pi^k}. \tag{4.11}$$

Let us then fix $(\hat{\theta}_1, \cdots, \hat{\theta}_k)$ in our subsequent arguments.

Step 2. Let V be the complex space spanned by the column vectors of $\hat{A}(k)$ and let $V^{\perp}$ be the orthogonal complement of V in $\mathbb{C}^{k+1}$. It is clear that $V^{\perp}$ is a one-dimensional complex space. We let v be a unit vector in $V^{\perp}$ and denote by $P_{V^{\perp}}$ the orthogonal projection onto $V^{\perp}$ in $\mathbb{C}^{k+1}$. Note that $\|P_{V^{\perp}} u\|_2 = |v^* u|$ for $u \in \mathbb{C}^{k+1}$, where v^* is the conjugate transpose of v. We have

$$\min_{\hat{\alpha}_t \in \mathbb{C}^k} ||\hat{A}(k)\hat{\alpha}_t - A\alpha_t||_2 = ||P_{V^{\perp}}(A\alpha_t)||_2 = |v^* A\alpha_t| = |\sum_{j=1}^{k+1} a_{j,t} v^* \phi_k(e^{i\theta_j})| = |\beta_t|, \tag{4.12}$$

where $\beta_t = \sum_{j=1}^{k+1} a_{j,t} v^* \phi_k(e^{i\theta_j}),\ t = 1, \cdots, T.$ Denote by $\beta = (\beta_1, \beta_2, \cdots, \beta_T)^{\top}$. Thus, we only need to estimate the lower bound of $||\beta||_{\infty}$. By (4.12), we have $\beta = B\hat{\eta}$, where B is given by (4.10) and $\hat{\eta} = (v^*\phi_k(e^{i\theta_1}), v^*\phi_k(e^{i\theta_2}), \cdots, v^*\phi_k(e^{i\theta_{k+1}}))^{\top}$. By the definition of $\sigma_{\infty,\min}(B)$, we have

$$||\beta||_{\infty} \geq \sigma_{\infty,\min}(B)||\hat{\eta}||_{\infty}.$$

On the other hand, by Theorem 4.1, we obtain that

$$||\hat{\eta}||_{\infty} \geq \frac{1}{2^k}||\eta_{k+1,k}(e^{i\theta_1}, \cdots, e^{i\theta_{k+1}}, e^{i\hat{\theta}_1}, \cdots, e^{i\hat{\theta}_k})||_{\infty},$$

where $\eta_{k+1,k}$ is defined by (4.4). Combining this with Lemma 4.1, we get

$$||\hat{\eta}||_{\infty} \geq \frac{1}{2^k}\xi(k)(\frac{2\theta_{\min}}{\pi})^k.$$

It then follows that

$$||\beta||_{\infty} \geq \frac{\sigma_{\infty,\min}(B)\xi(k)(\theta_{\min})^k}{\pi^k},$$

which proves (4.11) and hence the theorem. □

Theorem 4.3. *Let $k \geq 2$ and $\theta_j \in \mathbb{R}$, $j = 1, \cdots, k$, be k different points with*

$$\theta_{\min} = \min_{p \neq j} |\theta_p - \theta_j|_{2\pi} > 0.$$

Assume that there are k distinct points $\hat{\theta}_1, \cdots, \hat{\theta}_k \in \mathbb{R}$ satisfying

$$\max_{t=1,\cdots,T} ||\hat{A}\hat{\alpha}_t - A\alpha_t||_2 < \sigma,$$

where $\hat{\alpha}_t = (\hat{a}_{1,t}, \cdots, \hat{a}_{k,t})^{\top}$, $\alpha_t = (a_{1,t}, \cdots, a_{k,t})^{\top}$ and

$$\hat{A} = \big(\phi_k(e^{i\hat{\theta}_1}), \cdots, \phi_k(e^{i\hat{\theta}_k})\big), \quad A = \big(\phi_k(e^{i\theta_1}), \cdots, \phi_k(e^{i\theta_k})\big).$$

Then

$$||\eta_{k,k}(e^{i\theta_1},\cdots,e^{i\theta_k},e^{i\hat\theta_1},\cdots,e^{i\hat\theta_k})||_{\infty} < \frac{2^k}{\sigma_{\infty,\min}(B)}\sigma,$$

where

$$B = \begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{k,1} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1,T} & a_{2,T} & \cdots & a_{k,T} \end{pmatrix}. \tag{4.13}$$

Proof. Let V be the complex space spanned by the column vectors of $\hat A$ and let $V^{\perp}$ be the orthogonal complement of V in $\mathbb{C}^{k+1}$. Let v be a unit vector in $V^{\perp}$ and denote by $P_{V^{\perp}}$ the orthogonal projection onto $V^{\perp}$ in $\mathbb{C}^{k+1}$. Similarly to Step 2 in the proof of Theorem 4.2, we obtain that

$$\min_{\hat\alpha_t\in\mathbb{C}^k}||\hat A\hat\alpha_t - A\alpha_t||_2 = ||P_{V^{\perp}}(A\alpha_t)||_2 = |v^*A\alpha_t| = |\sum_{j=1}^k a_{j,t}v^*\phi_k(e^{i\theta_j})| = |\beta_t|, \tag{4.14}$$

where $\beta_t = \sum_{j=1}^k a_{j,t}v^*\phi_k(e^{i\theta_j}),\ t=1,\cdots,T$. Denote by $\beta = (\beta_1,\beta_2,\cdots,\beta_T)^{\top}$, we have $\beta = B\hat\eta$, where B is given by (4.13) and $\hat\eta = (v^*\phi_k(e^{i\theta_1}), v^*\phi_k(e^{i\theta_2}),\cdots,v^*\phi_k(e^{i\theta_k}))^{\top}$. By the definition of $\sigma_{\infty,\min}(B)$, we arrive at

$$||\beta||_{\infty} \geq \sigma_{\infty,\min}(B)||\hat\eta||_{\infty}.$$

On the other hand, by Theorem 4.1, we get

$$||\hat\eta||_{\infty} \geq \frac{1}{2^k}||\eta_{k,k}(e^{i\theta_1},\cdots,e^{i\theta_k},e^{i\hat\theta_1},\cdots,e^{i\hat\theta_k})||_{\infty},$$

and hence the theorem is proved. □

5. Proof of Theorem 2.1

The proof of Theorem 2.1 is divided into four steps.

Step 1. We only prove the theorem for problem (2.4) and the other two cases can be proved in the same manner. We first prove that the solution to (2.4) is a discrete measure corresponding to at least n point scatterers. For $\rho = \sum_{j=1}^k \hat a_j\delta_{\hat y_j}$ and $\mu = \sum_{j=1}^n a_j\delta_{y_j}$, we set $\hat\mu_t = I_t\rho = \sum_{j=1}^k \hat a_{j,t}\delta_{\hat y_j}$ and $\mu_t = \sum_{j=1}^n I_t(y_j)a_j\delta_{y_j}$. We shall prove that if $k<n$, then for any $\hat y_j\in\mathbb{R}, \hat a_{j,t}\in\mathbb{C}, j=1,\cdots,k, t=1,\cdots,T$,

$$\max_{t=1,\cdots,T}||\mathscr{F}[\hat\mu_t]-\mathscr{F}[\mu_t]||_2 > 2\sigma. \tag{5.1}$$

For ease of presentation, we fix $\hat y_j, \hat a_{j,t}$'s in the subsequent arguments. In view of $||\mathbf{W}_t||_2<\sigma$, from (5.1) we further have

$$\max_{t=1,\cdots,T}||\mathscr{F}[\hat\mu_t]-\mathbf{Y}_t||_2 > \sigma, \tag{5.2}$$

whereby any solution corresponding to only $k<n$ point scatterers cannot be a solution to (2.4). We now begin our proof. Let $h=\frac{2\Omega}{n}$ and $g(\omega)=\mathscr{F}[\hat\mu_t](\omega)-\mathscr{F}[\mu_t](\omega)$ for each t, we have

$$
\begin{aligned}
&\frac{1}{2\Omega}\int_{-\Omega}^{\Omega}|g(\omega)|^2d\omega=\frac{1}{2\Omega}\sum_{j=1}^{n}\int_{-\Omega+(j-1)h}^{-\Omega+jh}|g(\omega)|^2d\omega\\
=&\frac{1}{2\Omega}\int_0^h\sum_{j=1}^{n}|g(\omega+(j-1)h-\Omega)|^2d\omega.
\end{aligned}
$$

With this decomposition in hand, in order to prove (5.1) we shall prove in the following steps that for all ω in $[0,h]$,

$$
\max_{t=1,\cdots,T}\frac{1}{n}\sum_{j=1}^{n}|\mathscr{F}[\hat\mu_t](\omega+(j-1)h-\Omega)-\mathscr{F}[\mu_t](\omega+(j-1)h-\Omega)|^2>4\sigma^2. \tag{5.3}
$$

Without loss of generality, we only show (5.3) for $\omega=0$. For $k<n$, we consider

$$
\big(\mathscr{F}[\hat\mu_t](\omega_1),\mathscr{F}[\hat\mu_t](\omega_2),\cdots,\mathscr{F}[\hat\mu_t](\omega_{n-1})\big)^\top-\big(\mathscr{F}[\mu_t](\omega_1),\mathscr{F}[\mu_t](\omega_2),\cdots,\mathscr{F}[\mu_t](\omega_{n-1})\big)^\top, \tag{5.4}
$$

where $\omega_j=(j-1)h-\Omega,\ j=1,\cdots,n-1$. We write (5.4) as

$$
\hat\Phi\hat\alpha_t-\Phi\alpha_t,
$$

where $\hat\alpha_t=(\hat a_{1,t},\cdots,\hat a_{k,t})^\top$, $\alpha_t=(I_t(y_1)a_1,\cdots,I_t(y_n)a_n)^\top$ and

$$
\hat\Phi=\begin{pmatrix} e^{i\hat y_1\omega_1} & \cdots & e^{i\hat y_k\omega_1}\\ e^{i\hat y_1\omega_2} & \cdots & e^{i\hat y_k\omega_2}\\ \vdots & \vdots & \vdots\\ e^{i\hat y_1\omega_{n-1}} & \cdots & e^{i\hat y_k\omega_{n-1}}\end{pmatrix},\quad \Phi=\begin{pmatrix} e^{iy_1\omega_1} & \cdots & e^{iy_n\omega_1}\\ e^{iy_1\omega_2} & \cdots & e^{iy_n\omega_2}\\ \vdots & \vdots & \vdots\\ e^{iy_1\omega_{n-1}} & \cdots & e^{iy_n\omega_{n-1}}\end{pmatrix}.
$$

We shall prove that the following estimate holds:

$$
\max_{t=1,\cdots,T}\frac{1}{\sqrt{n}}||\hat\Phi\hat\alpha_t-\Phi\alpha_t||_2>2\sigma, \tag{5.5}
$$

and consequently arrive at (5.3).

Step 2. We let $\theta_j=y_j\frac{2\Omega}{n}, j=1,\cdots,n$ and $\hat\theta_j=\hat y_j\frac{2\Omega}{n}$. From the following decompositions:

$$
\begin{aligned}
\hat\Phi&=\big(\phi_{n-1}(e^{i\hat\theta_1}),\cdots,\phi_{n-1}(e^{i\hat\theta_k})\big)\mathrm{diag}(e^{-i\hat y_1\Omega},\cdots,e^{-i\hat y_k\Omega}),\\
\Phi&=\big(\phi_{n-1}(e^{i\theta_1}),\cdots,\phi_{n-1}(e^{i\theta_n})\big)\mathrm{diag}(e^{-iy_1\Omega},\cdots,e^{-iy_n\Omega}),
\end{aligned} \tag{5.6}
$$

where $\phi(\cdot)$ is defined as in (4.1), we readily obtain that

$$
\max_{t=1,\cdots,T}||\hat\Phi\hat\alpha_t-\Phi\alpha_t||_2=\max_{t=1,\cdots,T}||\hat D\hat\gamma_t-D\tilde\gamma_t||_2, \tag{5.7}
$$

where $\hat\gamma_t=(\hat a_{1,t}e^{-i\hat y_1\Omega},\cdots,\hat a_{k,t}e^{-i\hat y_k\Omega})^\top,\gamma_t=(I_t(y_1)a_1e^{-iy_1\Omega},\cdots,I_t(y_n)a_ne^{-iy_n\Omega})^\top$, $\hat D=\big(\phi_{n-1}(e^{i\hat\theta_1}),\cdots,\phi_{n-1}(e^{i\hat\theta_k})\big)$ and $D=\big(\phi_{n-1}(e^{i\theta_1}),\cdots,\phi_{n-1}(e^{i\theta_n})\big)$. We consider I in (2.7) and denote $B=I\mathrm{diag}(a_1e^{-iy_1\Omega},\cdots,a_ne^{-iy_n\Omega})$. Applying Theorem 4.2, we get

$$
\max_{t=1,\cdots,T}||\hat D\hat\gamma_t-D\gamma_t||_2\geq\frac{\sigma_{\infty,\min}(B)\xi(n-1)(\theta_{\min})^{n-1}}{\pi^{n-1}},
$$

where $\theta_{\min} = \min_{j\neq p} |\theta_j - \theta_p|_{2\pi}$. On the other hand, by the definition of $\sigma_{\infty,\min}$, we have

$$
\begin{aligned}
\sigma_{\infty,\min}(I) m_{\min} = & \min_{||\alpha||_\infty \geq m_{\min}} ||I\alpha||_\infty \\
\leq & \min_{||\alpha||_\infty \geq 1} ||I \mathrm{diag}(a_1 e^{-iy_1\Omega}, \cdots, a_n e^{-iy_n\Omega})\alpha||_\infty \quad (\text{by } m_{\min} = \min_{1\leq j\leq n} |a_j|) \\
= & \sigma_{\infty,\min}(B).
\end{aligned} \tag{5.8}
$$

Thus,

$$
\max_{t=1,\cdots,T} ||\hat{D}\hat{\gamma}_t - D\gamma_t||_2 \geq \frac{m_{\min}\sigma_{\infty,\min}(I)\xi(n-1)(\theta_{\min})^{n-1}}{\pi^{n-1}}.
$$

By (5.7), it follows that

$$
\max_{t=1,\cdots,T} ||\hat{\Phi}\hat{\alpha}_t - \Phi\alpha_t||_2 \geq \frac{m_{\min}\sigma_{\infty,\min}(I)\xi(n-1)(\theta_{\min})^{n-1}}{\pi^{n-1}}.
$$

On the other hand, recall that $d_{\min} = \min_{j\neq p} |y_j - y_p|_{\frac{n\pi}{\Omega}}$. Using the relation $\theta_j = y_j \frac{2\Omega}{n}$, we have $\theta_{\min} = \frac{2\Omega}{n} d_{\min}$. Then the separation condition (2.9) and $\frac{1}{\sigma_{\infty,\min}(I)} \frac{\sigma}{m_{\min}} \leq 1$ imply that

$$
\theta_{\min} \geq \frac{4.4e\pi}{n}\Big(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n}} \geq \frac{4.4e\pi}{n}\Big(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n-1}} > \pi\Big(\frac{2\sqrt{n}}{\xi(n-1)\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n-1}},
$$

where here we have used Lemma C.1 for deriving the last inequality. Therefore,

$$
\max_{t=1,\cdots,T} ||\hat{\Phi}\hat{\alpha}_t - \Phi\alpha_t||_2 > 2\sqrt{n}\sigma,
$$

whence (5.5) is proved.

Step 3. By the above results, the solution of (2.4) corresponds exactly to n point scatterers. Suppose that the solution is $\rho = \sum_{j=1}^n \hat{a}_j \delta_{\hat{y}_j}$ and $\hat{\mu}_t = I_t\rho = \sum_{j=1}^n \hat{a}_{j,t}\delta_{\hat{y}_j}$. We now prove the stability of the location recovery. Similarly to Step 1, using the constraints in (2.4)

$$
||\mathscr{F}[\hat{\mu}_t] - \mathbf{Y}_t||_2 < \sigma, \quad 1 \leq t \leq T,
$$

we can derive that

$$
\frac{1}{2\Omega}\int_0^h \max_{t=1,\cdots,T} \sum_{j=1}^n |\mathscr{F}[\hat{\mu}_t](\omega + (j-1)h - \Omega) - \mathscr{F}[\mu_t](\omega + (j-1)h - \Omega)|^2 d\omega < 4\sigma^2.
$$

Hence, there exists $\omega_0 \in [0, h]$ $(h = \frac{2\Omega}{n})$ such that

$$
\max_{t=1,\cdots,T} \frac{1}{n}\sum_{j=1}^n |\mathscr{F}[\hat{\mu}_t](\omega_0 + (j-1)h - \Omega) - \mathscr{F}[\mu_t](\omega_0 + (j-1)h - \Omega)|^2 < 4\sigma^2. \tag{5.9}
$$

Without loss of generality, we suppose that $\omega_0 = 0$ and consider

$$
\big(\mathscr{F}[\hat{\mu}_t](\omega_1), \mathscr{F}[\hat{\mu}_t](\omega_2), \cdots, \mathscr{F}[\hat{\mu}_t](\omega_n)\big)^\top - \big(\mathscr{F}[\mu_t](\omega_1), \mathscr{F}[\mu_t](\omega_2), \cdots, \mathscr{F}[\mu_t](\omega_n)\big)^\top = \hat{\Phi}\hat{\alpha}_t - \Phi\alpha_t,
$$

where $\omega_j = (j-1)h - \Omega$, $\hat{\alpha}_t = (\hat{a}_{1,t}, \cdots, \hat{a}_{n,t})^\top$, $\alpha_t = (I_t(y_1)a_1, \cdots, I_t(y_n)a_n)^\top$ and

$$\hat{\Phi} = \begin{pmatrix} e^{i\hat{y}_1\omega_1} & \cdots & e^{i\hat{y}_n\omega_1} \\ e^{i\hat{y}_1\omega_2} & \cdots & e^{i\hat{y}_n\omega_2} \\ \vdots & \vdots & \vdots \\ e^{i\hat{y}_1\omega_n} & \cdots & e^{i\hat{y}_n\omega_n} \end{pmatrix}, \quad \Phi = \begin{pmatrix} e^{iy_1\omega_1} & \cdots & e^{iy_n\omega_1} \\ e^{iy_1\omega_2} & \cdots & e^{iy_n\omega_2} \\ \vdots & \vdots & \vdots \\ e^{iy_1\omega_n} & \cdots & e^{iy_n\omega_n} \end{pmatrix}.$$

By (5.9), it is clear that

$$\max_{t=1,\cdots,T} ||\hat{\Phi}\hat{\alpha}_t - \Phi\alpha_t||_2 < 2\sqrt{n}\sigma.$$

Note that

$$\max_{t=1,\cdots,T} ||\hat{\Phi}\hat{\alpha}_t - \Phi\alpha_t||_2 = \max_{t=1,\cdots,T} ||\hat{D}\hat{\gamma}_t - D\gamma_t||_2, \tag{5.10}$$

where $\hat{\gamma}_t = (\hat{a}_{1,t}e^{-i\hat{y}_1\Omega}, \cdots, \hat{a}_{n,t}e^{-i\hat{y}_n\Omega})^\top, \gamma_t = (I_1(y_1)a_1e^{-iy_1\Omega}, \cdots, I_n(y_n)a_ne^{-iy_n\Omega})^\top$, $\hat{D} = \big(\phi_n(e^{i\hat{\theta}_1}), \cdots, \phi_n(e^{i\hat{\theta}_n})\big)$ and $D = \big(\phi_n(e^{i\theta_1}), \cdots, \phi_n(e^{i\theta_n})\big)$. Thus,

$$\max_{t=1,\cdots,T} ||\hat{D}\hat{\gamma}_t - D\gamma_t||_2 < 2\sqrt{n}\sigma. \tag{5.11}$$

We can apply Theorem 4.3 to get

$$||\eta_{n,n}(e^{i\theta_1}, \cdots, e^{i\theta_n}, e^{i\hat{\theta}_1}, \cdots, e^{i\hat{\theta}_n})||_\infty < \frac{2^{n+1}\sqrt{n}\sigma}{\sigma_{\infty,\min}(B)}, \tag{5.12}$$

where $\eta_{n,n}$ is defined by (4.4) and $B = I\text{diag}(a_1e^{-iy_1\Omega}, \cdots, a_ne^{-iy_n\Omega})$. By (5.8), it follows that

$$\sigma_{\infty,\min}(I)m_{\min} \le \sigma_{\infty,\min}(B).$$

Thus, we have

$$||\eta_{n,n}(e^{i\theta_1}, \cdots, e^{i\theta_n}, e^{i\hat{\theta}_1}, \cdots, e^{i\hat{\theta}_n})||_\infty < \frac{2^{n+1}}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}. \tag{5.13}$$

Step 4. We apply Lemma 4.2 to estimate $|\hat{\theta}_j - \theta_j|_{2\pi}$'s. For this purpose, let $\epsilon = \frac{2\sqrt{n}\pi^n}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}$. It is clear that $||\eta_{n,n}||_\infty < (\frac{2}{\pi})^n\epsilon$ and we only need to check the following condition:

$$\theta_{\min} \ge \Big(\frac{4\epsilon}{\lambda(n)}\Big)^{\frac{1}{n}}, \quad \text{or equivalently } (\theta_{\min})^n \ge \frac{4\epsilon}{\lambda(n)}. \tag{5.14}$$

Indeed, by $\theta_{\min} = \frac{2\Omega}{n}d_{\min}$ and the separation condition (2.9),

$$\theta_{\min} \ge \frac{4.4\pi e}{n}\Big(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n}} \ge \Big(\frac{8\sqrt{n}\pi^n}{\lambda(n)}\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n}}. \tag{5.15}$$

Here, we have used Lemma C.2 for deriving the last inequality. Then, we get (5.14). Therefore, we can apply Lemma 4.2 to get that, after reordering $\hat{\theta}_j$'s,

$$\Big|\hat{\theta}_j - \theta_j\Big|_{2\pi} < \frac{\theta_{\min}}{2}, \text{ and } \Big|\hat{\theta}_j - \theta_j\Big|_{2\pi} < \frac{2^n\sqrt{n}\pi^n}{(n-2)!(\theta_{\min})^{n-1}}\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}, \; j = 1, \cdots, n. \tag{5.16}$$

Finally, we estimate $|\hat{y}_j - y_j|_{\frac{n\pi}{\Omega}}$. Since $|\hat{\theta}_j - \theta_j|_{2\pi} < \frac{\theta_{\min}}{2}$, we have after reordering the $\hat{y}_j'$s,

$$|\hat{y}_j - y_j|_{\frac{n\pi}{\Omega}} < \frac{d_{\min}}{2}.$$

On the other hand, $\left|\hat{y}_j - y_j\right|_{\frac{n\pi}{\Omega}} = \frac{n}{2\Omega}\left|\hat{\theta}_j - \theta_j\right|_{2\pi}$. Combining (5.16) and (C.1), a direct calculation shows that

$$\left|\hat{y}_j - y_j\right|_{\frac{n\pi}{\Omega}} < \frac{C(n)}{\Omega}\left(\frac{\pi}{\Omega d_{\min}}\right)^{n-1}\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}},$$

where $C(n) = 2\sqrt{2}ne^n\sqrt{\pi}$.

6. Proof of Theorem 3.1

6.1. Number and location recoveries in one-dimensional case

We first introduce some results for the number and location recoveries of the one-dimensional case, which will help us to derive the stability results for two-dimensional super-resolution. Unlike Theorem 2.1, the stability results here consider Euclidean distance between point scatterers.

For source $\mu = \sum_{j=1}^n a_j\delta_{y_j}$ and illumination patterns I_t's, the measurements are

$$\mathbf{Y}_t(\omega) = \mathscr{F}[I_t\mu](\omega) + \mathbf{W}_t(\omega) = \sum_{j=1}^n I_t(y_j)a_je^{iy_j\omega} + \mathbf{W}_t(\omega), \quad 1 \le t \le T,\ \omega \in [-\Omega, \Omega], \tag{6.1}$$

where $\mathscr{F}[I_t\mu]$ denotes the Fourier transform of $I_t\mu$ and $\mathbf{W}_t(\omega)$ is the noise with $||\mathbf{W}_t||_\infty < \sigma$.

Theorem 6.1. *Suppose the measurements $\mathbf{Y}_t$'s in (6.1) are generated from $\mu = \sum_{j=1}^n a_j\delta_{y_j}$, $y_j \in \mathbb{R}$ where y_j's are in an interval $\mathcal{O}$ of length $\frac{c_0 n\pi}{\Omega}$ with $c_0 \ge 1$ and satisfy*

$$d_{\min} := \min_{p\neq j}\left|y_p - y_j\right| \ge \frac{4.4c_0e\pi}{\Omega}\Big(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n}} \tag{6.2}$$

with $\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}} \le 1$. Then there is no $k < n$ locations $\hat{y}_j \in \mathbb{R}, j = 1,\cdots,k$ such that there exists $\hat{\mu}_t = \sum_{j=1}^k \hat{a}_{j,t}\delta_{\hat{y}_j}$'s so that

$$||\mathscr{F}[\hat{\mu}_t] - \mathbf{Y}_t||_\infty < \sigma, \quad t = 1,\cdots,T.$$

Proof. Let $\mu_t = \sum_{j=1}^n a_jI_t(y_j)\delta_{y_j}$. Similar to Step 1 and Step 2 in the proof of Theorem 2.1, we only need to prove that if $k < n$, then for any $\hat{y}_j \in \mathbb{R}, \hat{a}_{j,t} \in \mathbb{C}, j = 1,\cdots,k, t = 1,\cdots,T$,

$$\max_{t=1,\cdots,T}||\mathscr{F}[\hat{\mu}_t] - \mathscr{F}[\mu_t]||_\infty > 2\sigma. \tag{6.3}$$

Specifically, for $k < n$, we consider

$$\big(\mathscr{F}[\hat{\mu}_t](\omega_1), \mathscr{F}[\hat{\mu}_t](\omega_2),\cdots,\mathscr{F}[\hat{\mu}_t](\omega_{n-1})\big)^\top - \big(\mathscr{F}[\mu_t](\omega_1), \mathscr{F}[\mu_t](\omega_2),\cdots,\mathscr{F}[\mu_t](\omega_{n-1})\big)^\top, \tag{6.4}$$

where $\omega_j = (j-1)h - \Omega,\ j = 1, \cdots, n-1$ with $h = \frac{\Omega}{c_0 n}$. We write (6.4) as

$$\hat{\Phi}\hat{\alpha}_t - \Phi\alpha_t,$$

where $\hat{\alpha}_t = (\hat{a}_{1,t}, \cdots, \hat{a}_{k,t})^\top$, $\alpha_t = (I_t(y_1)a_1, \cdots, I_t(y_n)a_n)^\top$ and

$$\hat{\Phi} = \begin{pmatrix} e^{i\hat{y}_1\omega_1} & \cdots & e^{i\hat{y}_k\omega_1} \\ e^{i\hat{y}_1\omega_2} & \cdots & e^{i\hat{y}_k\omega_2} \\ \vdots & \vdots & \vdots \\ e^{i\hat{y}_1\omega_{n-1}} & \cdots & e^{i\hat{y}_k\omega_{n-1}} \end{pmatrix}, \quad \Phi = \begin{pmatrix} e^{iy_1\omega_1} & \cdots & e^{iy_n\omega_1} \\ e^{iy_1\omega_2} & \cdots & e^{iy_n\omega_2} \\ \vdots & \vdots & \vdots \\ e^{iy_1\omega_{n-1}} & \cdots & e^{iy_n\omega_{n-1}} \end{pmatrix}.$$

We shall prove that the following estimate holds:

$$\max_{t=1,\cdots,T} \frac{1}{\sqrt{n}} ||\hat{\Phi}\hat{\alpha}_t - \Phi\alpha_t||_2 > 2\sigma, \tag{6.5}$$

and consequently it yields (6.3). Let $\theta_j = y_j h = y_j \frac{\Omega}{c_0 n}$ and $\hat{\theta}_j = \hat{y}_j h = \hat{y}_j \frac{\Omega}{c_0 n}$. Similar to Step 2 in the proof of Theorem 2.1, we can have

$$\max_{t=1,\cdots,T} ||\hat{\Phi}\hat{\alpha}_t - \Phi\alpha_t||_2 \geq \frac{m_{\min}\sigma_{\infty,\min}(I)\xi(n-1)(\theta_{\min})^{n-1}}{\pi^{n-1}},$$

where $\theta_{\min} = \min_{p\neq j} |\theta_j - \theta_p|_{2\pi}$. Because y_j's are in an interval of length $\frac{c_0 n\pi}{\Omega}$, by $\theta_j = y_j h$ we have $\theta_{\min} = \min_{p\neq j} |\theta_j - \theta_p|_{2\pi} = d_{\min}\frac{\Omega}{c_0 n}$. Then the separation condition (6.2) and $\frac{1}{\sigma_{\infty,\min}(I)\frac{\sigma}{m_{\min}}} \leq 1$ imply that

$$\theta_{\min} \geq \frac{4.4e\pi}{n}\Big(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n}} \geq \frac{4.4e\pi}{n}\Big(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n-1}} > \pi\Big(\frac{2\sqrt{n}}{\xi(n-1)\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n-1}},$$

where here we have used Lemma C.1 for deriving the last inequality. Therefore,

$$\max_{t=1,\cdots,T} ||\hat{\Phi}\hat{\alpha}_t - \Phi\alpha_t||_2 > 2\sqrt{n}\sigma,$$

whence we prove (6.5). □

Theorem 6.2. *Suppose that the measurements* $\mathbf{Y}_t$*'s in (6.1) are generated from* $\mu = \sum_{j=1}^n a_j\delta_{y_j}, y_j \in \mathbb{R}$*, where* y_j*'s are in an interval* $\mathcal{O}$ *of length* $\frac{c_0 n\pi}{\Omega}$ *with* $c_0 \geq 1$ *and satisfy*

$$d_{\min} := \min_{p\neq j}\Big|y_p - y_j\Big| \geq \frac{4.4c_0 e\pi}{\Omega}\Big(\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}\Big)^{\frac{1}{n}} \tag{6.6}$$

with $\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}} \leq 1$*. Moreover, for* $\hat{\mu}_t = \sum_{j=1}^n \hat{a}_{j,t}\delta_{\hat{y}_j}, \hat{y}_j \in \mathcal{O}$ *satisfying* $||\mathscr{F}[\hat{\mu}_t] - \mathbf{Y}_t||_\infty < \sigma, t = 1, \cdots, T$*, after reordering the* $\hat{y}_j$*'s, we have*

$$\Big|\hat{y}_j - y_j\Big| < \frac{d_{\min}}{2}, \tag{6.7}$$

and

$$\Big|\hat{y}_j - y_j\Big| < \frac{C(n)}{\Omega}SRF^{n-1}\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}}, \quad 1 \leq j \leq n, \tag{6.8}$$

where $C(n) = 2^n\sqrt{2\pi}nc_0^{n-1}e^n$ *and* $SRF = \frac{\pi}{\Omega d_{\min}}$ *is the super-resolution factor.*

Proof. Let $\mu_t = \sum_{j=1}^n a_j I_t(y_j)\delta_{y_j}$. Similar to the proof of Theorem 2.1, we consider

$$\left(\mathscr{F}[\hat\mu_t](\omega_1),\mathscr{F}[\hat\mu_t](\omega_2),\cdots,\mathscr{F}[\hat\mu_t](\omega_n)\right)^\top - \left(\mathscr{F}[\mu_t](\omega_1),\mathscr{F}[\mu_t](\omega_2),\cdots,\mathscr{F}[\mu_t](\omega_n)\right)^\top = \hat\Phi\hat\alpha_t - \Phi\alpha_t,$$

where $\omega_j = (j-1)h-\Omega$ with $h=\frac{\Omega}{c_0 n}$, $\hat\alpha_t = (\hat a_{1,t},\cdots,\hat a_{n,t})^\top$, $\alpha_t = (I_t(y_1)a_1,\cdots,I_t(y_n)a_n)^\top$ and

$$\hat\Phi = \begin{pmatrix} e^{i\hat y_1\omega_1} & \dots & e^{i\hat y_n\omega_1}\\ e^{i\hat y_1\omega_2} & \dots & e^{i\hat y_n\omega_2}\\ \vdots & \vdots & \vdots\\ e^{i\hat y_1\omega_n} & \dots & e^{i\hat y_n\omega_n}\end{pmatrix}, \quad \Phi = \begin{pmatrix} e^{i y_1\omega_1} & \dots & e^{i y_n\omega_1}\\ e^{i y_1\omega_2} & \dots & e^{i y_n\omega_2}\\ \vdots & \vdots & \vdots\\ e^{i y_1\omega_n} & \dots & e^{i y_n\omega_n}\end{pmatrix}.$$

By the constraint on the noise, it is clear that

$$\max_{t=1,\cdots,T} ||\hat\Phi\hat\alpha_t - \Phi\alpha_t||_2 < 2\sqrt{n}\sigma.$$

Let $\theta_j = y_j h$ and $\hat\theta_j = \hat y_j h$. Similar to the proof of Theorem 2.1, we can prove that, after reordering $\hat\theta_j$'s,

$$\left|\hat\theta_j-\theta_j\right|_{2\pi} < \frac{\theta_{\min}}{2}, \text{ and } \left|\hat\theta_j-\theta_j\right|_{2\pi} < \frac{2^n\sqrt{n}\pi^n}{(n-2)!(\theta_{\min})^{n-1}}\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}},\ j=1,\cdots,n. \tag{6.9}$$

Finally, we estimate $|\hat y_j - y_j|$. Since $|\hat\theta_j-\theta_j|_{2\pi} < \frac{\theta_{\min}}{2}$ and $\hat y_j$'s, y_j's are in $\mathscr{O}$, we have after reordering the $\hat y_j$'s,

$$|\hat y_j - y_j| < \frac{d_{\min}}{2}.$$

On the other hand, $\left|\hat y_j - y_j\right| = \frac{nc_0}{\Omega}\left|\hat\theta_j-\theta_j\right|_{2\pi}$. Combining (6.9) and (C.1), a direct calculation shows that

$$\left|\hat y_j - y_j\right| < \frac{C(n)}{\Omega}\left(\frac{\pi}{\Omega d_{\min}}\right)^{n-1}\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}},$$

where $C(n) = 2^n\sqrt{2\pi}nc_0^{n-1}e^n$. □

6.2. Projection lemmas

Next we introduce two auxiliary lemmas whose ideas are from [23]. We introduce some notation. For $0<\theta\le\frac{\pi}{2}$ and $N=\lfloor\frac{\pi}{\theta}\rfloor$, we denote the unit vectors in $\mathbb{R}^2$ by

$$\mathbf{v}(\tau\theta) = \left(\cos(\tau\theta),\sin(\tau\theta)\right)^T, \quad 1\le\tau\le N. \tag{6.10}$$

It is obvious that there are N different unit vectors of the form (6.10).

For a vector $\mathbf{v}\in\mathbb{R}^2$, we denote $\mathscr{P}_{\mathbf{v}}$ the projection to the one-dimensional space spanned by $\mathbf{v}$. We have the following lemmas.

Lemma 6.1. *Let $n \geq 2$ and $\mathbf{y}_1, \cdots, \mathbf{y}_n$ be n different points in $\mathbb{R}^2$. Let $d_{\min} = \min_{p \neq j} ||\mathbf{y}_p - \mathbf{y}_j||_2$ and $\Delta = \frac{\pi}{(n+2)(n+1)}$. Then there exist $n+1$ unit vectors $\mathbf{v}_q$'s such that $0 \leq \mathbf{v}_p \cdot \mathbf{v}_j \leq \cos(2\Delta)$ for $p \neq j$ and*

$$\min_{p \neq j, 1 \leq p, j \leq n} ||\mathscr{P}_{\mathbf{v}_q}(\mathbf{y}_p) - \mathscr{P}_{\mathbf{v}_q}(\mathbf{y}_j)||_2 \geq \frac{2\Delta d_{\min}}{\pi}, \quad q = 1, \cdots, n+1. \tag{6.11}$$

Proof. Note that there are at most $\frac{n(n-1)}{2}$ different vectors of the form $\mathbf{u}_{pj} = \mathbf{y}_p - \mathbf{y}_j, p < j$. For each $\mathbf{u}_{pj}$, consider the set $N(\mathbf{u}_{pj}, \Delta) = \left\{\mathbf{v} \middle| ||\mathbf{v}||_2 = 1, \mathbf{v} \in \mathbb{R}^2, |\mathbf{v} \cdot \mathbf{u}| < ||\mathbf{u}||_2 \sin \Delta \right\}$. Let $\theta = 2\Delta$ and introduce the vectors $\mathbf{v}(\tau\theta)$ as in (6.10). It is clear that $0 \leq \mathbf{v}(\tau_1\theta) \cdot \mathbf{v}(\tau_2\theta) \leq \cos(\theta)$ for $\tau_1 \neq \tau_2$. Thus if $|\mathbf{v}(\tau_1\theta) \cdot \mathbf{u}| < ||\mathbf{u}||_2 \sin \Delta$, then, for other $\tau_2 \neq \tau_1$, we have $|\mathbf{v}(\tau_2\theta) \cdot \mathbf{u}| \geq ||\mathbf{u}||_2 \sin \Delta$. We can derive that each set $N(\mathbf{u}_{pj}, \Delta)$ contains at most one of the vectors $\mathbf{v}(\tau\theta)$'s. As a result, $\cup_{p<j, 1 \leq j, p \leq n} N(\mathbf{u}_{pj}, \Delta)$ contains at most $\frac{n(n-1)}{2}$ vectors of the form $\mathbf{v}(\tau\theta)$.

Next recall that there are N different vectors of the form in (6.10), where $N = \lfloor \frac{\pi}{\theta} \rfloor$. Since

$$\theta = 2\Delta = \frac{2\pi}{(n+2)(n+1)},$$

we have

$$N \geq \lfloor \frac{\pi}{\theta} \rfloor = \frac{(n+2)(n+1)}{2} > \frac{(n+1)^2}{2}.$$

Note that $\frac{(n+1)^2}{2} - \frac{n(n-1)}{2} = n+1$, we can find $n+1$ vectors of the form $\mathbf{v}(\tau\theta)$ that are not contained in the set $\cup_{p<j, 1 \leq j, p \leq n} N(\mathbf{u}_{pj}, \Delta)$. That is, we can find $n+1$ unit vectors, say, $\mathbf{v}_q$, $1 \leq q \leq n+1$, which satisfy (6.11). □

Lemma 6.2. *For a vector $\mathbf{u} \in \mathbb{R}^2$, and two unit vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ satisfying $0 \leq \mathbf{v}_1 \cdot \mathbf{v}_2 \leq \cos(\theta)$, we have*

$$|\mathbf{v}_1 \cdot \mathbf{u}|^2 + |\mathbf{v}_2 \cdot \mathbf{u}|^2 \geq (1 - \cos(\theta))||\mathbf{u}||_2^2. \tag{6.12}$$

Proof. For $\mathbf{u} \in \mathbb{R}^2$, and two unit vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ satisfying $0 \leq \mathbf{v}_1 \cdot \mathbf{v}_2 \leq \cos\theta$, we have

$$\left|\left|(\mathbf{v}_1 \cdot \mathbf{u}, \ \mathbf{v}_2 \cdot \mathbf{u})^T\right|\right|_2^2 = \left|\left|\begin{pmatrix}\mathbf{v}_1^T \\ \mathbf{v}_2^T\end{pmatrix} \cdot \mathbf{u}\right|\right|_2^2 \geq \sigma_{\min}^2\left(\begin{pmatrix}\mathbf{v}_1^T \\ \mathbf{v}_2^T\end{pmatrix}\right)||\mathbf{u}||_2^2 \geq (1 - \cos\theta)||\mathbf{u}||_2^2, \tag{6.13}$$

where the last inequality follows from calculating $\sigma_{\min}\left(\begin{pmatrix}\mathbf{v}_1^T \\ \mathbf{v}_2^T\end{pmatrix}\right)$. □

6.3. PROOF OF THEOREM 3.1

Proof. We only prove the theorem for problem (3.3). The other cases can be proved in a similar manner. Let the measurements $\mathbf{Y}_t$'s be generated by $\mu = \sum_{j=1}^n a_j \delta_{\mathbf{y}_j}, \mathbf{y}_j \in \mathscr{O}$ satisfying the minimum separation condition

$$d_{\min}^{(2)} := \min_{p \neq j, 1 \leq p, j \leq n} \left|\left|\mathbf{y}_p - \mathbf{y}_j\right|\right|_2 \geq \frac{2.2 c_0 e \pi (n+2)(n+1)}{\Omega} \left(\frac{1}{\sigma_{\infty,\min}(I)} \frac{\sigma}{m_{\min}}\right)^{\frac{1}{n}}. \tag{6.14}$$

Let $\Delta = \frac{\pi}{(n+2)(n+1)}$. By Lemma 6.1, there exist $n+1$ unit vectors $\mathbf{v}_q$'s so that

$$0 \leq \mathbf{v}_p \cdot \mathbf{v}_j \leq \cos 2\Delta, 1 \leq p < j \leq n,$$

and for each q,

$$\min_{p \neq j} \left|\left| \mathscr{P}_{\mathbf{v}_q}(\mathbf{y}_p) - \mathscr{P}_{\mathbf{v}_q}(\mathbf{y}_j) \right|\right|_2 \geq d_{\min}^{(1)}, \tag{6.15}$$

where

$$d_{\min}^{(1)} = \min_{p \neq j} \left|\left| \mathbf{y}_p - \mathbf{y}_j \right|\right|_2 \frac{2\Delta}{\pi} = \frac{2d_{\min}^{(2)}}{(n+2)(n+1)} \geq \frac{4.4c_0 e\pi}{\Omega} \Big(\frac{1}{\sigma_{\infty,\min}(I)} \frac{\sigma}{m_{\min}} \Big)^{\frac{1}{n}}. \tag{6.16}$$

Because the projected point scatterers in these one-dimensional subspaces are separated by a distance beyond (6.16), by the constraints on the measurements and Theorem 6.1, any solution to (3.3) must contain n point scatterers. Assume that $\rho = \sum_{j=1}^n \hat{a}_j \delta_{\hat{\mathbf{y}}_j}, \hat{\mathbf{y}}_j \in \mathscr{O}$ is a solution to (3.3). Then we consider

$$\mu_t = I_t \mu = \sum_{j=1}^n a_j I_t(\mathbf{y}_j) \delta_{\mathbf{y}_j}, \quad \hat{\mu}_t = I_t \rho = \sum_{j=1}^n \hat{a}_{j,t} \delta_{\hat{\mathbf{y}}_j},$$

and the projected measures

$$\sum_{j=1}^n a_j I_t(\mathbf{y}_j) \delta_{\mathscr{P}_{\mathbf{v}_q}(\mathbf{y}_j)}, \quad \sum_{j=1}^n a_{j,t} \delta_{\mathscr{P}_{\mathbf{v}_q}(\hat{\mathbf{y}}_j)}$$

in the one-dimensional subspace spanned by $\mathbf{v}_q$. We also consider the corresponding measurements $\mathbf{Y}_t(\boldsymbol{\omega})$, $\boldsymbol{\omega}$ in the subspace spanned by $\mathbf{v}_q$. It is clear that for each q, $\mathscr{P}_{\mathbf{v}_q}(\mathbf{y}_j)$'s are in an interval (i.e., in a one-dimensional subspace) of length $\frac{c_0 n\pi}{\Omega}$ and the separation condition (6.16) is satisfied. On the other hand, the measurement constraints in (3.3) still hold for the measurements $\mathbf{Y}_t(\boldsymbol{\omega})$ in that subspace. By Theorem 6.2, we can conclude that for each q, we have a permutation τ_q of $\{1, \cdots, n\}$ so that

$$\left|\left| \mathscr{P}_{\mathbf{v}_q}(\hat{\mathbf{y}}_{\tau_q(j)}) - \mathscr{P}_{\mathbf{v}_q}(\mathbf{y}_j) \right|\right|_2 < \frac{C(n)}{\Omega} \Big(\frac{\pi}{d_{\min}^{(1)} \Omega} \Big)^{n-1} \frac{1}{\sigma_{\infty,\min}(I)} \frac{\sigma}{m_{\min}}, \quad 1 \leq j \leq n, \tag{6.17}$$

where $C(n) = 2^n \sqrt{2\pi} n c_0^{n-1} e^n$.

Note that, for fixed j in (6.17), we have $(n+1)$ different $\tau_q(j)$'s, while $\hat{\mathbf{y}}_p$'s take at most n values. Therefore, by the pigeonhole principle, for each fixed $\mathbf{y}_j$, we can find two different q's, say, q_1 and q_2, such that $\hat{\mathbf{y}}_{\tau_{q_1}(j)} = \hat{\mathbf{y}}_{\tau_{q_2}(j)} = \hat{\mathbf{y}}_{p_j}$ for some p_j. Since $0 \leq \mathbf{v}_{q_1} \cdot \mathbf{v}_{q_2} \leq \cos 2\Delta$, we can apply Lemma 6.2 to get

$$\left|\left| \hat{\mathbf{y}}_{p_j} - \mathbf{y}_j \right|\right|_2 < \frac{\sqrt{2}}{\sqrt{1-\cos(2\Delta)}} \frac{C(n)}{\Omega} \Big(\frac{\pi}{d_{\min}^{(1)} \Omega} \Big)^{n-1} \frac{1}{\sigma_{\infty,\min}(I)} \frac{\sigma}{m_{\min}}, \quad 1 \leq j \leq n.$$

Using the inequality

$$1 - \cos 2\Delta \geq \frac{8}{\pi^2} \Delta^2 \geq \frac{8}{(n+2)^2(n+1)^2},$$

we further obtain

$$\left|\left|\hat{\mathbf{y}}_{p_j}-\mathbf{y}_j\right|\right|_2<\frac{(n+2)(n+1)C(n)}{2\Omega}\Big(\frac{\pi}{d_{\min}^{(1)}\Omega}\Big)^{n-1}\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}},\quad 1\leq j\leq n. \tag{6.18}$$

Combining together (6.14) and (6.16), we can verify by a direct calculation that

$$\left|\left|\hat{\mathbf{y}}_{p_j}-\mathbf{y}_j\right|\right|_2<\frac{1}{2}d_{\min}^{(2)},\quad j=1,\cdots,n.$$

Thus we can reorder the $\hat{\mathbf{y}}_j$'s so that

$$\left|\left|\hat{\mathbf{y}}_{j}-\mathbf{y}_j\right|\right|_2<\frac{1}{2}d_{\min}^{(2)},\quad j=1,\cdots,n.$$

By (6.18), we obtain that

$$\left|\left|\hat{\mathbf{y}}_{j}-\mathbf{y}_j\right|\right|_2<\frac{((n+2)(n+1)/2)^nC(n)}{\Omega}\Big(\frac{\pi}{d_{\min}^{(2)}\Omega}\Big)^{n-1}\frac{1}{\sigma_{\infty,\min}(I)}\frac{\sigma}{m_{\min}},\quad 1\leq j\leq n,$$

which follows from (6.18) and the fact that $d_{\min}^{(1)}=\frac{2d_{\min}^{(2)}}{(n+2)(n+1)}$. □

A. PROOFS OF LEMMAS 4.1 AND 4.2

A.1. AUXILIARY LEMMA

The following result is useful for proving Lemmas 4.1 and 4.2.

Lemma A.1. *For* $0<p\leq q<\min\{p+\pi,2\pi\}$ *and sufficiently small* $\Delta>0$*, we have*

$$|(1-e^{ip-i\Delta})(1-e^{iq+i\Delta})|-|(1-e^{ip})(1-e^{iq})|<0. \tag{A.1}$$

Proof. We only prove (A.1) in the case when $p\leq\pi$. When $p>\pi$, proving (A.1) is equivalent to showing

$$|(1-e^{i(2\pi-q)-i\Delta})(1-e^{i(2\pi-p)+i\Delta})|-|(1-e^{i(2\pi-q)})(1-e^{i(2\pi-p)})|<0.$$

Thus, the case when $p>\pi$ can be reduced to the case $p<\pi$. We first prove the result for $p\neq q$. Introduce the function $g(\Delta)=|(1-e^{ip-i\Delta})(1-e^{iq+i\Delta})|-|(1-e^{ip})(1-e^{iq})|$. We only need to show that

$$\frac{dg(\Delta)}{d\Delta}\Big|_{\Delta=0}<0,\quad \text{for } 0<p<\pi,\ p<q<p+\pi.$$

We calculate that

$$\frac{dg(\Delta)}{d\Delta}\Big|_{\Delta=0}=\frac{d|(1-e^{i(p-\Delta)})(1-e^{i(q+\Delta)})|}{d\Delta}\Big|_{\Delta=0}=\frac{d|1-e^{i(p-\Delta)}-e^{i(q+\Delta)}+e^{i(p+q)}|}{d\Delta}\Big|_{\Delta=0}=\frac{d\sqrt{g_1(\Delta)}}{d\Delta}\Big|_{\Delta=0},$$

where

$$g_1(\Delta)=\Big[1+\cos(p+q)-\cos(p-\Delta)-\cos(q+\Delta)\Big]^2+\Big[\sin(p+q)-\sin(p-\Delta)-\sin(q+\Delta)\Big]^2.$$

Thus, we have

$$\begin{aligned}
&\frac{dg(\Delta)}{d\Delta}\Big|_{\Delta=0} = \frac{g_1^{'}(0)}{2\sqrt{g_1(0)}}\\
=&\frac{\Big[(1+\cos(p+q)-\cos(p)-\cos(q))(-\sin(p)+\sin(q))+(\sin(p+q)-\sin(p)-\sin(q))(\cos(p)-\cos(q))\Big]}{\sqrt{g_1(0)}}\\
=&\frac{(1+\cos(p+q))(\sin(q)-\sin(p))+\sin(p+q)(\cos(p)-\cos(q))+2\sin(p-q)}{\sqrt{g_1(0)}}\\
=&\frac{2\sin(p-q)+2\sin(q)-2\sin(p)}{\sqrt{g_1(0)}} = 2\frac{\sin(p-q)+\sin(q-p)\cos(p)+\cos(q-p)\sin(p)-\sin(p)}{\sqrt{g_1(0)}}\\
=&2\frac{\sin(q-p)(\cos(p)-1)+\sin(p)(\cos(q-p)-1)}{\sqrt{g_1(0)}}\\
<&0 \quad (\text{when } 0<p\le\pi \text{ and } p<q<p+\pi).
\end{aligned}$$

This proves the lemma for $p \neq q$. When $p = q$, $\frac{dg(\Delta)}{d\Delta}\Big|_{\Delta=0} = 0$, and we shall show

$$\frac{d^2g(\Delta)}{d\Delta^2}\Big|_{\Delta=0} < 0, \quad \text{for } 0 < p < 2\pi,$$

which proves (A.1) for $p = q$. We calculate that

$$\frac{dg(\Delta)}{d\Delta}\Big|_{\Delta=0} = \frac{d|(1-e^{i(p-\Delta)})(1-e^{i(p+\Delta)})|}{d\Delta}\Big|_{\Delta=0} = \frac{d|1-e^{i(p-\Delta)}-e^{i(p+\Delta)}+e^{i(2p)}|}{d\Delta}\Big|_{\Delta=0} = \frac{d\sqrt{g_2(\Delta)}}{d\Delta}\Big|_{\Delta=0},$$

where

$$g_2(\Delta) = \Big[1+\cos(2p)-\cos(p-\Delta)-\cos(p+\Delta)\Big]^2 + \Big[\sin(2p)-\sin(p-\Delta)-\sin(p+\Delta)\Big]^2.$$

Thus, we obtain that

$$\frac{d^2g(\Delta)}{d\Delta^2}\Big|_{\Delta=0} = \frac{d\frac{g_2'(\Delta)}{2\sqrt{g_2(\Delta)}}}{d\Delta}\Big|_{\Delta=0} = \frac{2g_2^{''}(0)g_2(0)-[g_2^{'}(0)]^2}{4g_2(0)^{\frac{3}{2}}} = \frac{g_2^{''}(0)g_2(0)}{2g_2(0)^{\frac{3}{2}}} = \frac{g_2^{''}(0)}{2g_2(0)^{\frac{1}{2}}}$$

and

$$\begin{aligned}
&\frac{g_2''(0)}{2} = \Big[2\cos(p)(1+\cos(2p)-2\cos(p))+2\sin(p)(\sin(2p)-2\sin(p))\Big]\\
=&2[\cos(p)(1+\cos(2p))+\sin(p)\sin(2p)-2]\\
=&2(2\cos(p)-2)<0.
\end{aligned}$$

This completes the proof. □

The proof of Lemma 4.1 is divided into two steps.

Proof. We only need to consider $\theta_j, \hat{\theta}_p \in [0, 2\pi), 1 \le j \le k+1, 1 \le p \le k$. We can also suppose that $\theta_1 < \theta_2 < \cdots < \theta_{k+1}$ and $\hat{\theta}_1 \le \hat{\theta}_2 \le \cdots \le \hat{\theta}_k$.

Step 1. Define

$$\eta_j(\hat{\theta}_1, \cdots, \hat{\theta}_k) = \Pi_{q=1}^k (e^{i\theta_j} - e^{i\hat{\theta}_q}), \ j = 1, \cdots, k+1.$$

Hence, $||\eta_{k+1,k}(e^{i\theta_1}, \cdots, e^{i\theta_{k+1}}, e^{i\hat{\theta}_1}, \cdots, e^{i\hat{\theta}_k})||_\infty = \max_{1 \le j \le k+1} |\eta_j(\hat{\theta}_1, \cdots, \hat{\theta}_k)|$. We only need to show that

$$\min_{\hat{\theta}_j, 1 \le j \le k} \max_{j=1,\cdots,k+1} |\eta_j(\hat{\theta}_1, \cdots, \hat{\theta}_k)| \ge \xi(k) d_{\min}^k. \tag{A.2}$$

It is easy to verify this result for $k = 1$. For $k \ge 2$, we argue as follows. It is clear that a minimizer of (A.2) does exist (but may not be unique). Let $(\hat{\theta}_1, \cdots, \hat{\theta}_k)$ be a minimizer of (A.2) with $\hat{\theta}_1 \le \hat{\theta}_2 \cdots \le \hat{\theta}_k$. Because $\{\theta_1, \cdots, \theta_{k+1}\}$ can separate $k+1$ disjoint regions of $[0, 2\pi)$, say $In_1 = [\theta_1, \theta_2), \cdots, In_2 = [\theta_k, \theta_{k+1}), In_{k+1} = [\theta_k, 2\pi) \cup [0, \theta_1)$, but $\{\hat{\theta}_1, \cdots, \hat{\theta}_k\}$ only has k points, there is at least one region In_p so that $In_p \cap \{\hat{\theta}_1, \cdots, \hat{\theta}_k\} = \emptyset$. Without loss of generality, we suppose that $In_{k+1} \cap \{\hat{\theta}_1, \cdots, \hat{\theta}_k\} = \emptyset$ (otherwise we can realize this by rotating the $e^{i\theta_j}$'s). Thus, $\hat{\theta}_1, \cdots, \hat{\theta}_k \in [\theta_1, \theta_k]$. We then have the following claim.

Claim. *Each interval* $[\theta_j, \theta_{j+1})$ *contains only one* $\hat{\theta}_q$ *in* $\{\hat{\theta}_1, \cdots, \hat{\theta}_k\}$.

We prove the claim by considering the following cases.

Case 1: There exists j_0, p such that $\hat{\theta}_{j_0}, \hat{\theta}_{j_0+1} \in (\theta_p, \theta_{p+1})$.

Denote j_1 the number in $\{1, \cdots, k+1\}$ such that

$$\max_{j=1,\cdots,k+1} |\eta_j(\hat{\theta}_1, \cdots, \hat{\theta}_k)| = |\eta_{j_1}(\hat{\theta}_1, \cdots, \hat{\theta}_k)| > 0. \tag{A.3}$$

Let $\Delta > 0$ be sufficiently small. Then, for $j_1 \in \{1, \cdots, p\}$,

$$\begin{aligned}
&|\eta_{j_1}(\hat{\theta}_1, \cdots, \hat{\theta}_{j_0} - \Delta, \hat{\theta}_{j_0+1} + \Delta, \hat{\theta}_{j_0+2}, \cdots, \hat{\theta}_k)| - |\eta_{j_1}(\hat{\theta}_1, \cdots, \hat{\theta}_k)| \\
=& \Big[|(e^{i\theta_{j_1}} - e^{i\hat{\theta}_{j_0} - i\Delta})(e^{i\theta_{j_1}} - e^{i\hat{\theta}_{j_0+1} + i\Delta})| - |(e^{i\theta_{j_1}} - e^{i\hat{\theta}_{j_0}})(e^{i\theta_{j_1}} - e^{i\hat{\theta}_{j_0+1}})| \Big] \Pi_{q=1, q \ne j_0, j_0+1}^k |(e^{i\theta_{j_1}} - e^{i\hat{\theta}_q})| \\
=& \Big[|(1 - e^{i\hat{\theta}_{j_0} - i\theta_{j_1} - i\Delta})(1 - e^{i\hat{\theta}_{j_0+1} - i\theta_{j_1} + i\Delta})| - |(1 - e^{i\hat{\theta}_{j_0} - i\theta_{j_1}})(1 - e^{i\hat{\theta}_{j_0+1} - i\theta_{j_1}})| \Big] \Pi_{q=1, q \ne j_0, j_0+1}^k |(e^{i\theta_{j_1}} - e^{i\hat{\theta}_q})| \\
<& 0 \quad \text{(for sufficiently small } \Delta \text{, by (A.3) and Lemma A.1)}.
\end{aligned}$$

For $j_1 \in \{p+1, \cdots, k+1\}$, we similarly have

$$\begin{aligned}
&|\eta_{j_1}(\hat{\theta}_1, \cdots, \hat{\theta}_{j_0} - \Delta, \hat{\theta}_{j_0+1} + \Delta, \hat{\theta}_{j_0+2}, \cdots, \hat{\theta}_k)| - |\eta_{j_1}(\hat{\theta}_1, \cdots, \hat{\theta}_k)| \\
=& \Big[|(e^{i\theta_{j_1}} - e^{i\hat{\theta}_{j_0} - i\Delta})(e^{i\theta_{j_1}} - e^{i\hat{\theta}_{j_0+1} + i\Delta})| - |(e^{i\theta_{j_1}} - e^{i\hat{\theta}_{j_0}})(e^{i\theta_{j_1}} - e^{i\hat{\theta}_{j_0+1}})| \Big] \Pi_{q=1, q \ne j_0, j_0+1}^k |(e^{i\theta_{j_1}} - e^{i\hat{\theta}_q})| \\
=& \Big[|(1 - e^{i\hat{\theta}_{j_1} - i\theta_{j_0} + i\Delta})(1 - e^{i\hat{\theta}_{j_1} - i\theta_{j_0+1} - i\Delta})| - |(1 - e^{i\hat{\theta}_{j_1} - i\theta_{j_0}})(1 - e^{i\hat{\theta}_{j_1} - i\theta_{j_0+1}})| \Big] \Pi_{q=1, q \ne j_0, j_0+1}^k |(e^{i\theta_{j_1}} - e^{i\hat{\theta}_q})| \\
<& 0 \quad \text{(for sufficiently small } \Delta \text{, by (A.3) and Lemma A.1)}.
\end{aligned}$$

Thus, choosing sufficiently small $\Delta > 0$, we can make

$$\max_{1\leq j\leq k+1} |\eta_j(\hat{\theta}_1,\cdots,\hat{\theta}_{j_0}-\Delta,\hat{\theta}_{j_0+1}+\Delta,\hat{\theta}_{j_0+2},\cdots,\hat{\theta}_k)| < \max_{1\leq j\leq k+1} |\eta_j(\hat{\theta}_1,\cdots,\hat{\theta}_k)|,$$

which contradicts the fact that $(\hat{\theta}_1,\cdots,\hat{\theta}_k)$ is a minimizer of (A.2). This means that Case 1 will not occur for $(\hat{\theta}_1,\cdots,\hat{\theta}_k)$.

Case 2: There exists j_0, p such that $\hat{\theta}_{j_0} = \theta_p, \theta_{j_0+1} \in (\theta_p, \theta_{p+1})$.
We still denote j_1 the number in $\{1,\cdots,k+1\}$ such that

$$\max_{j=1,\cdots,k+1} |\eta_j(\hat{\theta}_1,\cdots,\hat{\theta}_k)| = |\eta_{j_1}(\hat{\theta}_1,\cdots,\hat{d}_k)| > 0.$$

Since $\hat{\theta}_{j_0} = \theta_p$, $\eta_p(\hat{\theta}_1,\cdots,\hat{\theta}_k) = 0$, we only need to consider $j_1 \in \{1,\cdots,p-1\}$ and $j_1 \in \{p+1,\cdots,k+1\}$. For $j_1 = 1,\cdots,p-1$ and $j_1 = p+1,\cdots,k+1$, similarly to the analysis in Case 1, we have

$$|\eta_{j_1}(\hat{\theta}_1,\cdots,\hat{\theta}_{j_0}-\Delta,\hat{\theta}_{j_0+1}+\Delta,\hat{\theta}_{j_0+2},\cdots,\hat{\theta}_k)| < |\eta_{j_1}(\hat{\theta}_1,\cdots,\hat{\theta}_k)|,$$

for sufficiently small $\Delta > 0$. Thus, choosing sufficiently small $\Delta > 0$, we can make

$$\max_{1\leq j\leq k+1} |\eta_j(\hat{\theta}_1,\cdots,\hat{\theta}_{j_0}-\Delta,\hat{\theta}_{j_0+1}+\Delta,\hat{\theta}_{j_0+2},\cdots,\hat{\theta}_k)| < \max_{1\leq j\leq k+1} |\eta_j(\hat{\theta}_1,\cdots,\hat{\theta}_k)|,$$

which contradicts the fact that $(\hat{\theta}_1,\cdots,\hat{\theta}_k)$ is a minimizer of (A.2). This means that Case 2 will not occur for $(\hat{\theta}_1,\cdots,\hat{\theta}_k)$.

Case 3: There exists j_0, p such that $\hat{\theta}_{j_0} = \hat{\theta}_{j_0+1} = \theta_p$.
Denote j_1 the number in $\{1,\cdots,k+1\}$ such that

$$\max_{j=1,\cdots,k+1} |\eta_j(\hat{\theta}_1,\cdots,\hat{\theta}_k)| = |\eta_{j_1}(\hat{\theta}_1,\cdots,\hat{d}_k)| > 0. \tag{A.4}$$

Since $\eta_p(\hat{\theta}_1,\cdots,\hat{\theta}_k) = 0$, we only consider $j_1 \in \{1,\cdots,p-1\}$ and $j_1 \in \{p+1,\cdots,k+1\}$. Let $\Delta > 0$, for $j_1 \in \{1,\cdots,p-1\}$, we have

$$\begin{aligned}
&|\eta_{j_1}(\hat{\theta}_1,\cdots,\theta(\hat{d}_{j_0-1}),\hat{\theta}_{j_0}-\Delta,\hat{\theta}_{j_0+1}+\Delta,\cdots,\hat{\theta}_k)| - |\eta_{j_1}(\hat{\theta}_1,\cdots,\hat{\theta}_k)|\\
=&\Big[|(e^{i\theta_{j_1}}-e^{i\hat{\theta}_{j_0}-i\Delta})(e^{i\theta_{j_1}}-e^{i\hat{\theta}_{j_0+1}+i\Delta})| - |(e^{i\theta_{j_1}}-e^{i\hat{\theta}_{j_0}})(e^{i\theta_{j_1}}-e^{i\hat{\theta}_{j_0+1}})|\Big]\Pi_{q=1,q\neq j_0,j_0+1}^k |(e^{i\theta_{j_1}}-e^{i\hat{\theta}_q})|\\
=&\Big[|(e^{i\theta_{j_1}}-e^{i\theta_p-i\Delta})(e^{i\theta_{j_1}}-e^{i\theta_p+i\Delta})| - |(e^{i\theta_{j_1}}-e^{i\theta_p})(e^{i\theta_{j_1}}-e^{i\theta_p})|\Big]\Pi_{q=1,q\neq j_0,j_0+1}^k |(e^{i\theta_{j_1}}-e^{i\hat{\theta}_q})|\\
=&\Big[|(1-e^{i\theta_p-i\theta_{j_1}-i\Delta})(1-e^{i\theta_p-i\theta_{j_1}+i\Delta})| - |(1-e^{i\theta_p-\theta_{j_1}})(1-e^{i\theta_p-i\theta_{j_1}})|\Big]\Pi_{q=1,q\neq j_0,j_0+1}^k |(e^{i\theta_{j_1}}-e^{i\hat{\theta}_q})|\\
<&0 \quad \text{(for sufficiently small } \Delta\text{, by (A.4) and Lemma A.1).}
\end{aligned}$$

For $j_1 \in \{p+1,\cdots,k\}$, we have the same result in exactly the same way. Thus, choosing sufficiently small $\Delta > 0$, we can make

$$\max_{1\leq j\leq k+1} |\eta_j(\hat{\theta}_1,\cdots,\hat{\theta}_{j_0}-\Delta,\hat{\theta}_{j_0+1}+\Delta,\hat{\theta}_{j_0+2},\cdots,\hat{\theta}_k)| < \max_{1\leq j\leq k+1} |\eta_j(\hat{\theta}_1,\cdots,\hat{\theta}_k)|,$$

which contradicts the fact that $(\hat{\theta}_1,\cdots,\hat{\theta}_k)$ is a minimizer of (A.2). This means that Case 3 will not occur for $(\hat{\theta}_1,\cdots,\hat{\theta}_k)$. Combining these results, we proved the claim.

Step 2. By the claim in Step 1, the minimizer $(\hat{\theta}_1, \cdots, \hat{\theta}_k)$ and $(\theta_1, \cdots, d_{k+1})$ satisfy the interlacing relation

$$0 \le \theta_1 \le \hat{\theta}_1 < \theta_2 \le \hat{\theta}_2 < \cdots < \theta_k \le \hat{\theta}_k \le \theta_{k+1} < 2\pi. \tag{A.5}$$

Note that because $\min_{j\neq q} |\theta_j - \theta_q|_{2\pi} = \theta_{\min}$, if $|\theta_j - \hat{\theta}_j|_{2\pi} \le \frac{\theta_{\min}}{2}$, then $|\theta_{j+1} - \hat{\theta}_j|_{2\pi} > \frac{\theta_{\min}}{2}$. It is clear that by the interlacing relation (A.5), there must be some $1 \le j \le k$ such that $|\hat{\theta}_j - \theta_j|_{2\pi} > \frac{\theta_{\min}}{2}$ (Case 1) or $|\theta_{k+1} - \hat{\theta}_k|_{2\pi} > \frac{\theta_{\min}}{2}$ (Case 2). In what follows, we only prove this statement in Case 1. Case 2 can be handled in the same manner. Let $\hat{\theta}_{j_0}$ be the first point (starting from $\hat{\theta}_1, \hat{\theta}_2, \cdots$) such that

$$|\hat{\theta}_{j_0} - \theta_{j_0}|_{2\pi} > \frac{\theta_{\min}}{2}. \tag{A.6}$$

Without loss of generality, we suppose that $\theta_{j_0} > \pi$. Then, we decompose the rest of the θ_j's into two sets: $set_1 = \{\theta_{j_{-1}}, \cdots, \theta_{j_{-p}}\}$ contains all the θ_j's in $[\theta_{j_0} - \pi, \theta_{j_0})$ and $set_2 = \{\theta_{j_1}, \cdots, \theta_{j_{k-1-p}}\}$ contains the rest of the θ_j's except θ_1. Note that by this decomposition, $\{e^{i\theta_{j_{-1}}}, \cdots, e^{i\theta_{j_{-p}}}\}$ contains points in the same half circle of the unit circle and $\{e^{i\theta_{j_1}}, \cdots, e^{i\theta_{j_{k-1-p}}}\}$ contains points in the other half circle. We remark that we just ignore θ_1 when considering set_1, set_2 under the setting that $\theta_{j_0} > \pi$. For the other cases, the results can be proved in the same manner. The points in the two sets are arranged in such a way that, for $q = 1, 2, \cdots$,

$$|e^{i\theta_{j_{-(q+1)}}} - e^{i\theta_{j_0}}| > |e^{i\theta_{j_{-q}}} - e^{i\theta_{j_0}}|, \quad |e^{i\theta_{j_{q+1}}} - e^{i\theta_{j_0}}| > |e^{i\theta_{j_q}} - e^{i\theta_{j_0}}|. \tag{A.7}$$

We also denote the corresponding point $\hat{\theta}_j$ in $[\theta_{j_{-(q+1)}}, \theta_{j_{-q}})$, for $q = 1, 2, \cdots$, as $\hat{\theta}_{j_{-(q+1)}}$ (respectively, $\hat{\theta}_j$ in $[\theta_{j_q}, \theta_{j_{q+1}})$, for $q = 1, 2, \cdots$, as $\hat{\theta}_{j_q}$). Now, the following estimate holds:

$$\begin{aligned}|\Pi_{q=1}^k (e^{i\theta_{j_0}} - e^{i\hat{\theta}_q})| &= |e^{i\theta_{j_0}} - e^{i\hat{\theta}_{j_0}}||\Pi_{q=1}^p (e^{i\theta_{j_0}} - e^{i\hat{\theta}_{j_{-q}}})||\Pi_{s=1}^{k-1-p}(e^{i\theta_{j_0}} - e^{i\hat{\theta}_{j_s}})| \\ &> \frac{\theta_{\min}}{\pi}|\Pi_{q=1}^p (e^{i\theta_{j_0}} - e^{i\hat{\theta}_{j_{-q}}})||\Pi_{s=1}^{k-1-p}(e^{i\theta_{j_0}} - e^{i\hat{\theta}_{j_s}})|.\end{aligned}$$

Here, we have obtained the last inequality by (A.6) and used that, for $x, y \in \mathbb{R}$,

$$|e^{ix} - e^{iy}| \ge \frac{2}{\pi}|x - y|_{2\pi}. \tag{A.8}$$

Moreover, since by assumption $\hat{\theta}_{j_0}$ is the first point satisfying (A.6), we get $|\theta_{j_{-1}} - \hat{\theta}_{j_{-1}}|_{2\pi} \le \frac{\theta_{\min}}{2}$ and thus, $|\theta_{j_0} - \hat{\theta}_{j_{-1}}|_{2\pi} > \frac{\theta_{\min}}{2}$. Hence, by (A.6) and (A.8), we have

$$|\Pi_{q=1}^k (e^{i\theta_{j_0}} - e^{i\hat{\theta}_q})| > (\frac{\theta_{\min}}{\pi})^2 |\Pi_{q=2}^p (e^{i\theta_{j_0}} - e^{i\hat{\theta}_{j_{-q}}})||\Pi_{s=1}^{k-1-p}(e^{i\theta_{j_0}} - e^{i\hat{\theta}_{j_s}})|.$$

On the other hand, by the interlacing relation (A.5), for $q \ge 2$ we have

$$|\theta_{j_0} - \hat{\theta}_{-q}|_{2\pi} > |\theta_{j_0} - \theta_{-q+1}|_{2\pi}$$

and for $s \ge 1$, $|\theta_{j_0} - \hat{\theta}_s|_{2\pi} > |\theta_{j_0} - \theta_s|_{2\pi}$. Thus, we obtain the following estimate:

$$|\Pi_{q=1}^k (e^{i\theta_{j_0}} - e^{i\hat{\theta}_q})| > (\frac{\theta_{\min}}{\pi})^2 |\Pi_{q=2}^p (e^{i\theta_{j_0}} - e^{i\theta_{j_{-q+1}}})||\Pi_{s=1}^{k-1-p}(e^{i\theta_{j_0}} - e^{i\theta_{j_s}})|.$$

Note that by $\min_{j\neq q}|\theta_j-\theta_q|_{2\pi}=\theta_{\min}$,

$$|e^{i\theta_{j_0}}-e^{i\theta_{j-q}}|>\frac{2q}{\pi}\theta_{\min}, q=1,2,\cdots \quad \text{and} \quad |e^{i\theta_{j_0}}-e^{i\theta_{j_s}}|>\frac{2s}{\pi}\theta_{\min}, s=1,2,\cdots.$$

Therefore, we obtain that

$$|\Pi_{q=1}^{k}(e^{i\theta_{j_0}}-e^{i\hat{\theta}_q})|>(p-1)!(k-1-p)!(\frac{2\theta_{\min}}{\pi})^{k-2}(\frac{\theta_{\min}}{\pi})^2.$$

Minimizing $(p-1)!(k-1-p)!$ over $p=1,\cdots,k-1$ gives

$$\min_{p=1,\cdots,k}(p-1)!(k-1-p)!\geq\begin{cases}(\frac{k-1}{2})!(\frac{k-3}{2})! & k \text{ is odd},\\ (\frac{k-2}{2}!)^2 & k \text{ is even}.\end{cases} \tag{A.9}$$

Thus, for $k\geq 2$, $||\eta_{k+1,k}||_\infty\geq\xi(k)(\frac{2\theta_{\min}}{\pi})^k$. □

A.3. Proof of Lemma 4.2

The proof of Lemma 4.2 is divided into three steps.

Proof. We only prove the lemma for $k\geq 3$. The case $k=2$ can be deduced in a similar manner. Moreover, we only need to consider $\theta_j,\hat{\theta}_j\in[0,2\pi), j=1,\cdots,k$.

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

$$\begin{aligned}&\eta_{k,k}(e^{i\theta_1},\cdots,e^{i\theta_k},e^{i\hat{\theta}_1},\cdots,e^{i\hat{\theta}_k})\\ =&\text{diag}\Big(|e^{i\theta_1}-e^{i\hat{\theta}_{p_0}}|,\cdots,|e^{i\theta_k}-e^{i\hat{\theta}_{p_0}}|\Big)\eta_{k,k-1}(e^{i\theta_1},\cdots,e^{i\theta_k},e^{i\hat{\theta}_1},\cdots,e^{i\hat{\theta}_{p_0-1}},\\ &e^{i\hat{\theta}_{p_0+1}},\cdots,e^{i\hat{\theta}_k}).\end{aligned}$$

Combining Lemma 4.1 and (A.8), we obtain that

$$||\eta_{k,k}||_\infty\geq\frac{\theta_{\min}}{\pi}||\eta_{k,k-1}||_\infty\geq\frac{\xi(k-1)}{2}(\frac{2\theta_{\min}}{\pi})^k.$$

By the formula of $\xi(k)$ in (4.3), we can verify directly that $\frac{\xi(k-1)}{2}\geq\frac{\xi(k-2)}{4}$. Therefore,

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

where we have used (4.6) in the inequality above. This is in contradiction with (4.5) and hence the claim is proved.

Step 2. We claim that for each $\theta_j, 1\leq j\leq k$, there exists one and only one $\hat{\theta}_p$ such that $|\theta_j-\hat{\theta}_p|_{2\pi}<\frac{\theta_{\min}}{2}$. It suffices to show that for each $\theta_j, 1\leq j\leq k$, there is only one $\hat{\theta}_p$ such that $|\theta_j-\hat{\theta}_p|_{2\pi}<\frac{\theta_{\min}}{2}$. By contradiction, suppose there exist p_1,p_2, and j_0 such that

$$|\theta_{j_0}-\hat{\theta}_{p_1}|_{2\pi}<\frac{\theta_{\min}}{2}, |\theta_{j_0}-\hat{\theta}_{p_2}|_{2\pi}<\frac{\theta_{\min}}{2}.$$

Then, for all $j \neq j_0$, we have

$$|\theta_j - \hat{\theta}_{p_1}|_{2\pi} |\theta_j - \hat{\theta}_{p_2}|_{2\pi} \geq \frac{\theta_{\min}^2}{4}, \text{ and } |(e^{i\theta_j} - e^{i\hat{\theta}_{p_1}})(e^{i\theta_j} - e^{i\hat{\theta}_{p_2}})| \geq \frac{\theta_{\min}^2}{\pi^2}. \tag{A.10}$$

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

$$\begin{aligned}
&\eta_{k,k}(e^{i\theta_1}, \cdots, e^{i\theta_k}, e^{i\hat{\theta}_1}, \cdots, e^{i\hat{\theta}_k}) \\
=&\text{diag}\left(|e^{i\theta_1} - e^{i\hat{\theta}_{p_1}}||e^{i\theta_1} - e^{i\hat{\theta}_{p_2}}|, \cdots, |e^{i\theta_k} - e^{i\hat{\theta}_{p_1}}||e^{i\theta_k} - e^{i\hat{\theta}_{p_2}}|\right)\eta_{k-1,k-2},
\end{aligned}$$

where

$$\begin{aligned}
\eta_{k-1,k-2} =& \eta_{k-1,k-2}(e^{i\theta_1}, \cdots, e^{i\theta_{j_0-1}}, e^{i\theta_{j_0+1}}, \cdots, e^{i\theta_k}, e^{i\hat{\theta}_1}, \cdots, e^{i\hat{\theta}_{p_1-1}}, \\
& e^{i\hat{\theta}_{p_1+1}}, \cdots, e^{i\hat{\theta}_{p_2-1}}, e^{i\hat{\theta}_{p_2+1}}, \cdots, e^{i\hat{\theta}_k}).
\end{aligned}$$

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

$$||\eta_{k,k}||_\infty \geq \frac{\theta_{\min}^2}{\pi^2}||\eta_{k-1,k-2}||_\infty.$$

Using Lemma 4.1 and (4.6), we further get

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

which contradicts (4.5). This contradiction proves our claim.

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

$$|\hat{\theta}_j - \theta_j|_{2\pi} < \frac{\theta_{\min}}{2}, \quad j = 1, \cdots, k. \tag{A.11}$$

We now prove (4.9). For each $j_0 \in \{1, \cdots, k\}, \theta_{j_0} > \pi$, we decompose the rest of the θ_j's into two sets: $set_1 = \{\theta_{j_{-1}}, \cdots, \theta_{j_{-p}}\}$ contains all the θ_j's in $[\theta_{j_0} - \pi, \theta_{j_0})$ and $set_2 = \{\theta_{j_1}, \cdots, \theta_{j_{k-1-p}}\}$ contains the rest of the θ_j's. For $\theta_{j_0} \leq \pi$, we can prove (4.9) in the same manner. The points in the two sets are arranged to satisfy that for $q = 1, 2, \cdots$,

$$|e^{i\theta_{j_{-(q+1)}}} - e^{i\theta_{j_0}}| > |e^{i\theta_{j_{-q}}} - e^{i\theta_{j_0}}|, \quad |e^{i\theta_{j_{q+1}}} - e^{i\theta_{j_0}}| > |e^{i\theta_{j_q}} - e^{i\theta_{j_0}}|.$$

By (A.11) and (A.8), it is clear that

$$|e^{i\theta_{j_0}} - e^{i\hat{\theta}_{j_q}}| \geq \begin{cases} \frac{(-q-\frac{1}{2})2\theta_{\min}}{\pi}, & q \leq -1, \\ \frac{(q-\frac{1}{2})2\theta_{\min}}{\pi}, & q \geq 1. \end{cases} \tag{A.12}$$

Thus,

$$
\begin{aligned}
&|(e^{i\theta_{j_0}}-e^{i\hat{\theta}_{j_0}})(e^{i\theta_{j_0}}-e^{i\hat{\theta}_{j-1}})\cdots(e^{i\theta_{j_0}}-e^{i\hat{\theta}_{j-p}})(e^{i\theta_{j_0}}-e^{i\hat{\theta}_{j_1}}\cdots(e^{i\theta_{j_0}}-e^{i\hat{\theta}_{j_{k-1-p}}}))|\\
\geq&|e^{i\theta_{j_0}}-e^{i\hat{\theta}_{j_0}}|\Big(\Pi_{1\leq q\leq p}\frac{2q-1}{2}\frac{2\theta_{\min}}{\pi}\Big)\Big(\Pi_{1\leq q\leq k-1-p}\frac{2q-1}{2}\frac{2\theta_{\min}}{\pi}\Big)\quad\Big(\text{by (A.12)}\Big)\\
=&|e^{i\theta_{j_0}}-e^{i\hat{\theta}_{j_0}}|(\frac{\theta_{\min}}{\pi})^{k-1}(2p-1)!!(2(k-1-p)-1)!!\\
\geq&|e^{i\theta_{j_0}}-e^{i\hat{\theta}_{j_0}}|(\frac{\theta_{\min}}{\pi})^{k-1}(k-2)!\qquad\Big(\text{by } (2p-1)!!(2(k-1-p)-1)!!\geq(k-2)!\Big).
\end{aligned}
$$

This together with (4.5) yields

$$
|e^{i\hat{\theta}_j}-e^{i\theta_j}|(\frac{\theta_{\min}}{2})^{k-1}(k-2)!<\frac{2}{\pi}\epsilon,\quad j=1,2,\cdots,k.
$$

Thus (4.9) follows and the proof of the lemma is complete. □

B. PROOF OF PROPOSITION 2.1

We first introduce an auxiliary lemma that was derived in [24].

Lemma B.1. *Let $t_1,\cdots,t_k$ be k different real numbers and let t be a real number. We have*

$$
\big(D_k(k-1)^{-1}\phi_{k-1}(t)\big)_j=\Pi_{1\leq q\leq k,q\neq j}\frac{t-t_q}{t_j-t_q},
$$

where $D_k(k-1):=\big(\phi_{k-1}(t_1),\cdots,\phi_{k-1}(t_k)\big)$ and $\phi_{k-1}(\cdot)$ is defined by (4.1).

The proof of Proposition 2.1 is divided into two steps.

Proof. **Step 1.** Let τ be the one in (2.14), $y_1=-\tau,y_2=-2\tau,\cdots,y_n=-n\tau$ and $\hat{y}_1=0,\hat{y}_2=\tau,\cdots,\hat{y}_n=(n-1)\tau$. For any measure $\mu=\sum_{j=1}^n a_j\delta_{y_j}$, the corresponding illuminated measure is

$$
I_t\mu=\sum_{j=1}^n I_t(y_j)a_j\delta_{y_j},
$$

where I_t is the t-th illumination pattern. Denoting $\hat{I}_t\rho=\rho_t=\sum_{j=1}^n\hat{a}_{j,t}\delta_{\hat{y}_j}$, by the definition of $||\cdot||_2$ we have

$$
||\mathscr{F}[\hat{I}_t\rho]-\mathscr{F}[I_t\mu]||_2\leq\max_{x\in[-1,1]}|\mathscr{F}[\hat{I}_t\rho-I_t\mu](\Omega x)|,\quad x\in[-1,1]. \tag{B.1}
$$

Thus we next estimate $|\mathscr{F}[\hat{I}_t\rho-I_t\mu](\Omega x)|$ to prove the proposition. By Taylor expansion,

$$
\mathscr{F}[\hat{I}_t\rho-I_t\mu](\Omega x)=\sum_{k=0}^{\infty}\big(Q_k(\hat{I}_t\rho)-Q_k(I_t\mu)\big)\frac{(ix)^k}{k!}, \tag{B.2}
$$

where $Q_k(I_t\mu)=\sum_{j=1}^n I_t(y_j)a_j(\Omega y_j)^k$, $Q_k(\hat I_t\rho)=\sum_{j=1}^n \hat a_{j,t}(\Omega\hat y_j)^k$. Note that $Q_k(\hat I_t\rho)=Q_k(I_t\mu), k=1,\cdots,n-1$, is a linear system that can be rewritten as

$$\big(\phi_{n-1}(\Omega\hat y_1),\cdots,\phi_{n-1}(\Omega\hat y_n)\big)\hat\alpha_t=\big(\phi_{n-1}(\Omega y_1),\cdots,\phi_{n-1}(\Omega y_n)\big)\alpha_t, \tag{B.3}$$

where $\hat\alpha_t=(\hat a_{1,t},\cdots,\hat a_{n,t})^T$ and $\alpha_t=(I_t(y_1)a_1,\cdots,I_t(y_n)a_n)^T$. Since $\phi_{n-1}(\Omega\hat y_j), 1\le j\le n$, are linearly independent, for each t, we can find some $\hat a_{j,t}$'s (i.e., some $\hat I_t$'s) so that $Q_k(\hat I_t\rho)=Q_k(I_t\mu), k=0,\cdots,n-1$. We next estimate $Q_k(\hat I_t\rho)$ and $Q_k(I_t\mu)$.

Step 2. Under the scenario of Step 1, we obtain

$$\hat\alpha_t=\big(\phi_{n-1}(\Omega\hat y_1),\cdots,\phi_{n-1}(\Omega\hat y_n)\big)^{-1}\big(\phi_{n-1}(\Omega y_1),\cdots,\phi_{n-1}(\Omega y_n)\big)\alpha_t,\ 1\le t\le T.$$

By Lemma B.1, we arrive at

$$\begin{aligned}
\max_{p=1,\cdots,n}\left|\hat a_{p,t}\right| &\le \max_{p=1,\cdots,n}\sum_{j=1}^n\left|\Pi_{q=1,q\neq p}^n\frac{y_j-\hat y_q}{\hat y_p-\hat y_q}I_t(y_j)a_j\right|\le \max_{p,j=1,\cdots,n}\left|\Pi_{q=1,q\neq p}^n\frac{y_j-\hat y_q}{\hat y_p-\hat y_q}\right|nm_{\min}\\
&\le\frac{(n+1)(n+2)\cdots(2n-1)}{\zeta(n)}nm_{\min}=\frac{(2n-1)!}{n!\zeta(n)}nm_{\min}\quad\big(\zeta(\cdot)\text{ is defined in (4.3)}\big)\\
&\le\frac{e2^{3n-\frac12}}{\pi^{\frac32}(n-1)}nm_{\min}\quad(\text{using inequality (C.1)}).
\end{aligned}$$

We have $\sum_{j=1}^n|\hat a_{j,t}|\le\frac{e2^{3n-\frac12}}{\pi^{\frac32}(n-1)}n^2m_{\min}, 1\le t\le T$. Therefore, we have

$$\begin{aligned}
\left|\max_{x\in[-1,1]}\mathscr F[\hat I_t\rho-I_t\mu](\Omega x)\right| &\le\sum_{k\ge n}\frac{|Q_k(I_t\mu)|+|Q_k(\hat I_t\rho)|}{k!}\le\sum_{k\ge n}(\sum_{j=1}^n|I_t(y_j)a_j|+|\hat a_{j,t}|)\frac{(n\Omega\tau)^k}{k!}\\
&\le\Big(\frac{e2^{3n-\frac12}}{\pi^{\frac32}(n-1)}n^2+n\Big)m_{\min}\frac{(n\Omega\tau)^n}{n!}\sum_{k\ge n}\frac{(n\Omega\tau)^{k-n}n!}{k!}\\
&<1.06\Big(\frac{e2^{3n-\frac12}}{\pi^{\frac32}(n-1)}n^2+n\Big)m_{\min}\frac{(n\Omega\tau)^n}{n!}\quad\big(\text{by (2.14) we have }\Omega\tau<0.05\big)\\
&<1.06\Big(\frac{e2^{3n-\frac12}}{\pi^{\frac32}(n-1)}n^2+n\Big)m_{\min}\frac{e^n}{\sqrt{2\pi n}}(\Omega\tau)^n\quad\big(\text{by (C.1)}\big)\\
&<\sigma\quad(\text{by (2.14)}).
\end{aligned}$$

Together with (B.1), this completes the proof. □

C. Some estimations

In this section, we present some estimations that are used in this paper. We first recall the following Stirling approximation of factorial

$$\sqrt{2\pi}n^{n+\frac12}e^{-n}\le n!\le en^{n+\frac12}e^{-n}. \tag{C.1}$$

Then, we state the following results.

Lemma C.1. *Let $\xi(n-1)$ be defined as in (4.3). For $n \geq 2$, we have*

$$\Big(\frac{2\sqrt{n}n^{n-1}}{\xi(n-1)}\Big)^{\frac{1}{n-1}} < 4.4e.$$

Proof. For $n = 2,3,4,5$, it is easy to check that the above inequality holds. Using (C.1), we have for odd $n \geq 7$,

$$\begin{aligned}\Big(\frac{2\sqrt{n}n^{n-1}}{\xi(n-1)}\Big)^{\frac{1}{n-1}} &= \Big(\frac{2\sqrt{n}n^{n-1}}{\frac{(\frac{n-3}{2}!)^2}{4}}\Big)^{\frac{1}{n-1}} \leq \Big(\frac{8\sqrt{n}n^{n-1}}{\pi(\frac{n-3}{2})^{n-2}e^{-(n-3)}}\Big)^{\frac{1}{n-1}} \\ &= \Big(\frac{8\sqrt{n}}{\pi(\frac{n-3}{2})^{-1}e^2}\Big)^{\frac{1}{n-1}}\frac{2en}{n-3} \\ &< 4.4e,\end{aligned}$$

and for even $n \geq 6$,

$$\begin{aligned}\Big(\frac{2\sqrt{n}n^{n-1}}{\xi(n-1)}\Big)^{\frac{1}{n-1}} &= \Big(\frac{2\sqrt{n}n^{n-1}}{\frac{(\frac{n-2}{2})!(\frac{n-4}{2})!}{4}}\Big)^{\frac{1}{n-1}} \leq \Big(\frac{8\sqrt{n}n^{n-1}}{\pi(\frac{n-2}{2})^{\frac{n-1}{2}}(\frac{n-4}{2})^{\frac{n-3}{2}}e^{-(n-3)}}\Big)^{\frac{1}{n-1}} \\ &= \Big(\frac{8\sqrt{n}}{\pi(\frac{n-4}{2})^{-1}e^2}\Big)^{\frac{1}{n-1}}\frac{2en}{\sqrt{n-2}\sqrt{n-4}} \\ &< 4.4e.\end{aligned}$$

This completes the proof. □

Lemma C.2. *Let $\lambda(n)$ be defined as in (4.7). For $n \geq 2$, we have*

$$\Big(\frac{8\sqrt{n}n^n}{\lambda(n)}\Big)^{\frac{1}{n}} < 4.4e.$$

Proof. For $n = 2,3,\cdots,14$, it is easy to check that the above inequality holds. Using (C.1), we have for even $n \geq 16$,

$$\begin{aligned}\Big(\frac{8\sqrt{n}n^n}{\xi(n-2)}\Big)^{\frac{1}{n}} &= \Big(\frac{8\sqrt{n}n^n}{(\frac{n-4}{2}!)^2/4}\Big)^{\frac{1}{n}} \leq \Big(\frac{32\sqrt{n}n^n}{\pi(\frac{n-4}{2})^{n-3}e^{-(n-4)}}\Big)^{\frac{1}{n}} \\ &= \Big(\frac{32\sqrt{n}}{\pi(\frac{n-4}{2})^{-3}e^4}\Big)^{\frac{1}{n}}\frac{2en}{n-4} \\ &< 4.4e,\end{aligned}$$

and for odd $n \geq 15$,

$$\begin{aligned}\Big(\frac{8\sqrt{n}n^n}{\xi(n-2)}\Big)^{\frac{1}{n}} &= \Big(\frac{8\sqrt{n}n^n}{\frac{(\frac{n-3}{2})!(\frac{n-5}{2})!}{4}}\Big)^{\frac{1}{n}} \leq \Big(\frac{32\sqrt{n}n^n}{\pi(\frac{n-3}{2})^{\frac{n-2}{2}}(\frac{n-5}{2})^{\frac{n-4}{2}}e^{-(n-4)}}\Big)^{\frac{1}{n}} \\ &= \Big(\frac{32\sqrt{n}}{\pi(\frac{n-3}{2})^{-1}(\frac{n-5}{2})^{-2}e^4}\Big)^{\frac{1}{n}}\frac{2en}{\sqrt{n-3}\sqrt{n-5}} \\ &< 4.4e.\end{aligned}$$

This completes the proof. □

Lemma C.3. *Let* $0 \le s, s_1, s_2 \le 1$,

$$A = \begin{pmatrix} 1 & s \\ s & 1 \end{pmatrix},$$

then $\sigma_{\infty,\min}(A) = 1 - s$.

Proof. By definition of $\sigma_{\infty,\min}(A)$, it follows that

$$\begin{aligned}
\sigma_{\infty,\min}(A) &= \min_{x,y\in\mathbb{C},\max(|x|,|y|)\ge 1} \max(|x+sy|,|sx+y|) \\
&= \min_{|y|\le 1, y\in\mathbb{C}} \max(|1+sy|,|s+y|) \\
&= \min_{0\le r\le 1,\theta\in[0,2\pi)} \sqrt{\max((1+sr\cos\theta)^2 + s^2r^2\sin^2\theta, (s+r\cos\theta)^2 + r^2\sin^2\theta)} \\
&= \min_{0\le r\le 1,\theta\in[0,2\pi)} \sqrt{\max(1+s^2r^2+2sr\cos\theta, s^2+r^2+2sr\cos\theta)} \\
&= \min_{0\le r\le 1} \sqrt{\max(1+s^2r^2-2sr, s^2+r^2-2sr)} \\
&= \min_{0\le r\le 1} \max(1-sr, |s-r|) \\
&= 1-s.
\end{aligned}$$

□

References

[1] 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)*, pages 617–621. IEEE, 2015.

[2] Jean-Marc Azais, Yohann De Castro, and Fabrice Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.

[3] 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.

[4] Dmitry Batenkov, Gil Goldman, and Yosef Yomdin. Super-resolution of near-colliding point sources. *Information and Inference: A Journal of the IMA*, 05 2020. iaaa005.

[5] E. Betzig, G.H. Patterson, R. Sougrat, O.W. Lindwasser, S. Olenych, J.S. Bonifacino, M.W. Davidson, J. Lippincott-Schwartz, and H.F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313:1642–1645, 2006.

[6] E. Betzig, G.H. Patterson, R. Sougrat, O.W. Lindwasser, S. Olenych, J.S. Bonifacino, M.W. Davidson, J. Lippincott-Schwartz, and H.F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313:1642–1645, 2006.

[7] Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[8] Sitan Chen and Ankur Moitra. Algorithmic foundations for the diffraction limit. pages 490–503, 2021.

[9] 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.

[10] Laurent Demanet and Nam Nguyen. The recoverability limit for superresolution via sparsity. *arXiv preprint arXiv:1502.01385*, 2015.

[11] Arnold Jan Den Dekker and A Van den Bos. Resolution: a survey. *JOSA A*, 14(3):547–557, 1997.

[12] 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.

[13] David L. Donoho. Superresolution via sparsity constraints. *SIAM journal on mathematical analysis*, 23(5):1309–1331, 1992.

[14] Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.

[15] M.G. Gustafsson. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.*, 198:82–87, 2000.

[16] R. Heintzmann. Saturated patterned excitation microscopy with two-dimensional excitation patterns. *Micron.*, 34:283–291, 2003.

[17] S.W. Hell, S. Jakobs, and L. Kastrup. Imaging and writing at the nanoscale with focused visible light through saturable optical transitions. *Appl. Phys. A.*, 77:859–860, 2003.

[18] S.W. Hell and M. Kroug. Ground-state-depletion fluorscence microscopy: A concept for breaking the diffraction resolution limit. *Appl. Phys. B.*, 60:495–497, 1995.

[19] S.W. Hell, S.J. Sahl, M. Bates, X. Zhuang, R. Heintzmann, M.J. Booth, J. Bewersdorf, G. Shtengel, H. Hess, P. Tinnefeld, A. Honigmann, S. Jakobs, I. Testa, L. Cognet, B. Lounis, H. Ewers, S.J. Davis, C. Eggeling, D. Klenerman, K.I. Willig, G. Vicidomini, M. Castello, A. Diaspro, and T. Cordes. The 2015 super-resolution microscopy roadmap. *J. Phys. D. Appl. Phys.*, 48:443001, 2015.

[20] S.W. Hell and J. Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.*, 19:780–782, 1994.

[21] S.T. Hess, T.P.K. Girirajan, and M.D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.*, 91:4258–4272, 2006.

[22] 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.

[23] Ping Liu and Hai Zhang. A mathematical theory of computational resolution limit in multi-dimensional spaces. *Inverse Problems*, 37(10):104001, 2021.

[24] 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.

[25] Ping Liu and Hai Zhang. A mathematical theory of computational resolution limit in one dimension. *Applied and Computational Harmonic Analysis*, 56:402–446, 2022.

[26] T. Mangeat and et al. Super-resolved live-cell imaging using random illumination microscopy. *Cell Reports Methods*, 1(1):100009, 2021.

[27] J. Min, J. Jang, D. Keum, S.-W. Ryu, K.-H. Choi, C.and Jeong, and J.C. Ye. Fluorescent microscopy beyond diffraction limits using speckle illumination and joint support recovery. *Sci Rep.*, 3:2075, 2013.

[28] 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, pages 821–830, 2015.

[29] Veniamin I Morgenshtern. Super-resolution of positive sources on an arbitrarily fine grid. *J. Fourier Anal. Appl.*, 28(1):Paper No. 4., 2021.

[30] Veniamin I. Morgenshtern and Emmanuel J. Candes. Super-resolution of positive sources: The discrete setup. *SIAM Journal on Imaging Sciences*, 9(1):412–444, 2016.

[31] E. Mudry, K. Belkebir, J. Girard, J. Savatier, E. Le Moal, C. Nicoletti, M. Allain, and A. Sentenac. Structured illumination microscopy using unknown speckle patterns. *Nature Photonics*, 6:312–315, 2012.

[32] Clarice. Poon and Gabriel. Peyré. Multidimensional sparse super-resolution. *SIAM Journal on Mathematical Analysis*, 51(1):1–44, 2019.

[33] M.J. Rust, M. Bates, and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nat. Methods*, 3:793–796, 2006.

[34] Ola Sabet, Sanghyeon Yu, Riccardo Murri, Max Hess, Habib Ammari, and Lucas Pelkmans. Compressive super-resolution imaging at scale by brownian excitation amplitude modification. *submitted*, 2022.

[35] L. Schermelleh, P.M. Carlton, S. Haase, L. Shao, L. Winoto, P. Kner, B. Burke, M. C. Cardoso, D.A. Agard, M.G.L. Gustafsson, H. Leonhardt, and J.W. Sedat. Subdiffraction multicolor imaging of the nuclear periphery with 3d structured illumination microscopy. *Science*, 320:1332–1336, 2008.

[36] Morteza Shahram and Peyman Milanfar. Imaging below the diffraction limit: a statistical analysis. *IEEE Transactions on image processing*, 13(5):677–689, 2004.

[37] Morteza Shahram and Peyman Milanfar. Statistical analysis of achievable resolution in incoherent imaging. In *Signal and Data Processing of Small Targets 2003*, volume 5204, pages 1–9. International Society for Optics and Photonics, 2004.

[38] 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.

[39] O. Solomon, M. Mutzafi, M. Segev, and Y.C. Eldar. Sparsity-based super-resolution microscopy from correlation information. *Opt. Express*, 26:18238–18269, 2018.

[40] Gongguo Tang. Resolution limits for atomic decompositions via markov-bernstein type inequalities. In *2015 International Conference on Sampling Theory and Applications (SampTA)*, pages 548–552. IEEE, 2015.

[41] Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht. Near minimax line spectral estimation. *IEEE Transactions on Information Theory*, 61(1):499–512, 2014.

[42] 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.

[43] V. Westphalsilvio, O. Rizzolimarcel, A. Lauterbachdirk, J. Kaminereinhard, and S.W. Hell. Video-rate far-field optical nanoscopy dissects synaptic vesicle movementvideo-rate far-field optical nanoscopy dissects synaptic vesicle movement. *Science*, 320:246–249, 2008.

[44] L. Zhu, W. Zhang, D. Elnatan, and B. Huang. Faster storm using compressed sensing. *Nature methods*, 9(7):721–723, 2012.