Appl. Comput. Harmon. Anal. 51 (2021) 118–156

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Stable super-resolution limit and smallest singular value of restricted Fourier matrices

Weilin Li [a,*], Wenjing Liao [b]

[a] *Courant Institute of Mathematical Sciences, New York University, United States of America*
[b] *School of Mathematics, Georgia Institute of Technology, United States of America*

ARTICLE INFO

Article history:
Received 30 October 2018
Received in revised form 15 October 2020
Accepted 15 October 2020
Available online 21 October 2020
Communicated by Charles K. Chui

MSC:
42A10
42A15
94A08
94A15
94A20

Keywords:
Super-resolution
Vandermonde matrix
Fourier matrix
Minimum singular value
Subspace methods
MUSIC
Min-max error
Sparse recovery
Polynomial interpolation
Uncertainty principles

ABSTRACT

We consider the inverse problem of recovering the locations and amplitudes of a collection of point sources represented as a discrete measure, given $M+1$ of its noisy low-frequency Fourier coefficients. Super-resolution refers to a stable recovery when the distance Δ between the two closest point sources is less than $1/M$. We introduce a clumps model where the point sources are closely spaced within several clumps. Under this assumption, we derive a non-asymptotic lower bound for the minimum singular value of a Vandermonde matrix whose nodes are determined by the point sources. Our estimate is given as a weighted ℓ^2 sum, where each term only depends on the configuration of each individual clump. The main novelty is that our lower bound obtains an exact dependence on the *Super-Resolution Factor* $SRF = (M\Delta)^{-1}$. As noise level increases, the *sensitivity of the noise-space correlation function in the MUSIC algorithm* degrades according to a power law in SRF where the exponent depends on the cardinality of the largest clump. Numerical experiments validate our theoretical bounds for the minimum singular value and the sensitivity of MUSIC. We also provide lower and upper bounds for a min-max error of super-resolution for the grid model, which in turn is closely related to the minimum singular value of Vandermonde matrices.

1. Introduction

1.1. Background

This paper studies the inverse problem of recovering a collection of point sources from its noisy low-frequency Fourier coefficients. Suppose S point sources with amplitudes $x = \{x_j\}_{j=1}^S \in \mathbb{C}^S$ are located on

* Corresponding author.
E-mail addresses: weilinli@cims.nyu.edu (W. Li), wliao60@gatech.edu (W. Liao).

https://doi.org/10.1016/j.acha.2020.10.004
1063-5203/

an unknown discrete set $\Omega = \{\omega_j\}_{j=1}^S$ in the periodic interval $\mathbb{T} = [0, 1)$. This collection of points sources can be represented by a discrete measure,

$$\mu(\omega) := \sum_{j=1}^{S} x_j \delta_{\omega_j}(\omega), \tag{1.1}$$

where δ_{ω_j} denotes the Dirac measure supported at ω_j and Ω is the support of μ, denoted $\mathrm{supp}(\mu)$. A uniform array of $M+1$ sensors collects measurements of the point sources. Suppose the k-th sensor collects the k-th noisy Fourier coefficient of μ:

$$y_k := \widehat{\mu}(k) + \eta_k := \int_{\mathbb{T}} e^{-2\pi i k\omega}\, d\mu(\omega) + \eta_k = \sum_{j=1}^{S} x_j e^{-2\pi i k\omega_j} + \eta_k, \qquad k = 0, 1, \ldots, M, \tag{1.2}$$

where $\widehat{\mu}$ is the Fourier transform of μ and η_k represents some unknown noise at the k-th sensor. Our goal is to accurately recover μ, which consists of the support Ω and the amplitudes $x \in \mathbb{C}^S$, from the noisy low-frequency Fourier data $y = \{y_k\}_{k=0}^M \in \mathbb{C}^{M+1}$.

The measurement vector y can be expressed as

$$y = \Phi x + \eta, \tag{1.3}$$

where Φ is the $(M+1) \times S$ *Fourier* or *Vandermonde* matrix (with nodes $e^{-2\pi i\omega_j}$ on the unit circle):

$$\Phi := \Phi(\Omega, M) := \Phi_M(\Omega) := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{-2\pi i\omega_1} & e^{-2\pi i\omega_2} & \cdots & e^{-2\pi i\omega_S} \\ \vdots & \vdots & \vdots & \vdots \\ e^{-2\pi iM\omega_1} & e^{-2\pi iM\omega_2} & \cdots & e^{-2\pi iM\omega_S} \end{bmatrix}. \tag{1.4}$$

While it is convenient to re-formulate the measurement vector y in the linear system (1.3), we do not have access to the sensing matrix Φ because it depends on the unknown Ω. This inverse problem is referred to as single-snapshot spectral estimation, as only one snapshot of measurements is taken by the sensors. This problem appears in many interesting imaging and signal processing applications, including remote sensing [1], inverse scattering [2,3], Direction-Of-Arrival (DOA) estimation [4,5] and spectral analysis [6].

A key step is to estimate the support Ω and then the amplitudes x can be recovered as the least-squares solution of (1.3). In the noisy case, the stability of this inverse problem crucially depends on Ω. The *minimum separation* of Ω has been widely used to describe the stability of this inverse problem. It is defined as

$$\Delta := \Delta(\Omega) := \min_{1 \le j < k \le S} |\omega_j - \omega_k|_{\mathbb{T}}, \quad \text{where} \quad |\omega|_{\mathbb{T}} := \min_{n \in \mathbb{Z}} |\omega - n|.$$

The Heisenberg uncertainty principle implies that the spatial and frequency localization of signals are inversely proportional. When we have access to only M Fourier coefficients of μ, the recovery is sensitive to noise whenever $\Delta < 1/M$. In the imaging community, $1/M$ is called the *Rayleigh Length* (*RL*), and it is regarded as the resolution that a standard imaging system can resolve [7]. Super-resolution refers to the capability of recovering point sources when $\Delta < 1/M$. The *super-resolution factor* (SRF) is $1/(M\Delta)$, which characterizes the maximum number of points in Ω that is contained in $1/M$.

The first recovery method was invented by Prony [8]. The classical Prony's method [8] can recover μ exactly in the noiseless setting, but it is very sensitive to noise [9]. Numerous modifications were attempted to improve its numerical behavior [10–12]. In the signal processing community, a class of subspace methods has been widely used in applications, including MUltiple SIgnal Classification (MUSIC) [5], Estimation

of Signal Parameters via Rotational Invariance Technique (ESPRIT) [13], and the Matrix Pencil Method (MPM) [14]. In the past ten years, super-resolution was addressed with an optimization approach, such as the total variation minimization (TV-min) [15–19].

Existing mathematical theories on the recovery of μ can be divided into three main categories: (1) The *well-separated case* is when $\Delta \geq 1/M$, and in which case, we say that Ω is *well-separated*. There are many polynomial-time algorithms that *provably* recover μ with high accuracy. These methods include total variation minimization (TV-min) [15–19], greedy algorithms [20–24], and subspace methods [14,25–28,13,5]. These works address the discretization error and basis mismatch issues [29,21,30] arising in compressed sensing [31,32]. (2) The *super-resolution regime* is when $\Delta < 1/M$. There are two main approaches to achieve super-resolution. The optimization-based methods require certain assumptions, such as positivity [33–35]or certain sign constraints [36] of μ. The classical subspace methods [5,13,14] can recover complex measures and have super-resolution phenomenon. Meanwhile, there are many open problems regarding the stability of subspace methods with a single snapshot of measurements. (3) Super resolution is addressed from an information theoretic point of view in [37,38], where the point sources are located on a grid of $\mathbb{R}$ with spacing $1/N$ and measurements are the noisy continuous Fourier transform of the points. These works derived asymptotic lower and upper bounds for a min-max error, when the grid spacing is sufficiently small. Recently, [39] studied the off-the-grid min-max error when the support contains a single cluster of nodes and other nodes are well separated.

This paper addresses two important questions: *(1) What is the fundamental limit of super-resolution?* To quantify the fundamental difficulty of super-resolution, we use the concept of min-max error introduced by Donoho [37]. The min-max error is the reconstruction error incurred by the best possible algorithm in the worst case scenario. For technical reasons, we assume that the point sources are located on a grid with spacing $1/N$ when we study the min-max error. We follow the theme in [37,38] to relate the min-max error with $\sigma_{\min}(\Phi(\Omega, M))$ for the worst subset Ω on the grid. *(2) What is the stability of subspace methods?* The focus of this paper is the MUSIC algorithm [5]. MUSIC can recover *complex* measures and is well known for its super-resolution capability [40]. However, there are many unanswered questions related to its stability when $\Delta < 1/M$. In the classical Direction-Of-Arrival setting [4] with multiple snapshots of measurements, the Cramér-Rao bound [41–43] gives a stability of MUSIC with respect to noise, #Snapshot and the source separation Δ, in the asymptotic scenario when #Snapshot $\to \infty$ and $\Delta \to 0$. These works can not be directly applied to the single-snapshot setting. This paper aims to establish a sensitivity analysis for the MUSIC algorithm, which is different from the Cramér-Rao bound since the noise statistics are not emphasized. Recently, it was shown in [26] that the perturbation of the noise-space correlation function in the single-snapshot MUSIC is closely related with $\sigma_{\min}(\Phi)$. The key question is to accurately estimate $\sigma_{\min}(\Phi)$ for a given support Ω.

1.2. Contributions

This paper has three main contributions: accurate lower bounds for $\sigma_{\min}(\Phi)$ under geometric assumptions of the support set, improvements to the min-max error of super-resolution by sparsity constraints, and a sensitivity analysis of the noise-space correlation function in the MUSIC algorithm. We informally summarize our main results here and postpone the formal definitions until later.

(1) We consider a geometric clumps model for Ω. Assume that: (1) Ω can be written as the disjoint union of A finite sets called clumps, where each clump is contained in an interval of length $1/M$ and the distance between any two clumps is at least β/M; (2) the minimum separation of Ω is at least α/M.
Theorem 2.7 shows that if λ_a denotes the cardinality of the a-th clump, for any $\alpha > 0$ and sufficiently large $\beta > 0$, there exist constants $\{C_a(\lambda_a, M)\}_{a=1}^{A}$ such that

$$\sigma_{\min}(\Phi) \geq \sqrt{M} \left(\sum_{a=1}^{A} \left(C_a(\lambda_a, M)\, \alpha^{-\lambda_a+1} \right)^2 \right)^{-1/2}. \tag{1.5}$$

The explicit formula for C_a is given in equation (2.5) and we discuss its dependence on λ_a and M in Remark 2.9. The main novelty of this bound is that the exponent on $\alpha^{-\lambda_a+1} = \mathrm{SRF}^{\lambda_a-1}$ only depends on λ_a as opposed to S, which shows that $\sigma_{\min}(\Phi)$ depends on the local geometry of Ω. We provide an upper bound for $\sigma_{\min}(\Phi)$ in Proposition 2.10 and numerical experiments in Section 2.3 to show that the dependence on α is tight.

(2) We derive an estimate of the min-max error under sparsity constraints, which illustrates why geometric assumptions on Ω are both natural and necessary. Suppose S and Δ are known and for technical reasons, we assume that Ω lies on a fine grid with spacing $1/N$. The min-max error $\mathcal{E}(M, N, S, \delta)$ for this model is defined to be the error incurred by the best possible recovery algorithm(s) when the error is measured with respect to the worst case measure and noise η satisfying $\|\eta\|_2 \leq \delta$.
Theorem 3.8 provides explicit constants $A(M, S)$ and $B(M, S)$ such that

$$A(M,S) \left(\frac{N}{M}\right)^{2S-1} \delta \leq \mathcal{E}(M, N, S, \delta) \leq B(M,S) \left(\frac{N}{M}\right)^{2S-1} \delta. \tag{1.6}$$

The dominant factor in both inequalities is $(N/M)^{2S-1} = \mathrm{SRF}^{2S-1}$. Hence, without any prior geometric assumptions on Ω, no algorithm can accurately estimate every measure and noise, unless δ is smaller than SRF^{-2S+1}.

(3) We provide a sensitivity analysis of the noise-space correlation function in MUSIC under the clumps model. MUSIC amounts to recovering the point sources from the S smallest local minima of a noise-space correlation function $\mathcal{R}$. Corollary 4.5 shows that when Ω satisfies the clumps assumption in Theorem 2.7, then for arbitrary noise η, in order to guarantee an ε-perturbation of $\mathcal{R}$ in the supremum norm, the noise-to-signal ratio that MUSIC can tolerate obeys the following scaling law,

$$\frac{\|\eta\|_2}{x_{\min}} \propto \sqrt{M} \left(\sum_{a=1}^{A} \left(C_a(\lambda_a, M/2)\, \alpha^{-\lambda_a+1} \right)^2 \right)^{-1} \varepsilon. \tag{1.7}$$

Our result shows that the sensitivity of the noise-space correlation function in MUSIC is exponential in $1/\mathrm{SRF}$, and importantly, the exponent depends on λ_a instead of the sparsity S. This estimate is verified by numerical experiments. A perturbation bound for deterministic noise is given in Corollary 4.5.

Let us briefly discuss the implications of our main results to super-resolution. Inequalities (1.6) show that, super-resolution solely based on sparsity and the minimum separation is impossible unless the noise is smaller than SRF^{-2S+1}. The results can be greatly improved if the structure of Ω is exploited. Inequality (1.7), derived from (1.5), indicates that the perturbation of $\mathcal{R}$ is small if the noise is smaller than $\mathrm{SRF}^{-2\lambda+1}$ as opposed to SRF^{-2S+1} where $\lambda = \max_a \lambda_a$ and $S = \sum_a \lambda_a$. This rigorously confirms prior numerical evidence that MUSIC can succeed in the super-resolution regime, if λ is sufficiently small. Although our analysis only pertains to perturbations of $\mathcal{R}$, this is an inherent feature of MUSIC.

1.3. Outline

The remainder of this paper is organized as follows. Since this paper encapsulates three main topics, we present each in its own section and related work is located in the last subsection. Our estimates for $\sigma_{\min}(\Phi)$ and proof strategy are contained in Section 2 and the min-max error under sparsity constraints is studied in Section 3. Numerical experiments are included highlighting their accuracy. Section 4 explains the MUSIC algorithm, and includes a new sensitivity analysis of the noise-space correlation function under the clumps

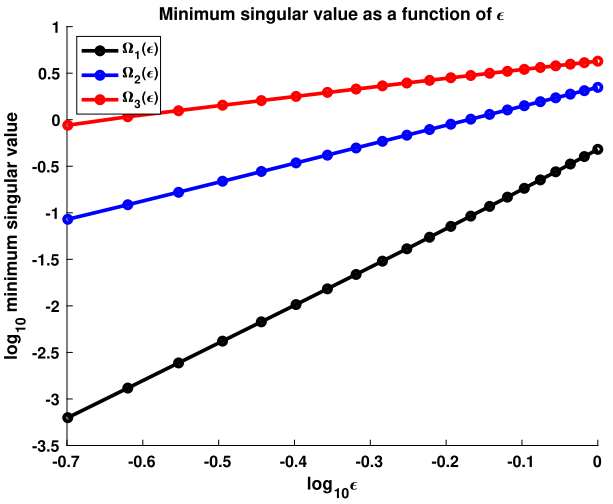


Fig. 1. Consider the sets $\Omega_1(\varepsilon)$, $\Omega_2(\varepsilon)$, $\Omega_3(\varepsilon)$ defined in (2.1) each with $\Delta(\varepsilon) = 0.01\varepsilon$. The functions $\varepsilon \mapsto \sigma_{\min}(\Phi_M(\Omega_j(\varepsilon)))$ is exponential with very different exponents. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

model. Numerical simulations are provided to validate our theoretical analysis on MUSIC. Appendices A, B, and C contain the proofs for all the theorems, propositions, and lemmas, respectively.

2. Minimum singular value of Vandermonde matrices

2.1. Lower bounds under a clumps model

The quantity $\sigma_{\min}(\Phi)$ is extremely sensitive to the "geometry" or configuration of Ω in the super-resolution regime $\Delta < 1/M$. To support this assertion, Fig. 1 provides examples of three sets

$$\begin{aligned} \Omega_1(\varepsilon) &= \varepsilon\{0, 0.01, 0.02, 0.03, 0.04\}, \\ \Omega_2(\varepsilon) &= \varepsilon\{0, 0.01, 0.02, 0.4, 0.5\}, \\ \Omega_3(\varepsilon) &= \varepsilon\{0, 0.01, 0.3, 0.4, 0.5\} \end{aligned} \tag{2.1}$$

where $0.2 \leq \varepsilon \leq 1$. We fix $M = 50$ and $S = 5$. These three support sets have the same cardinality and minimum separation, but the minimum singular values of their associated Vandermonde matrices have different behaviors. This simple numerical experiment demonstrates that it is impossible to accurately describe $\sigma_{\min}(\Phi)$ solely in terms of Δ. A more sophisticated description of the "geometry" of Ω is required.

We introduce a clumps model where Ω consists of well-separated subsets called *clumps*, where each clump contains several points that can be closely spaced.

Assumption 2.1 *(Clumps model).* We say that a finite set $\Omega \subseteq \mathbb{T}$ consists of A clumps with parameters (A, M, S, β) if the following hold.

(1) Ω has cardinality S and can be written as a disjoint union of A sets, $\Omega = \bigcup_{a=1}^{A} \Lambda_a$, where each *clump* Λ_a is a finite set contained in an open interval of length $1/M$.
(2) If $A > 1$, the distance between two clumps, defined as

$$\operatorname{dist}(\Lambda_m, \Lambda_n) := \min_{\omega_j \in \Lambda_m,\, \omega_k \in \Lambda_n} |\omega_j - \omega_k|_{\mathbb{T}},$$

satisfies $\min_{m\neq n}\operatorname{dist}(\Lambda_m,\Lambda_n)\geq \beta/M$.

Throughout the paper, we denote the cardinality of Λ_a by λ_a.

Definition 2.2 *(Complexity).* For any finite set $\Omega=\{\omega_k\}_{k=1}^S\subseteq\mathbb{T}$, its *complexity* at $\omega_j\in\Omega$ is the quantity,

$$\rho_j:=\rho_j(\Omega,M):=\prod_{\omega_k\in\Omega:\ 0<|\omega_k-\omega_j|_{\mathbb{T}}<1/M}\frac{1}{\pi M|\omega_j-\omega_k|_{\mathbb{T}}}.$$

Our Theorem 2.3 below is our most general lower bound for $\sigma_{\min}(\Phi)$, which is given in terms of a weighted ℓ^2 aggregate of the complexity of Ω at each point. The proof of Theorem 2.3 can be found in Appendix A.1.

Theorem 2.3. *Fix positive integers A,M,S with $M\geq 2S^2$. Suppose Ω satisfies Assumption 2.1 with parameters (A,M,S,β). If $A>1$, assume that*

$$\beta\geq\max_{1\leq a\leq A}\max_{\omega_j\in\Lambda_a}10\lambda_a^{5/2}(S\rho_j)^{1/(2\lambda_a)}. \tag{2.2}$$

For each $1\leq a\leq A$, we define the constant

$$B_a:=B_a(\lambda_a,M):=\frac{20\sqrt{2}}{19}\Big(1-\frac{\pi^2}{3\lambda_a^2}\Big)^{-(\lambda_a-1)/2}\Big(\frac{M}{\lambda_a}\Big)^{\lambda_a-1}\Big\lfloor\frac{M}{\lambda_a}\Big\rfloor^{-(\lambda_a-1)}.$$

Then the minimum singular value of $\Phi=\Phi(\Omega,M)$ defined in (1.4) satisfies

$$\sigma_{\min}(\Phi)\geq\sqrt{M}\Big(\sum_{a=1}^{A}\sum_{\omega_j\in\Lambda_a}(B_a\lambda_a^{\lambda_a-1}\rho_j)^2\Big)^{-1/2}. \tag{2.3}$$

Remark 2.4. The constant B_a is insensitive to the geometry of each clump Λ_a because it only depends on M and λ_a, and it is also independent of S. The dependence of B_a on M is weak because

$$\Big(\frac{M}{\lambda_a}\Big)\Big\lfloor\frac{M}{\lambda_a}\Big\rfloor^{-1}=1+o\Big(\frac{M}{\lambda_a}\Big)\quad\text{as}\quad\frac{M}{\lambda_a}\to\infty.$$

It is possible to upper bound B_a by a constant that does not depend on M. Clearly B_a does not dependent on M when $\lambda_a=1$, and for $\lambda_a\geq 2$, since $t/\lfloor t\rfloor\leq t/(t-1)\leq 2$ for $t>1$, we see that

$$B_a\leq\frac{20\sqrt{2}}{19}\Big(1-\frac{\pi^2}{3\lambda_a^2}\Big)^{-(\lambda_a-1)/2}2^{\lambda_a-1}.$$

We can think of B_a as a small universal constant because the function $n\mapsto(1-\pi^2/(3n^2))^{-(n-1)/2}$ defined on the integers $n\geq 2$ approaches a horizontal asymptote of 1 quickly as n increases. In the regime where each λ_a is of moderate size and M/λ_a is large, B_a is approximately $20\sqrt{2}/19\approx 1.4886$.

Remark 2.5. Although this is not the main point of the theorem, we can also apply it to the well-separated case. Assume that $\Delta\geq 10\sqrt{S}/M$. Then each clump Λ_a contains a single point, $B_a=20\sqrt{2}/19$ for each $1\leq a\leq A$, and $\rho_j=1$ for each $\omega_j\in\Omega$. We readily check that the conditions of Theorem 2.3 are satisfied, and thus,

$$\sigma_{\min}(\Phi)\geq\frac{19}{20\sqrt{2}}\sqrt{M}.$$

This shows that $\sigma_{\min}(\Phi)$ is on the order of $\sqrt{M}$ if Δ is about $\sqrt{S}$ times larger than $1/M$. This result is weaker than the one in [28], which was derived using extremal functions that are specialized to the well-separated case. Note that $\sqrt{M}$ is approximately the largest $\sigma_{\min}(\Phi)$ can be because $\sigma_{\max}(\Phi) \le \|\Phi\|_F = \sqrt{MS}$, where $\|\cdot\|_F$ is the Frobenius norm.

Theorem 2.3 is our most general lower bound for $\sigma_{\min}(\Phi)$ without a minimum separation condition. The bound in Theorem 2.3 can be reduced to a more explicit estimate if we consider the minimum separation of Ω.

Assumption 2.6 *(Clumps model with a minimum separation).* We say that a finite set $\Omega \subseteq \mathbb{T}$ satisfies a clumps model with a minimum separation with parameters (A, M, S, β, α) if Ω satisfies Assumption 2.1 with parameters (A, M, S, β) and moreover, the minimum separation of Ω satisfies $\Delta \ge \alpha/M$ with $\max_{1\le a\le A}(\lambda_a - 1) < 1/\alpha$.

Theorem 2.7 below is derived from Theorem 2.3 under Assumption 2.6, which is proved in Appendix A.2.

Theorem 2.7. *Fix positive integers A, M, S with $M \ge S^2$. Suppose Ω satisfies Assumption 2.6 with parameters (A, M, S, β, α). If $A > 1$, assume that*

$$\beta \ge \max_{1\le a\le A} \frac{20 S^{1/2} \lambda_a^{5/2}}{\alpha^{1/2}}. \tag{2.4}$$

For each $1 \le a \le A$, let $B_a := B_a(\lambda_a, M)$ be the constant defined in Theorem 2.3 and

$$C_a := C_a(\lambda_a, M) := B_a \left(\frac{\lambda_a}{\pi}\right)^{\lambda_a - 1} \Big(\sum_{j=1}^{\lambda_a} \prod_{k=1,\ k\ne j}^{\lambda_a} \frac{1}{(j-k)^2} \Big)^{1/2}. \tag{2.5}$$

Then the minimum singular value of $\Phi = \Phi(\Omega, M)$ defined in (1.4) satisfies

$$\sigma_{\min}(\Phi) \ge \sqrt{M} \left(\sum_{a=1}^{A} \left(C_a \alpha^{-\lambda_a + 1} \right)^2 \right)^{-1/2}. \tag{2.6}$$

Remark 2.8. The main difference between Theorems 2.3 and 2.7 is that, Theorem 2.3 is more general and accurate, while Theorem 2.7 is more concrete since it bounds $\sigma_{\min}(\Phi)$ in terms of $\alpha = 1/\mathrm{SRF}$. The conclusions in Theorem 2.3 and Theorem 2.7 are identical for the special case where each clump Λ_a consists of λ_a points consecutively spaced by α/M. For all other configurations of Ω, Theorem 2.3 provides a more accurate lower bound. The separation condition (2.2) is always weaker than (2.4).

Remark 2.9. According to Remark 2.4, the constant B_a can be thought of as a universal constant. For the behavior of C_a, we have a simple upper bound $C_a \le B_a \lambda_a^{1/2} (\lambda_a/\pi)^{\lambda_a - 1}$, which is a reasonable bound for small and moderate λ_a. A more refined argument (see Lemma C.2 in Appendix C.4) shows that

$$C_a \le B_a \Big(\frac{\lambda_a}{\pi}\Big)^{\lambda_a - 1} 2\pi e \sqrt{\lambda_a} \Big(\frac{\lambda_a}{e} - 1\Big)^{-\lambda_a} e^{\lambda_a} = \frac{2\pi^2 e B_a}{\sqrt{\lambda_a}} \Big(\frac{e\lambda_a}{\pi(\lambda_a - e)}\Big)^{\lambda_a} e^{\lambda_a}. \tag{2.7}$$

For large λ_a, the right hand side scales like $C B_a \lambda_a^{-1/2} c^{\lambda_a}$ for universal constants $C, c > 0$.

The main contribution of this theorem is the exponent on $\mathrm{SRF} = (M\Delta)^{-1} = 1/\alpha$, which depends on λ_a as opposed to the sparsity S. Let us look at a special case of Ω, where each clump Λ_a contains λ points

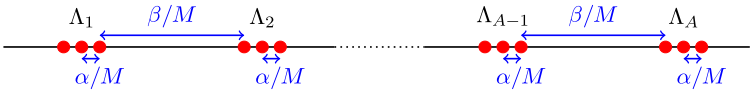


Fig. 2. An example of clumps model with a minimum separation.

equally spaced by α/M and the distance between clumps is β/M where β is properly chosen such that (2.4) holds. See Fig. 2 for an illustration.

In this case, Theorem 2.7 implies

$$\sigma_{\min}(\Phi(\Omega, M)) \geq C(\lambda) A^{-1/2} \underbrace{\sqrt{M}}_{\text{scaling}} \cdot \underbrace{\alpha^{\lambda-1}}_{(1/\text{SRF})^{\lambda-1}}, \tag{2.8}$$

where the constant $C(\lambda)$ only depends on λ. Here, $\sqrt{M}$ is a natural scaling factor because each column of $\Phi(\Omega, M)$ has Euclidean norm $\sqrt{M+1}$. Importantly, the lower bound scales like $\alpha^{\lambda-1} = (1/\text{SRF})^{\lambda-1}$ where λ is the cardinality of each clump instead of S, which matches our intuition that the conditioning of $\Phi(\Omega, M)$ should only depend on how complicated each individual clump is.

2.2. Upper bound for the minimum singular value

Our lower bounds in Theorems 2.3 and 2.7 show that, for the special support in Fig. 2, our lower bound for $\sigma_{\min}(\Phi)$ depends linearly on $\text{SRF}^{-\lambda+1}$. Proposition 2.10 below proves that this dependence is sharp. Its proof can be found in Appendix B.1 and uses a method similar to one in [37].

Proposition 2.10. *Fix positive integers M, S, λ such that $\lambda \leq S \leq M-1$. Let $\omega \in \mathbb{T}$ and $\alpha > 0$ such that*

$$\alpha \leq \frac{1}{C(\lambda)\sqrt{M+1}}, \quad \textit{where} \quad C(\lambda) = 2\pi \sum_{j=0}^{\lambda-1} \binom{\lambda-1}{j} \frac{j^\lambda}{\lambda!}. \tag{2.9}$$

Assume that $\Omega = \{\omega_j\}_{j=1}^S \subseteq \mathbb{T}$ contains the set,

$$\Lambda = \omega + \left\{0, \frac{\alpha}{M}, \ldots, \frac{(\lambda-1)\alpha}{M}\right\}.$$

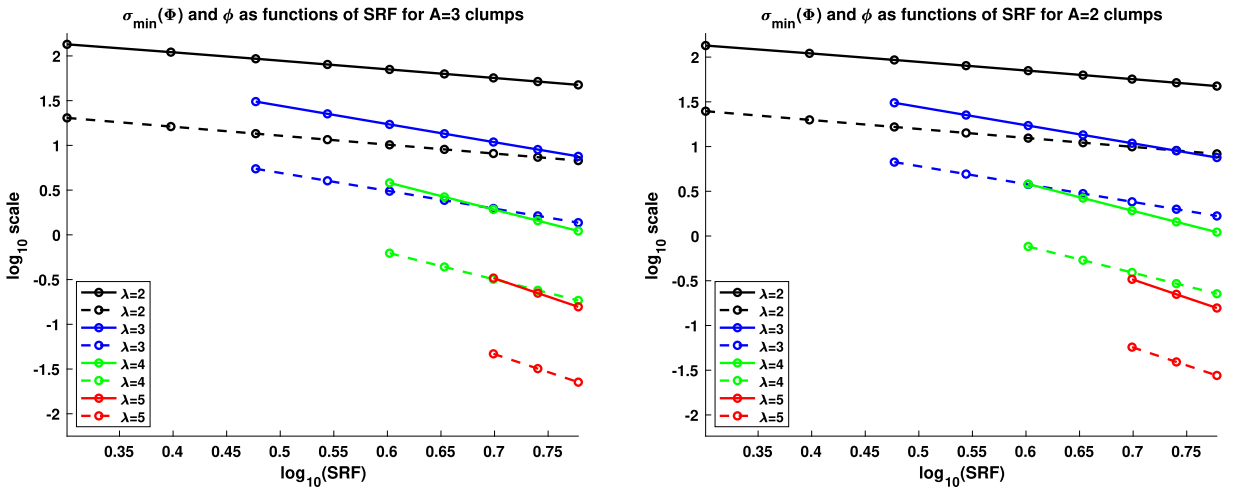
Let $\Phi = \Phi(\Omega, M)$ be the $(M+1) \times S$ Vandermonde matrix associated with Ω and M. Then

$$\sigma_{\min}(\Phi) \leq \binom{2\lambda-2}{\lambda-1}^{-1/2} 2\sqrt{M+1}\, (2\pi\alpha)^{\lambda-1}.$$

Proposition 2.10 shows that if Ω contains a set Λ, which consists of λ points equally spaced by α/M for a sufficiently small α, then $\sigma_{\min}(\Phi)$ depends on $\alpha^{\lambda-1}$. This implies that the dependence on α in the order of $\alpha^{\lambda-1}$ in our lower bound of (2.8) is tight.

2.3. Numerical accuracy of Theorems 2.3 and 2.7

To numerically evaluate the accuracy of Theorems 2.3 and 2.7, we consider the case where Ω consists of A clumps, each clump contains λ equally spaced points separated by α/M, and the clump separation is β/M. We fix M, vary $\text{SRF} = 1/\alpha$, and select β to be the right hand side of inequality (2.4). As discussed in Remark 2.8, both Theorem 2.3 and 2.7 provide the identical lower bound for $\sigma_{\min}(\Phi)$ for this example of Ω, which is of the form



(a) 2-clump model: $A = 2$ (b) 3-clump model: $A = 3$

Fig. 3. Plots of $\sigma_{\min}(\Phi)$ (solid lines) and ϕ in (2.10) (dashed lines), as functions of $\mathrm{SRF} = 1/\alpha$ when the other parameters A, λ, M are fixed. For all the curves, we set $M = 50000$ and let Ω consist of A clumps, where each clump contains λ points equally separated by α/M and the clump separation is given by the right hand side of inequality (2.4). We consider the following range of parameters: $2 \le A \le 3$, $2 \le \lambda \le 5$, and $\lambda \le \mathrm{SRF} \le 6$.

$$\sigma_{\min}(\Phi) \ge C(\lambda)\sqrt{\frac{M}{A}}\,\mathrm{SRF}^{-\lambda+1} =: \phi(A, M, \lambda, \mathrm{SRF}). \tag{2.10}$$

Fig. 3 displays plots of $\sigma_{\min}(\Phi)$ and ϕ as functions of SRF, for several values of A and λ. Their slopes are identical, which establishes that the dependence on SRF in inequality (2.10) is exact. In other words, the experiments verify that $\sigma_{\min}(\Phi)$ should only depend on the cardinality of each clump and not on the total number of points S under the theorems' assumptions.

2.4. Proof strategy by polynomial duality and interpolation

Our primary method for lower bounding $\sigma_{\min}(\Phi)$ is through a dual characterization via trigonometric interpolation. We begin with some notation and definitions. Let $\mathcal{P}(M)$ be the space of all smooth functions f on $\mathbb{T}$ such that for all $\omega \in \mathbb{T}$,

$$f(\omega) = \sum_{m=0}^{M} \widehat{f}(m) e^{2\pi i m\omega}.$$

We call f a *trigonometric polynomial* of degree at most M.

Definition 2.11 *(Polynomial interpolation set).* Given $\Omega = \{\omega_j\}_{j=1}^{S} \subseteq \mathbb{T}$ and $v \in \mathbb{C}^S$, the polynomial interpolation set, denoted by $\mathcal{P}(\Omega, M, v)$, is the set of $f \in \mathcal{P}(M)$ such that $f(\omega_j) = v_j$ for each $1 \le j \le S$.

We have the following duality between the minimum singular value of Fourier matrices and the polynomial interpolation set, and its proof is in Appendix B.2.

Proposition 2.12 *(Exact duality). Fix positive integers M and S such that $S \le M - 1$. For any set $\Omega = \{\omega_j\}_{j=1}^{S} \subseteq \mathbb{T}$, let $\Phi = \Phi(\Omega, M)$ be the $(M+1) \times S$ Vandermonde matrix associated with Ω and M. For any $w \in \mathbb{C}^S$, the set $\mathcal{P}(\Omega, M, w)$ is non-empty. If $\sigma_{\min}(\Phi) = \|\Phi v\|_2$ for some unit norm vector $v \in \mathbb{C}^S$, then*

$$\sigma_{\min}(\Phi) = \max_{f \in \mathcal{P}(M,\Omega,v)} \|f\|_{L^2(\mathbb{T})}^{-1}.$$

There are two main technical difficulties with using the proposition. First, in the extreme case that $S \ll M$, we expect $\mathcal{P}(\Omega, M, v)$ to contain a rich set of functions. However, we do not know much about this set, aside from it being convex. Moreover, this set is extremely dependent on Ω because we know that $\sigma_{\min}(\Phi)$ is highly sensitive to the configuration of Ω. Second, we have no information about v, a right singular vector associated with $\sigma_{\min}(\Phi)$. Yet, in order to invoke the duality result, we must examine the set $\mathcal{P}(\Omega, M, v)$.

It turns out that we can circumvent both of these issues, but doing so will introduce additional technicalities and difficulties. Proposition 2.13 below is a relaxation of the exact duality in Proposition 2.12, which will provide us with an extra bit of flexibility. Its proof can be found in Appendix B.3.

Proposition 2.13 *(Robust duality). Fix positive integers M and S such that $S \leq M-1$. For any set $\Omega = \{\omega_j\}_{j=1}^S \subseteq \mathbb{T}$, unit norm vector $v \in \mathbb{C}^S$, and $\varepsilon \in \mathbb{C}^S$ with $\|\varepsilon\|_2 < 1$, if there exists $f \in \mathcal{P}(M)$ such that $f(\omega_j) = v_j + \varepsilon_j$ for each $1 \leq j \leq S$, then*

$$\|\Phi v\|_2 \geq (1 - \|\varepsilon\|_2)\|f\|_{L^2(\mathbb{T})}^{-1}.$$

In order to use these results to derive a lower bound for $\sigma_{\min}(\Phi)$ for a given Ω, for each unit norm $v \in \mathbb{C}^S$, we construct a $f_v \in \mathcal{P}(\Omega, M, v)$ and then bound $\|f_v\|_{L^2(\mathbb{T})}$ uniformly in v. This process must be done carefully; otherwise we would obtain a loose lower bound for $\sigma_{\min}(\Phi)$. This construction is technical and our approach is inspired by uncertainty principles for trigonometric polynomials [44] and uniform dilation problems on the torus [45,46].

2.5. Related work on the conditioning of Vandermonde matrices

The spectral properties of a Vandermonde matrix greatly depend on its nodes. Real [47–50], random [51–53], and those located within the unit complex disk [54,55] have been studied. In this paper, we study tall and deterministic Vandermonde matrices whose nodes are on the complex unit circle, which are generalizations of harmonic Fourier matrices.

In the well-separated case $\Delta \geq 1/M$, Vandermonde matrices are well-conditioned [56,57,26,28]. There are fewer available results for the super-resolution regime $\Delta \leq 1/M$. On one extreme, there exists a set Ω for which the conditioning of Φ scales linearly in SRF^{S-1} when all other parameters are fixed [28]. On the other hand, Theorems 2.3 and 2.7 show that this is overly pessimistic under appropriate clumps model.

Classical work, such as [58], primarily focused on square Vandermonde matrices. One of the earliest results for rectangular ones that also incorporates geometric information follows by combining [54, Theorem 1] and [58, Theorem 1], which yields the estimate

$$\sigma_{\min}(\Phi) \geq \sqrt{\frac{M}{S}} \min_{1 \leq j \leq S} \prod_{k=1,\ k \neq j}^{S} \frac{|\omega_j - \omega_k|_{\mathbb{T}}}{2}. \tag{2.11}$$

When Ω consists of separated clumps with parameters (A, M, S, β), this inequality yields

$$\sigma_{\min}(\Phi) \geq 2^{-S} \sqrt{\frac{M}{S}} \min_{1 \leq a \leq A} \left(\frac{\beta}{M}\right)^{S-\lambda_a} \Delta^{\lambda_a - 1}.$$

When M is large, this lower bound is significantly worse than the inequality in Theorem 2.7 which depends on terms involving $\mathrm{SRF}^{-\lambda_a+1} = (M\Delta)^{\lambda_a-1}$.

The preceding discussion is closely related to an important difference between the complexity ρ_j and $\prod_{j \neq k} |\omega_j - \omega_k|^{-1}$, the reciprocal of the product term in (2.11). Notice that ρ_j is local in the sense it only

depends on the structure of Ω in a $1/M$ neighborhood of ω_j and it includes the πM scaling term. We include the π factor because we use the Fourier transform with the 2π convention in (1.2), but the M factor is very important. In the super-resolution and imaging communities, it is known that $\mathrm{SRF} = (M\Delta)^{-1}$, not Δ^{-1}, is the correct quantity to describe the feasibility of super-resolution. The complexity can be viewed as a refined notion of $\mathrm{SRF}^S = (M\Delta)^{-S}$ since ρ_j only depends on the structure of Ω near ω_j.

In the process of revising the first version of this manuscript, [59] independently derived lower bounds for $\sigma_{\min}(\Phi)$ with clustered nodes. Here the differences between their [59, Corollary 3.6] and our Theorem 2.7: (1) They assume that Ω consists of clumps that are all contained in an interval of length approximately $1/S^2$. For ours, the clumps can be spread throughout $\mathbb{T}$ and not have to be concentrated on a sub-interval. (2) They require the aspect ratio $M \geq 4S^3$, whereas we only need $M \geq 2S^2$. They also require an upper bound on M/S, which prohibits their Vandermonde matrix from being too tall. (3) Their lower bound is of the form, $\sigma_{\min}(\Phi) \geq C_S\sqrt{M} \cdot \mathrm{SRF}^{-\lambda+1}$, where C_S depends only on S and scales like S^{-2S}. According to (2.7), our constant C_a scales like $\lambda_a^{-1/2}c^{\lambda_a}$ for a universal constant $c > 0$, which is independent of S. (4) Our clump separation condition (2.4) in Theorem 2.7 depends on $\alpha = \mathrm{SRF}^{-1}$, which means that for large SRF, the clumps have to be significantly separated. In contrast, their clump separation condition only depends on S but all of their clumps must remain in an interval with length approximately $1/S^2$.

During the review period of this paper, the singular values of the same Vandermonde matrices with nearly colliding pairs were analyzed in [60] and [61, Corollary 4.2]. The nearly colliding pairs form a special case of the clumps model when each clump has at most two closely spaced points. In this special case, the separation condition between the clumps is improved to be independent of Δ in [60,61]. A recent work [62] refined the main strategy in the proofs of Theorems 2.3 and 2.7 to improve the constant C_a and strengthen our ℓ^2 bound to an ℓ^∞ one. They also provided an extension to the multi-dimensional case. Another recent paper [63] provides bounds for all singular values of Vandermonde matrices under a clumps model.

Lower bounds for the conditioning of arbitrarily sized $p \times q$ cyclically contiguous sub-matrices of the $N \times N$ discrete Fourier transform matrix was derived in [64]. This sub-matrix is equivalent to our $\Phi(\Omega, p)$, where $\Omega := \{k/N\}_{k=1}^q$. A comparison between our Proposition 2.10 and the main result in [64] can be found in the referenced paper.

3. Min-max error and worst case analysis

3.1. A grid model and the min-max error

In order to understand the fundamental limits of super-resolution without any geometric information about Ω, in this section, we study the min-max error under a grid and sparsity assumption on Ω. Our results will illustrate that it is not only natural, but also necessary to take the geometric information of Ω into account.

Suppose Ω has cardinality S and is a subset of $\{n/N\}_{n=0}^{N-1}$, which we refer to as the grid with spacing $1/N$. In the super-resolution literature, this is called the *on-the-grid* model. This assumption implicitly places a minimum separation requirement so that all point sources are separated by at least $1/N$ and $\mathrm{SRF} = N/M$. In comparison to the clumps model, the one considered in this section only has a sparsity constraint and places no geometric constraints on the support set. The grid assumption is purely for technical reasons.

We will define a min-max error proposed in [37]. Fix positive integers M, N, S such that $S \leq M \leq N$ and let $\delta > 0$. We define a set of vectors in $\mathbb{C}^{M+1}$,

$$\begin{aligned}\mathcal{Y} &:= \mathcal{Y}(M, N, S, \delta)\\ &= \Big\{y \in \mathbb{C}^{M+1} : \exists\, \mu \text{ supported in } \Omega \subseteq \{n/N\}_{n=0}^{N-1},\ |\Omega| = S, \Big(\sum_{k=0}^{M} |y_k - \widehat{\mu}(k)|^2\Big)^{1/2} \leq \delta\Big\}.\end{aligned}$$

Let $\mathcal{A} = \mathcal{A}(M, N, S, \delta)$ be the set of functions φ that maps each $y \in \mathcal{Y}$ to a discrete measure φ_y supported on the grid with spacing $1/N$.

Definition 3.1 *(Min-max error).* The ℓ^2 *min-max error* for the on-the-grid model is

$$\mathcal{E}(M, N, S, \delta) := \inf_{\varphi \in \mathcal{A}} \sup_{y \in \mathcal{Y}(M,N,S,\delta)} \Big(\sum_{n=0}^{N-1} \Big| \varphi_y\Big(\frac{n}{N}\Big) - \mu\Big(\frac{n}{N}\Big) \Big|^2 \Big)^{1/2}.$$

We interpret δ as the noise level and $\mathcal{Y}$ as the "signal space" formed by all possible δ perturbations of the Fourier coefficients of a measure with at most S Dirac masses on the grid. We interpret a function $\varphi \in \mathcal{A}$ as an "algorithm" that maps a given $y \in \mathcal{Y}$ to a measure φ_y that approximates μ. By taking the infimum over all possible algorithms (which includes those that are computationally intractable), the min-max error is the reconstruction error incurred by the best algorithm, when measured against the worst case signal and noise. This quantity describes the fundamental limit of super-resolution under sparsity constraints, and no algorithm can perform better than the min-max rate.

3.2. Lower and upper bounds of the min-max error

To lower and upper bound the min-max error, we follow the approach of [38] to connect the min-max error with $\sigma_{\min}(\Phi(\Omega, M))$ for the worst subset Ω on the grid.

Definition 3.2 *(Lower restricted isometry constant).* Fix positive integers M, N, S such that $S \le M \le N$. The lower restricted isometry constant of order S is defined as

$$\Theta(M, N, S) := \min_{\Omega \subseteq \{n/N\}_{n=0}^{N-1},\ |\Omega| = S} \sigma_{\min}(\Phi(\Omega, M)).$$

Remark 3.3. While $\Theta(M, N, S)$ is related to the lower bound of the *S-restricted isometry property* (RIP) in compressive sensing [65], there is a major difference. If we select M appropriately chosen rows of the $N \times N$ Discrete Fourier Transform (DFT), then every $M \times S$ sub-matrix is well-conditioned, see [66]. However, $\Phi(\Omega, M)$ uses the first M rows of the DFT matrix, so $\Phi(\Omega, M)$ may be ill-conditioned.

The following result establishes the relationship between the min-max error and the lower restricted isometry constant. An analogue of this result for a similar super-resolution problem on $\mathbb{R}$ was proved in [38]. Their proof carries over to this discrete setting with minor modifications. To keep this paper self-contained, we prove it in Appendix B.4 and it relies on the grid assumption.

Proposition 3.4 *(Min-max error and lower restricted isometry constant). Fix positive integers M, N, S such that $2S \le M \le N$, and let $\delta > 0$. Then,*

$$\frac{\delta}{2\Theta(M, N, 2S)} \le \mathcal{E}(M, N, S, \delta) \le \frac{2\delta}{\Theta(M, N, 2S)}.$$

With this result at hand, our main focus is to derive an accurate lower bound for $\Theta(M, N, S)$. We suspect that the minimum in $\Theta(M, N, S)$ occurs when all the points in Ω are consecutively spaced by $1/N$. If we could prove this conjecture, then we could simply apply Theorem 2.7 for the single clump case to derive an upper bound of $\Theta(M, N, S)$. Without this conjecture, our strategy is to apply our duality techniques in Section 2.4 for all $\binom{N}{S}$ possible choices of Ω and then uniformly bound over all these possibilities. There are exponentially many possible such Ω, so we must be extremely careful with the estimate in order to obtain an accurate lower bound for $\Theta(M, N, S)$. The following theorem is proved in Appendix A.3.

Theorem 3.5. *Fix positive integers M, N, S such that $S \geq 2$, $M \geq 2S$, and $N \geq \pi MS$. We define the constant,*

$$C(M,S) := \Big(\frac{12-\pi^2}{24}\Big)^{1/2}\Big(\sum_{j=1}^{S}\prod_{k\neq j,\,k=1}^{S}\frac{1}{(j-k)^2}\Big)^{-1/2}\frac{1}{\sqrt{S}}\Big(\frac{\pi}{S}\Big)^{S-1}\Big(\frac{M}{S}\Big)^{-(S-1)}\Big\lfloor\frac{M}{S}\Big\rfloor^{S-1}.$$

Then we have

$$\Theta(M,N,S) \geq C(M,S)\sqrt{M}\Big(\frac{M}{N}\Big)^{S-1}. \tag{3.1}$$

Remark 3.6. As described in Remark 2.4, when M is significantly larger than S, the $(M/S)^{-(S-1)}\lfloor M/S\rfloor^{S-1}$ factor in $C(M,S)$ can be safely ignored if M/S is sufficiently large. Alternatively, we can lower bound $C(M,S)$ independent of M since $t/\lfloor t\rfloor \leq t/(t-1) \leq 2$ for all $t > 2$ and so

$$C(M,S) \geq \Big(\frac{12-\pi^2}{24}\Big)^{1/2}\Big(\sum_{j=1}^{S}\prod_{k\neq j,\,k=1}^{S}\frac{1}{(j-k)^2}\Big)^{-1/2}\frac{1}{\sqrt{S}}\Big(\frac{\pi}{S}\Big)^{S-1}2^{-(S-1)}.$$

Now we concentrate on the dependence of $C(M,S)$ on S. Since $\prod_{k\neq j,\,k=1}^{S}(j-k)^{-2} \leq 1$ for each $1 \leq j \leq S$, we have the lower bound,

$$C(M,S) \geq \frac{1}{\pi}\Big(\frac{12-\pi^2}{24}\Big)^{1/2}\Big(\frac{\pi}{S}\Big)^{S}\Big(\frac{M}{S}\Big)^{-(S-1)}\Big\lfloor\frac{M}{S}\Big\rfloor^{S-1}.$$

For large S, a more accurate bound is (see Lemma C.2 in Appendix C.4)

$$C(M,S) \geq \frac{1}{2\pi^2 e}\Big(\frac{12-\pi^2}{24}\Big)^{1/2}e^{-S}\Big(\frac{S}{2}-1\Big)^{S}\Big(\frac{\pi}{S}\Big)^{S}\Big(\frac{M}{S}\Big)^{-(S-1)}\Big\lfloor\frac{M}{S}\Big\rfloor^{S-1}.$$

For large S, the right hand side scales as Cc^{-S} for universal constants $C, c > 0$.

Remark 3.7. Although we do not identify which Ω achieves the minimum in $\Theta(M,N,S)$, the theorem suggests that the minimum is attained by translates of $\Omega_* := \{n/N\}_{n=0}^{S-1}$. Indeed, the assumption $N \geq \pi MS$ implies that Ω_* is contained in an open interval of length $1/M$. We readily check that if $\rho_j = \rho_j(\Omega_*, M)$ is the complexity of $\omega_j \in \Omega_*$, then

$$\Big(\sum_{j=1}^{S}\rho_j^2\Big)^{1/2} = \Big(\sum_{j=1}^{S}\prod_{k\neq j}\frac{1}{(j-k)^2}\Big)^{1/2}\Big(\frac{N}{\pi M}\Big)^{S-1}.$$

Notice that this is precisely the quantities that appear in Theorem 3.5.

We next present the lower and upper bounds for the min-max error, which is proved in Appendix A.4.

Theorem 3.8. *Fix positive integers S, M, N and let $\delta > 0$.*

(a) *Assume that $M \geq 4S$ and $N \geq 2\pi MS$, and let $C(M,S)$ be the constant defined in Theorem 3.5. Then, we have the upper bound,*

$$\mathcal{E}(M,N,S,\delta) \leq \frac{2\delta}{C(M,2S)}\frac{1}{\sqrt{M}}\Big(\frac{N}{M}\Big)^{2S-1}.$$

(b) *Assume that* $M \geq 2S+1$, *and* $N/M \geq 2\pi C(2S)\sqrt{M+1}$, *where* $C(2S)$ *is the constant defined in Proposition 2.10. Then, we have the lower bound,*

$$\mathcal{E}(M,N,S,\delta) \geq \frac{\delta}{4}\binom{4S-2}{2S-1}^{1/2}\frac{1}{\sqrt{M+1}}\Big(\frac{N}{2\pi M}\Big)^{2S-1}.$$

3.3. Numerical accuracy of Theorem 3.5

Let $\theta(M,N,S)$ denote the right hand side of (3.1). Note that θ is only defined for $N/M \geq \pi S$. We make two important observations before numerically evaluating the accuracy of Theorem 3.5.

First, while we would like to compare θ directly with Θ, this is not computationally feasible because we would need to enumerate through all possible Ω, for numerous values of M, N, S. Instead, we compare θ with the quantity,

$$\psi(M,N,S) := \sigma_{\min}(\Phi(\Omega_*, M)),$$

where $\Omega_* := \{n/N\}_{n=0}^{S-1}$. Note that ψ serves as a useful substitute for Θ because of the inequalities

$$\theta(M,N,S) \leq \Theta(M,N,S) \leq \psi(M,N,S). \tag{3.2}$$

Second, while both θ and ψ depend on three parameters, Theorem 3.5 and Proposition 2.10 suggest that after safely ignoring the $\sqrt{M}$ scaling factor, they should only depend on two parameters, the super-resolution factor $\mathrm{SRF} = N/M$ and the sparsity S. Additionally, we can only reliably perform the experiments for modest size of SRF^{S-1}, or else numerical round off errors become significant.

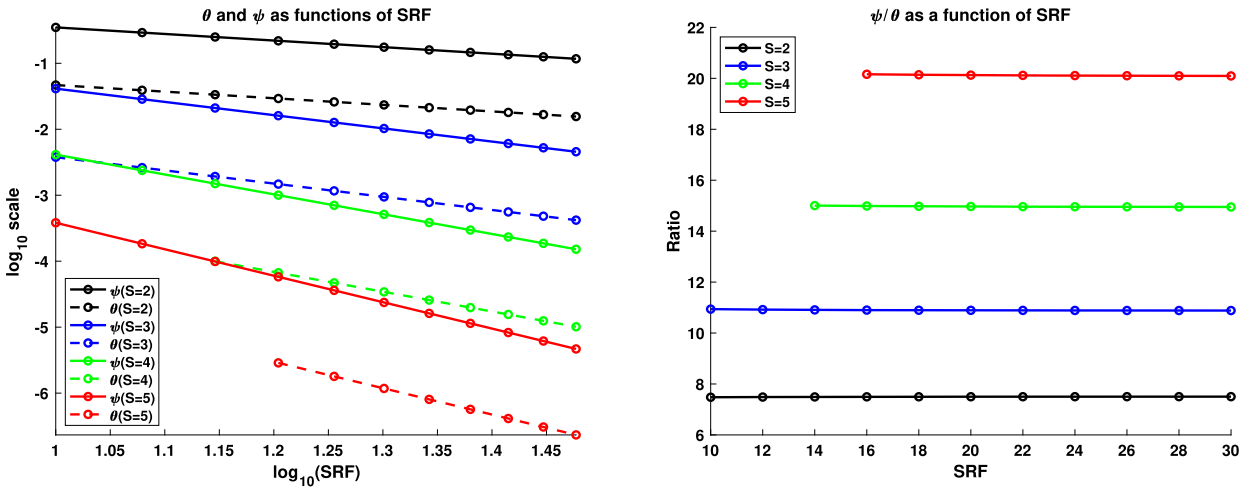
Fig. 4 (a) displays the values of θ and ψ as functions of SRF. Since the slope of both curves are identical, they verify that Theorem 3.5 correctly quantifies the dependence of the min-max error on SRF. Fig. 4 (b) displays the ratio between ψ and θ. As a consequence of inequalities (3.2), this experiment also indirectly provides us information about the ratio of Θ and θ. The lines in Fig. 4 (b) are horizontal, which confirms our theory that, if $\mathrm{SRF} \geq \pi S$, then there exists a $c(M,S) > 0$ such that

$$\Theta(M,N,S) = c(M,S)\sqrt{M}\Big(\frac{M}{N}\Big)^{S-1}.$$

3.4. Related work on super-resolution limit

A min-max formulation of super-resolution for measures on the grid can be traced back to [37] and related works [38,59]. The similarity is that all three consider *continuous* Fourier measurements and the main differences are their assumptions on μ. In [37], μ is supported in a grid of the real line $\mathbb{R}$ with constraints on the "density" of the support. In [38], μ is again supported on a grid in $\mathbb{R}$ but has S atoms without restrictions on its support. In [59], μ is supported a grid in $[0,1]$, has S atoms, and its support satisfies a clumps condition. Our setup is different from those considered in [37,38,59] in the sense that we consider *discrete* Fourier measurements. Even though the min-max error in Theorem 3.8, is strikingly similar to those found in [37,38,59], their results do not imply ours or vice versa.

Recently, the authors of [39] studied the min-max error with continuous Fourier measurements, when the support set contains a *single cluster* of λ nodes while the remaining $S-\lambda$ ones are well separated. In our notation, this corresponds to $\lambda_1 = \lambda$ and $\lambda_a = 1$ for all $a \geq 2$. Under appropriate conditions on the various parameters, [39, Theorem 2.10] shows that: (1) the min-max support error for the cluster nodes, as a function of SRF, scales linearly in $\mathrm{SRF}^{2\lambda-2}$; (2) the min-max amplitude error for the cluster nodes,



(a) $\theta(M,N,S)$ and $\psi(M,N,S)$ (b) $\psi(M,N,S)/\theta(M,N,S)$

Fig. 4. The left figure displays $\theta(M,N,S)$ and $\psi(M,N,S)$ as functions of SRF $= N/M$, while the right figure displays their ratio. For both experiments, we consider the range of parameters $2 \le S \le 5$, $\pi S \le$ SRF and $10 \le$ SRF ≤ 30, respectively.

scales linearly in $\mathrm{SRF}^{2\lambda-1}$. Their result does not imply ours or vice versa, since we consider different models. Nonetheless, their min-max amplitude error is consistent with our Theorem 3.8 if the support set contains a cluster of λ closely spaced points.

Regarding the detection of the number of sources, [67] provided an information-theoretic condition for the SRF and the noise level such that the number of sources can be correctly detected.

4. MUSIC and its super-resolution limit

In signal processing, a class of subspace methods, including MUSIC [5], is widely used due to their superior numerical performance. It was well known that MUSIC has super-resolution phenomenon [40]. The resolution limit of MUSIC was discovered by numerical experiments in [26], but has never been rigorously proved. A main contribution of this paper is to prove the sensitivity of the noise-space correlation function in MUSIC under the clumps model.

4.1. The MUSIC algorithm

MUSIC is built upon a Hankel matrix and its Vandermonde decomposition. For a fixed positive integer $L \le M$, we form the Hankel matrix of y:

$$\mathcal{H}(y) := \begin{bmatrix} y_0 & y_1 & \cdots & y_{M-L} \\ y_1 & y_2 & \cdots & y_{M-L+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & \cdots & y_M \end{bmatrix} \in \mathbb{C}^{(L+1)\times(M-L+1)}.$$

Denote the noiseless measurement vector by $y^0 = \Phi(\Omega, M)x$. For simplicity, we will denote $\Phi(\Omega, M)$ by Φ_M in this section. It is straightforward to verify that $\mathcal{H}(y^0)$ possesses the following Vandermonde decomposition:

$$\mathcal{H}(y^0) = \Phi_L X \Phi_{M-L}^T,$$

where $X = \mathrm{diag}(x) \in \mathbb{R}^{S\times S}$. We always assume that $M+1 \ge 2S$ and $S \le L \le M-S+1$ so that Φ_L and Φ_{M-L} have full column rank, and $\mathcal{H}(y^0)$ has rank S.

Algorithm 1 MUltiple SIgnal Classification (MUSIC).

Input: $y \in \mathbb{C}^{M+1}$, sparsity S and a integer L

1: Form Hankel matrix $\mathcal{H}(y) \in \mathbb{C}^{(L+1)\times(M-L+1)}$

2: Compute the SVD of $\mathcal{H}(y)$:

$$\mathcal{H}(y) = [\underbrace{\widehat{U}}_{(L+1)\times S} \quad \underbrace{\widehat{W}}_{(L+1)\times(L-S)}] \underbrace{\operatorname{diag}(\widehat{\sigma}_1, \ldots, \widehat{\sigma}_S, \widehat{\sigma}_{S+1}, \ldots)}_{(L+1)\times(M-L+1)} [\underbrace{\widehat{V}_1}_{(M-L+1)\times S} \quad \underbrace{\widehat{V}_2}_{(M-L+1)\times(M-L+1-S)}]^* \tag{4.1}$$

where $\widehat{\sigma}_1 \geq \widehat{\sigma}_2 \ldots \geq \widehat{\sigma}_S \geq \widehat{\sigma}_{S+1} \geq \ldots$ are the singular values of $\mathcal{H}(y)$.

3: Compute the imaging function $\widehat{\mathcal{J}}(\omega) = \|\phi_L(\omega)\|_2 / \|\widehat{W}^* \phi_L(\omega)\|_2,\ \omega \in [0, 1)$.

Output: $\widehat{\Omega} = \{\widehat{\omega}_j\}_{j=1}^S$ corresponding to the S largest local maxima of $\widehat{\mathcal{J}}$.

Table 1
Noise-space correlation functions and imaging functions in MUSIC, where W given in (4.2) and $\widehat{W}$ given in (4.1) of Algorithm 1 are the noise space of dimension $L - S$ in the noiseless and noisy case respectively.

	Noise-space correlation function	Imaging function
Noiseless case	$\mathcal{R}(\omega) = \frac{\|W^* \phi_L(\omega)\|_2}{\|\phi_L(\omega)\|_2}$	$\mathcal{J}(\omega) = \frac{1}{\mathcal{R}(\omega)} = \frac{\|\phi_L(\omega)\|_2}{\|W^* \phi_L(\omega)\|_2}$
Noisy case	$\widehat{\mathcal{R}}(\omega) = \frac{\|\widehat{W}^* \phi_L(\omega)\|_2}{\|\phi_L(\omega)\|_2}$	$\widehat{\mathcal{J}}(\omega) = \frac{1}{\widehat{\mathcal{R}}(\omega)} = \frac{\|\phi_L(\omega)\|_2}{\|\widehat{W}^* \phi_L(\omega)\|_2}$

In the noiseless case, let the Singular Value Decomposition (SVD) of $\mathcal{H}(y^0)$ be:

$$\mathcal{H}(y^0) = [\underbrace{U}_{L\times S} \quad \underbrace{W}_{L\times(L-S)}] \underbrace{\operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_S, 0, \ldots, 0)}_{L\times(M-L+1)} [\underbrace{V_1}_{(M-L+1)\times S} \quad \underbrace{V_2}_{(M-L+1)\times(M-L+1-S)}]^* \tag{4.2}$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_S$ are the non-zero singular values of $\mathcal{H}(y^0)$. The column spaces of U and W are the same as $\operatorname{Range}(\mathcal{H}(y^0))$ and $\operatorname{Range}(\mathcal{H}(y^0))^\perp$, which are called the signal space and the noise space respectively.

For any $\omega \in [0, 1)$ and integer L, we define the steering vector of length $L + 1$ at ω to be

$$\phi_L(\omega) = [1 \ e^{-2\pi i\omega} \ e^{-2\pi i 2\omega} \ \ldots \ e^{-2\pi i L\omega}]^T \in \mathbb{R}^{L+1},$$

and then $\Phi_L = [\phi_L(\omega_1) \ \ldots \ \phi_L(\omega_S)]$. The MUSIC algorithm is based on the following observation on the Vandermonde structure of Φ_L: if $S \leq L \leq M - S + 1$, then

$$\omega \in \{\omega_j\}_{j=1}^S \Longleftrightarrow \phi_L(\omega) \in \operatorname{Range}(\Phi_L) = \operatorname{Range}(\mathcal{H}(y^0)) = \operatorname{Range}(U). \tag{4.3}$$

We define a noise-space correlation function $\mathcal{R}(\omega)$ and an imaging function as its reciprocal (see Table 1 for definitions). The following lemma is based on (4.3).

Lemma 4.1. *Suppose $M + 1 \geq 2S$ and L is chosen such that $S \leq L \leq M - S + 1$. Then*

$$\omega \in \{\omega_j\}_{j=1}^S \Longleftrightarrow \mathcal{R}(\omega) = 0 \Longleftrightarrow \mathcal{J}(\omega) = \infty.$$

In the noiseless case, the source locations can be exactly identified through the zeros of the noise-space correlation function $\mathcal{R}(\omega)$ or the peaks of the imaging function $\mathcal{J}(\omega)$, as long as the number of measurements is at least twice the number of point sources to be recovered.

In the presence of noise, $\mathcal{H}(y^0)$ is perturbed to $\mathcal{H}(y)$ such that:

$$\mathcal{H}(y) = \mathcal{H}(y^0) + \mathcal{H}(\eta)$$

whose SVD is given by (4.1) in Algorithm 1. The noise-space correlation function and the imaging function are perturbed to $\widehat{\mathcal{R}}$ and $\widehat{\mathcal{J}}$ respectively (see Table 1 for the definitions). MUSIC in Algorithm 1 gives rise to a recovered support $\widehat{\Omega} = \{\widehat{\omega}_j\}_{j=1}^S$ corresponding to the S largest local maxima of $\widehat{\mathcal{J}}$.

4.2. Sensitivity of the noise-space correlation function in MUSIC

When the noise-to-signal ratio is low, the imaging function $\widehat{\mathcal{J}}$ still peaks around the true sources, but MUSIC can fail when the noise-to-signal ratio increases. The stability of MUSIC depends on the perturbation of the noise-space correlation function from $\mathcal{R}$ to $\widehat{\mathcal{R}}$ which is measured by $\|\widehat{\mathcal{R}} - \mathcal{R}\|_\infty := \max_{\omega\in[0,1)} |\widehat{\mathcal{R}}(\omega) - \mathcal{R}(\omega)|$. Thanks to the classical perturbation theory on singular subspaces by Wedin [68, Section 3], we have the following bound for $\|\widehat{\mathcal{R}} - \mathcal{R}\|_\infty$:

Proposition 4.2. *Suppose $M + 1 \geq 2S$ and L is chosen such that $S \leq L \leq M - S + 1$. If $2\|\mathcal{H}(\eta)\|_2 < x_{\min}\sigma_{\min}(\Phi_L)\sigma_{\min}(\Phi_{M-L})$, then*

$$\|\widehat{\mathcal{R}} - \mathcal{R}\|_\infty \leq \frac{2\|\mathcal{H}(\eta)\|_2}{x_{\min}\sigma_{\min}(\Phi_L)\sigma_{\min}(\Phi_{M-L})}. \tag{4.4}$$

The dependence on $\sigma_{\min}(\Phi_L)$ and $\sigma_{\min}(\Phi_{M-L})$ in (4.4) are crucial since they are small in the super-resolution regime. It is the best to set $L = \lfloor M/2 \rfloor$ to balance them. With Wedin's theorem, Proposition 4.2 improves Theorem 3 in [26], which upper bound $\|\widehat{\mathcal{R}} - \mathcal{R}\|_\infty$ in the order of $\|\mathcal{H}(\eta)\|_2/[\sigma^2_{\min}(\Phi_L)\sigma^2_{\min}(\Phi_{M-L})]$. Combining Proposition 4.2 and Theorem 2.7 gives rise to the following sensitivity bound of the noise-space correlation function under the clumps model:

Theorem 4.3. *Fix positive integers A, M, S. Let $M \geq S^2$ be an even integer and $L = M/2$. Suppose Ω satisfies Assumption 2.6 with parameters (A, L, S, β, α) for some $\alpha > 0$ and β with (2.4). For each $1 \leq a \leq A$, let $c_a = C_a(\lambda_a, L)$. If $4\|\mathcal{H}(\eta)\|_2 \sum_{a=1}^A (c_a \alpha^{-\lambda_a+1})^2 < x_{\min} M$, then*

$$\|\widehat{\mathcal{R}} - \mathcal{R}\|_\infty \leq \frac{4\|\mathcal{H}(\eta)\|_2}{x_{\min} M} \sum_{a=1}^A \left(c_a\, \alpha^{-\lambda_a+1}\right)^2.$$

Remark 4.4. In this paper, the stability of MUSIC is given in terms of the L^∞ perturbation of the noise-space correlation function. When the point sources are well separated, a source localization error $|\widehat{\omega}_j - \omega_j|$ was derived in terms of $\|\widehat{\mathcal{R}} - \mathcal{R}\|_\infty$ [26, Theorem 4]. In the clumps model, such a derivation is more complicated and we leave it as a future work.

Theorem 4.3 is stated for even M only for simplicity, and a similar result holds for odd M by setting $L = \lfloor M/2 \rfloor$. This theorem applies to any noise vector η. In the case of bounded noise, applying the inequality $\|\mathcal{H}(\eta)\|_2 \leq \|\mathcal{H}(\eta)\|_F \leq \sqrt{L}\|\eta\|_2$ gives rise to the following corollary:

Corollary 4.5. *Under the assumptions in Theorem 4.3,*

$$\|\widehat{\mathcal{R}} - \mathcal{R}\|_\infty \leq \frac{4\sqrt{L}\|\eta\|_2}{x_{\min} M} \sum_{a=1}^A \left(c_a\, \alpha^{-\lambda_a+1}\right)^2.$$

If η is independent gaussian noise, i.e., $\eta \sim \mathcal{N}(0, \sigma^2 I)$, the spectral norm of $\mathcal{H}(\eta)$ has been well studied in literature [69,70]. The following lemma [27, Theorem 4] is obtained from the matrix Bernstein inequality [71]:

Lemma 4.6. *If $\eta \sim \mathcal{N}(0, \sigma^2 I)$, then*

$$\mathbb{P}\left\{\|\mathcal{H}(\eta)\|_2 \geq t\right\} \leq (M+2)\exp\left(-\frac{t^2}{2\sigma^2 \max(L+1, M-L+1)}\right), \forall t > 0.$$

Combining Theorem 4.3 and Lemma 4.6 gives rise to the following explicit bound on the noise-to-signal ratio $\sigma/x_{\min}$ in order to guarantee a fixed ε-perturbation of $\mathcal{R}$.

Corollary 4.7. *Suppose $\eta \sim \mathcal{N}(0, \sigma^2 I)$. Fix $\varepsilon > 0$ and $\nu > 1$. Under the assumptions in Theorem 4.3, if*

$$\frac{\sigma}{x_{\min}} < \frac{M}{4\sqrt{\nu(M+2)\log(M+2)}} \left(\sum_{a=1}^{A} \left(c_a \alpha^{-\lambda_a+1}\right)^2\right)^{-1} \varepsilon, \tag{4.5}$$

then $\|\widehat{\mathcal{R}} - \mathcal{R}\|_\infty \leq \varepsilon$ with probability no less than $1 - (M+2)^{-(\nu-1)}$.

Corollary 4.7 is proved in Appendix A.5. It upper bounds the noise-to-signal ratio for which MUSIC can guarantee an ε-perturbation of $\mathcal{R}$. In the special case where each Λ_a contains λ equally spaced points with spacing α/M (see Fig. 2), (4.5) can be simplified to

$$\frac{\sigma}{x_{\min}} \propto \varepsilon \sqrt{\frac{M}{\log M}} \alpha^{2\lambda-2} = \varepsilon \sqrt{\frac{M}{\log M}} \left(\frac{1}{\mathrm{SRF}}\right)^{2\lambda-2}, \tag{4.6}$$

which shows that the noise-to-signal ratio that MUSIC can tolerate is exponential in 1/SRF. The key contribution of this paper is that, the exponent only depends on the cardinality of the clumps instead of the total sparsity S. These estimates are verified by numerical experiments in Section 4.3.

4.3. Numerical simulations on the super-resolution limit of MUSIC

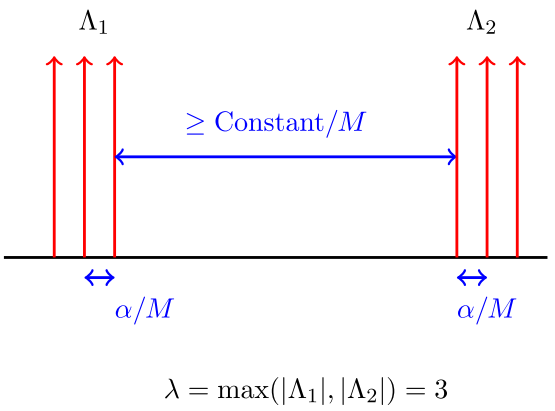
In our experiments, the true support Ω consists of $A = 1, 2, 3, 4$ clumps and each clump contains λ equally spaced point sources separated by $\Delta = \alpha/M$ where $1/\alpha$ is the SRF of Ω (see Fig. 5 (a) for an example). The clumps are separated at least by $10/M$. The x_i's are complex with unit magnitudes and random phases. Noise is gaussian: $\eta \sim \mathcal{N}(0, \sigma^2 I)$. We set $M = 100$ and let Δ vary so that the SRF varies from 1 to 10. We run MUSIC with the varying SRF and σ for 10 trials. The support error is measured by the matching distance between Ω and $\widehat{\Omega}$:

$$\mathrm{dist}_B(\Omega, \widehat{\Omega}) := \inf_{\text{bijection } \psi\colon\ \Omega \to \widehat{\Omega}} \sup_{\widehat{\omega} \in \widehat{\Omega}} |\widehat{\omega} - \psi(\omega)|_{\mathbb{T}}\,.$$

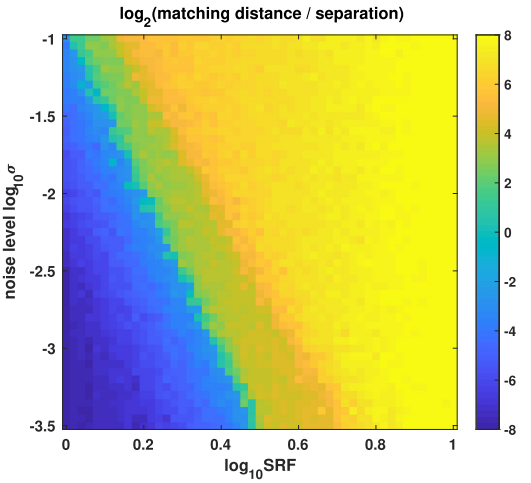
Fig. 5 (b) displays the average $\log_2[\mathrm{dist}_B(\Omega, \widehat{\Omega})/\Delta]$ over 10 trials with respect to $\log_{10}\mathrm{SRF}$ (x-axis) and $\log_{10}\sigma$ (y-axis) when Ω contains 2 clumps of 3 equally spaced point sources: $A = 2, \lambda = 3$. A clear phase transition demonstrates that MUSIC is capable of resolving closely spaced *complex-valued* objects as long as σ is below certain threshold.

In Fig. 6, we display the phase transition curves at which $\mathrm{dist}_B(\Omega, \widehat{\Omega}) \approx \Delta/2$ with respect to $\log_{10}\mathrm{SRF}$ (x-axis) and $\log_{10}\sigma$ (y-axis) when Ω contains $A = 1, 2, 3, 4$ clumps of $\lambda = 2, 3, 4, 5$ equally spaced point sources. All phase transition curves are almost straight lines, manifesting that the noise level σ that MUSIC can tolerate satisfies

$$\sigma \propto \mathrm{SRF}^{-q(\Omega)}. \tag{4.7}$$



(a) 2 clumps of 3 equally spaced point sources separated by $\Delta = \alpha/M$



(b) Phase transition of MUSIC

Fig. 5. (b) displays the average $\log_2[\mathrm{dist}_B(\Omega, \widehat{\Omega})/\Delta]$ over 10 trials with respect to $\log_{10}\mathrm{SRF}$ (x-axis) and $\log_{10}\sigma$ (y-axis) when $A = 2$ and $\lambda = 3$.

Table 2
Numerical simulations of $q(\Omega)$ in (4.7) on the super-resolution limit of MUSIC.

	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	Numerical $q(\Omega)$	Theoretical $q(\Omega)$
1-clump: $A = 1$	3.0019	5.1935	7.4176	10.4286	$2.45\lambda - 2.07$	$2\lambda - 2$
2-clump: $A = 2$	3.1287	5.2717	7.9371	10	$2.33\lambda - 1.56$	$2\lambda - 2$
3-clump: $A = 3$	3.1081	5.1826	7.299	8.5	$1.83\lambda - 0.38$	$2\lambda - 2$
4-clump: $A = 4$	3.0767	5.1731	7.3252	10	$2.29\lambda - 1.63$	$2\lambda - 2$

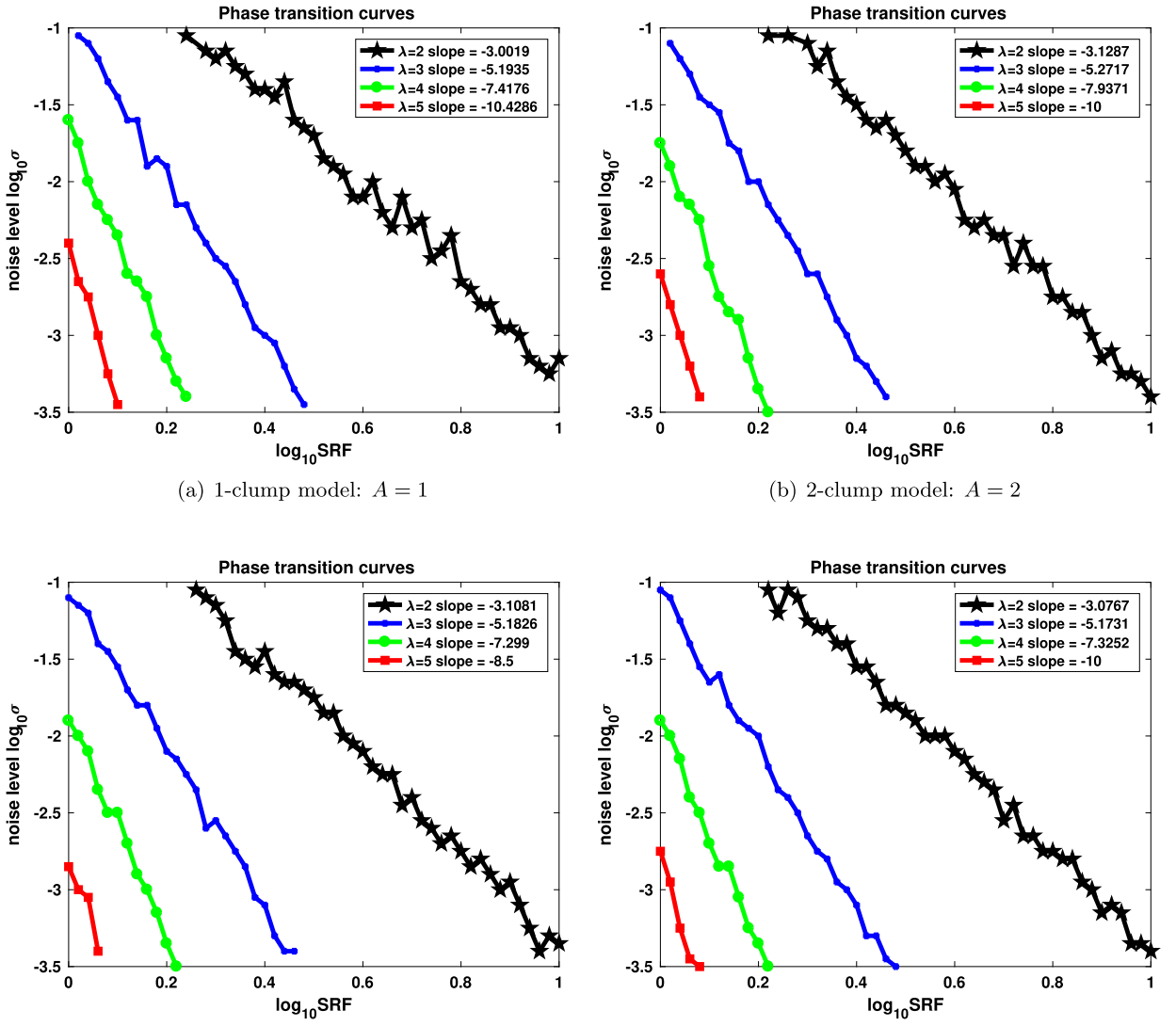
A least squares fitting of the curves by straight lines gives rise to the exponent $q(\Omega)$ numerically, summarized in Table 2. It is close to our theory (4.6) of $q(\Omega) = 2\lambda - 2$.

4.4. Related work on subspace methods

Subspace methods, including MUSIC [5], ESPRIT [13] and MPM [14], were initially proposed for the Direction-Of-Arrival (DOA) estimation [4]. In the DOA setting, the amplitude vector x is random with respect to time, and multiple snapshots of measurements are taken. The covariance matrix of y possesses a Vandermonde decomposition so that MUSIC, ESPRIT and MPM are applicable. The classical theories in [72,73] primarily analyze the stability of MUSIC and ESPRIT with respect to the number of snapshots, denoted by #Snapshot. They show that the asymptotic distribution of the squared error is on the order of $C \times \mathrm{noise}/\#\mathrm{Snapshot}$ where the constant C depends on $\sigma_{\min}(\Phi)$. It is not clear from [72,73] how large the implicit constant C is, while our paper gives an explicit characterization of the C under the clumps model.

The super-resolution problem considered in this paper corresponds to the single snapshot case, in which case, the key quantity of interest is the implicit constant C. In the single-snapshot case, there have been several works on the stability of subspace methods, which addressed the connection between $\sigma_{\min}(\Phi)$ and the support Ω. These works include MUSIC [26], ESPRIT [74] and MPM [28], but only apply to the well-separated case. During the review period of this paper, the present authors applied Theorem 2.7 to derive an error bound for ESPRIT in the super-resolution regime [75].

In literature, the resolution of MUSIC was addressed in [41–43]. These papers studied the frequency estimation error in the DOA setting with multiple snapshots of measurements. The Cramér-Rao lower bound is derived based on the statistics of random amplitudes. Thus, these results do not apply to the



(a) 1-clump model: $A = 1$ (b) 2-clump model: $A = 2$

(c) 3-clump model: $A = 3$ (d) 4-clump model: $A = 4$

Fig. 6. Phase transition curves: (a-d) show the average $\log_2[\mathrm{dist}_B(\Omega, \widehat{\Omega})/\Delta]$ over 10 trials with respect to $\log_{10}\mathrm{SRF}$ (x-axis) and $\log_{10}\sigma$ (y-axis) when Ω contains $A = 1, 2, 3, 4$ clumps of $\lambda = 2, 3, 4, 5$ equally spaced point sources. The slope is computed by a least squares fitting by a straight line.

single-snapshot case. It was explicitly mentioned in [41, after Equation (1.2)] that, "In effect, the analysis in [42] assumes the multi-experiment case, and hence may not apply to the single experiment case." Our theory shares some similarity with [42]. Equation (2) in [42] implies that, to guarantee the success of MUSIC for the estimation of λ closely spaced frequencies with separation Δ, the SNR needs to be at least $\Delta^{-2(\lambda-1)}$, which has the same dependence on Δ as the Equation (4.6) of our paper. Here are some differences between Equation (2) in [42] and Equation (4.6) of our paper: (1) Our Equation (4.6) implies SNR in the order of $\mathrm{SRF}^{2(\lambda-1)} = (M\Delta)^{-2(\lambda-1)}$, which is more accurate than $\Delta^{-2(\lambda-1)}$. In other words, our theory takes the advantage of a large M in applications. (2) Our theory does not rely on the randomness of x, and considers the clumps model, which is more general than just having λ closely spaced frequencies with separation Δ. (3) Our results are non-asymptotic and work for finite measurements M, while the theories in [41–43] are asymptotic and require #Snapshot $\to \infty$ and $\Delta \to 0$.

Finally we remark that the results in the present paper are not the same as the Cramér-Rao bounds for a single-snapshot case. In the single snapshot case, one assumes certain statistics of the random noise, and the

Cramér-Rao bound expresses a lower bound on the variance of unbiased estimators. The noise statistics play an important role in the Cramér-Rao bound. Our perturbation bound for MUSIC, such as Corollary 4.5, is deterministic and can be applied for all noise vectors, independently of the noise statistics.

Acknowledgments

Both authors express their sincere gratitude to Professors John Benedetto, Sinan Güntürk, and Kasso Okoudjou for valuable advice and insightful discussions. Weilin Li was supported by the Defense Threat Reduction Agency Grant 1-13-1-0015, and the Army Research Office Grants W911 NF-17-1-0014 and W911 NF-16-1-0008. He was also supported by The National Science Foundation Grant DMS-1440140 during the Spring 2017 semester while in residence at the Mathematical Sciences Research Institute in Berkeley, California, and by the University of Maryland's Ann G. Wylie Dissertation fellowship during the Spring 2018 semester. Wenjing Liao was supported by NSF DMS 1818751, a startup fund from Georgia Institute of Technology, and the AMS-Simons Travel Grants.

Appendix A. Proof of theorems and corollaries

A.1. Proof of Theorem 2.3

We first need to introduce the following function that is of great importance in Fourier analysis. For a positive integer P, we define the normalized *Fejér kernel* $F_P \in C^\infty(\mathbb{T})$ by the formula,

$$F_P(\omega) := \frac{1}{P+1} \sum_{m=-P}^{P} \Big(1 - \frac{|m|}{P+1}\Big) e^{2\pi i m\omega} = \frac{1}{(P+1)^2} \Big(\frac{\sin(\pi(P+1)\omega)}{\sin(\pi\omega)}\Big)^2.$$

The normalization is chosen so that $F_P(0) = 1$. We recall some basic facts about the Fejér kernel. Its $L^2(\mathbb{T})$ norm can be calculated using Parseval's formula, and so

$$\|F_P\|_{L^2(\mathbb{T})} = \frac{1}{(P+1)^2} \Big((P+1)^2 + 2\sum_{m=1}^{P} m^2\Big)^{1/2} \le \frac{1}{(P+1)^{1/2}}. \tag{A.1}$$

We can also provide a point-wise estimate. By the trigonometric inequality $|\sin(\pi\omega)| \ge 2|\omega|_{\mathbb{T}}$, we have

$$|F_P(\omega)| \le \frac{1}{2^2(P+1)^2|\omega|_{\mathbb{T}}^2}, \quad \text{for all} \quad \omega \in \mathbb{T}. \tag{A.2}$$

If we raise the Fejér kernel to a power R, then the function $(F_P(\omega))^R$ has better decay, but at the cost of increasing its frequency support. If we keep the product PR fixed, then increasing R leads to better decay at the expense of worse localization near the origin.

The proof of Theorem 2.3 relies on the quantitative properties of a set of polynomials $\{I_j\}_{j=1}^S$, with I_j depending on Ω and M, which we shall explicitly construct. The construction seems complicated, but the idea is very simple. For each $\omega_j \in \Omega$, we construct $I_j \in \mathcal{P}(M)$ such that it decays rapidly away from ω_j and

$$I_j(\omega_k) = \delta_{j,k}, \quad \text{for all} \quad \omega_k \in \Lambda_a.$$

The key is to carefully construct each I_j so that it has small norm; otherwise, the resulting lower bound for $\sigma_{\min}(\Phi)$ would be loose and have limited applicability. The construction of these polynomials is technical and it can be found in Appendix C.

Lemma A.1. *Suppose the assumptions of Theorem 2.3 hold and define the constant*

$$\widetilde{B}_a := \widetilde{B}_a(\lambda_a, M) := \Big(1 - \frac{\pi^2}{3\lambda_a^2}\Big)^{-(\lambda_a-1)/2} \Big(\frac{M}{\lambda_a}\Big)^{\lambda_a-1} \Big\lfloor \frac{M}{\lambda_a} \Big\rfloor^{-(\lambda_a-1)}.$$

For each $1 \le a \le A$ and each $\omega_j \in \Lambda_a$, there exists a $I_j \in \mathcal{P}(M)$ satisfying the following properties.

(a) $I_j(\omega_k) = \delta_{j,k}$ *for all* $\omega_k \in \Lambda_a$.
(b) $|I_j(\omega_k)| \le 1/(20S)$ *for all* $\omega_k \notin \Lambda_a$.
(c) $\|I_j\|_{L^2(\mathbb{T})} \le (2/M)^{1/2} \widetilde{B}_a \lambda_a^{\lambda_a - 1} \rho_j$.

Proof of Theorem 2.3. Let $\{I_j\}_{j=1}^S$ be the polynomials constructed in Lemma A.1. Let $v \in \mathbb{C}^S$ be a unit norm vector such that

$$\sigma_{\min}(\Phi) = \|\Phi v\|_2.$$

We define the trigonometric polynomial $I \in \mathcal{P}(M)$ by the formula,

$$I(\omega) := I(\omega, v) := \sum_{j=1}^S v_j I_j(\omega).$$

For each index $1 \le k \le S$, we define the quantity

$$\varepsilon_k := I(\omega_k) - v_k.$$

Since $I_j(\omega_j) = 1$, we have

$$\varepsilon_k = \sum_{j \ne k} v_j I_j(\omega_k).$$

Fix a $\omega_k \in \Omega$. Then $\omega_k \in \Lambda_a$ for some $1 \le a \le A$. By Cauchy Schwartz, the assumption that v is unit norm, the property that $I_j(\omega_k) = \delta_{j,k}$ for all $\omega_k \in \Lambda_a$, and the upper bound on $|I(\omega_k)|$ given in Lemma A.1, we deduce

$$|\varepsilon_k| \le \Big(\sum_{j \ne k} |I_j(\omega_k)|^2\Big)^{1/2} = \Big(\sum_{\omega_j \notin \Lambda_a} |I_j(\omega_k)|^2\Big)^{1/2} \le \frac{1}{20\sqrt{S}}.$$

This holds for each $1 \le k \le S$, so we have

$$\|\varepsilon\|_2 \le \sqrt{S}\|\varepsilon\|_\infty \le \frac{1}{20}.$$

The conditions of robust duality, Proposition 2.13, are satisfied, so we have

$$\sigma_{\min}(\Phi) = \|\Phi v\|_2 \ge \frac{19}{20} \|I(\cdot, v)\|_{L^2(\mathbb{T})}^{-1}.$$

To complete the proof, we need to upper bound $\|I(\cdot, v)\|_{L^2(\mathbb{T})}$ uniformly in v. We use Cauchy-Schwartz, that v has unit norm, and the norm bound for I_j given in Lemma A.1 to obtain the upper bound,

$$\|I(\cdot, v)\|_{L^2(\mathbb{T})} \le \Big(\sum_{j=1}^S \|I_j\|_{L^2(\mathbb{T})}^2\Big)^{1/2} \le \Big(\frac{2}{M}\Big)^{1/2} \Big(\sum_{j=1}^S (\widetilde{B}_a \lambda_a^{\lambda_a} \rho_j)^2\Big)^{1/2},$$

where $\widetilde{B}_a$ was defined in the referenced lemma. Combining the previous two inequalities shows that

$$\sigma_{\min}(\Phi) \geq \frac{19}{20\sqrt{2}}\sqrt{M}\Big(\sum_{j=1}^{S}(\widetilde{B}_a^2\lambda_a^{\lambda_a}\rho_j)^2\Big)^{-1/2} = \sqrt{M}\Big(\sum_{j=1}^{S}(B_a\lambda_a^{\lambda_a}\rho_j)^2\Big)^{-1/2},$$

where by definition, $B_a = \widetilde{B}_a(20\sqrt{2})/19$. □

A.2. Proof of Theorem 2.7

Proof. Fix an index $1 \leq a \leq A$ and $\omega_j \in \Lambda_a$. Recalling the definition of ρ_j and using that $\Delta \geq \alpha/M$, we see that

$$\rho_j = \prod_{\omega_k \in \Lambda_a \setminus \{\omega_j\}} \frac{1}{\pi M|\omega_k - \omega_j|_{\mathbb{T}}} \leq \Big(\frac{1}{\pi\alpha}\Big)^{\lambda_a - 1}.$$

This implies that

$$\frac{10\lambda_a^{5/2}(S\rho_j)^{1/(2\lambda_a)}}{M} \leq \frac{10\lambda_a^{5/2}S^{1/2}}{M\alpha^{1/2}}.$$

This in turn, shows that the separation condition (2.4) implies (2.2). Hence, the assumptions of Theorem 2.3 are satisfied, and we have

$$\sigma_{\min}(\Phi) \geq \sqrt{M}\Big(\sum_{a=1}^{A}\sum_{\omega_j\in\Lambda_a}(B_a\lambda_a^{\lambda_a-1}\rho_j)^2\Big)^{-1/2}.$$

We can write the right hand side in terms of α. Observe that if $\widetilde{\Lambda}_a = \{\widetilde{\omega}_j\}_{j=1}^{\lambda_a}$ contains λ_a points that are equispaced by α/M and $\widetilde{\rho}_j$ is the complexity of $\widetilde{\omega}_j$, then

$$\sum_{\omega_j\in\Lambda_a}\rho_j^2 \leq \sum_{\widetilde{\omega}_j\in\widetilde{\Lambda}_a}\widetilde{\rho}_j^2.$$

Thus we have the inequality,

$$\sum_{\omega_j\in\Lambda_a}\rho_j^2 \leq \sum_{j=1}^{\lambda_a}\Big(\prod_{k=1,\ k\neq j}^{\lambda_a}\frac{1}{(j-k)^2}\Big)\Big(\frac{1}{\pi\alpha}\Big)^{2\lambda_a-2}.$$

Combining the above inequalities completes the proof. □

A.3. Proof of Theorem 3.5

The crux of the proof is to construct, for each $\Omega \subseteq \{n/N\}_{n=0}^{N-1}$ with $|\Omega| = S$, a family of polynomials $\{H_j(\cdot,\Omega)\}_{j=1}^{S}$ with small $L^2(\mathbb{T})$ norms that satisfy an appropriate interpolation property. The construction is technical because it must be done carefully in order to obtain an accurate bound for the lower restricted isometry constant. The proof of the following lemma can be found in Appendix C.

Lemma A.2. *Suppose the assumptions of Theorem 3.5 hold and let $C(M,S)$ be the constant defined in the theorem. For each $\Omega \subseteq \{n/N\}_{n=0}^{N-1}$ and of cardinality S, there exist a family of polynomials $\{H_j(\cdot,\Omega)\}_{j=1}^{S} \subseteq \mathcal{P}(M)$ such that*

$$H_j(\omega_k, \Omega) = \delta_{j,k} \quad \textit{for all} \quad \omega_j, \omega_k \in \Omega.$$

Moreover, we have the upper bound,

$$\Big(\sum_{j=1}^{S} \|H_j(\cdot, \Omega)\|^2_{L^2(\mathbb{T})}\Big)^{1/2} \leq C(M,S)^{-1} \frac{1}{\sqrt{M}} \Big(\frac{N}{M}\Big)^{S-1}.$$

Proof of Theorem 3.5. By definition of the lower restricted isometry constant, there exists a set Ω of cardinality S and supported on the grid with spacing $1/N$ such that

$$\Theta(M,N,S) = \sigma_{\min}(\Phi(\Omega, M)).$$

Let $\{H_j(\Omega)\}_{j=1}^S$ be the family of polynomials given in Lemma A.2. Let $u = u(\Omega) \in \mathbb{C}^S$ be a unit norm vector such that

$$\sigma_{\min}(\Phi(\Omega, M)) = \|\Phi(\Omega, M)u\|_2.$$

We define the polynomial,

$$H(\omega) := H(\omega, u, \Omega) := \sum_{j=1}^{S} u_j H_j(\omega, \Omega).$$

Using the interpolation property of $\{H_j(\cdot, \Omega)\}_{j=1}^S$ guaranteed by Lemma A.2, we see that $H \in \mathcal{P}(\Omega, M, u)$. By exact duality, Proposition 2.12, we have

$$\sigma_{\min}(\Phi(\Omega, M)) = \max_{f \in \mathcal{P}(\Omega, M, u(\Omega))} \|f\|^{-1}_{L^2(\mathbb{T})} \geq \|H(\cdot, \Omega)\|^{-1}_{L^2(\mathbb{T})}.$$

Using Cauchy-Schwartz and that u is a unit norm vector, we have

$$\|H\|_{L^2(\mathbb{T})} \leq \Big(\sum_{j=1}^{S} \|H_j\|^2_{L^2(\mathbb{T})}\Big)^{1/2}.$$

Combining the previous inequalities and using the upper bound given in Lemma A.2 completes the proof of the theorem. □

A.4. Proof of Theorem 3.8

Proof. The upper bound for the min-max error is a direct consequence of Proposition 3.4 and Theorem 3.5. To obtain a lower bound for the min-max error, we first apply Proposition 2.10 to the case that Ω consists of $2S$ consecutive points spaced by $1/N$. We ready check that the size assumptions on M and N imply that the conditions of Proposition 2.10 are satisfied, and thus,

$$\Theta(M,N,2S) \leq 2\binom{4S-2}{2S-1}^{-1/2} \sqrt{M+1} \left(\frac{2\pi M}{N}\right)^{2S-1}.$$

Combining this with Proposition 3.4 establishes a lower bound for the min-max error. □

A.5. Proof of Corollary 4.7

Proof. According to Theorem 4.3, a sufficient condition for $\|\widehat{\mathcal{R}} - \mathcal{R}\|_\infty \leq \varepsilon$ is

$$\|\mathcal{H}(\eta)\|_2 \leq M x_{\min} \Big(\sum_{a=1}^{A} (c_a \alpha^{-\lambda_a+1})^2 \Big)^{-1} \varepsilon/4. \tag{A.3}$$

Lemma 4.6 implies that (A.3) holds with probability no less than $1-(M+2)^{-(\nu-1)}$ as long as $t = M x_{\min} \Big(\sum_{a=1}^{A} c_a^2 \alpha^{-2(\lambda_a-1)} \Big)^{-1} \varepsilon/4$ and $(M+2) e^{-\frac{t^2}{\sigma^2 (M+2)}} < (M+2)^{1-\nu}$, which is guaranteed by (4.5). □

Appendix B. Proof of propositions

B.1. Proof of Proposition 2.10

Proof. The argument relies on the variational form for the minimum singular value,

$$\sigma_{\min}(\Phi) = \min_{u \in \mathbb{C}^S, u \neq 0} \frac{\|\Phi u\|_2}{\|u\|_2}.$$

To obtain an upper bound, it suffices to consider a specific u, and our choice is inspired by Donoho [37]. Without loss of generality, we assume that $\omega = 0$. We re-index the set $\Omega = \{\omega_j\}_{j=1}^S$ so that

$$\omega_j = \frac{(j-1)\alpha}{M} \quad \text{for} \quad 1 \leq j \leq \lambda.$$

We consider the vector $u \in \mathbb{C}^S$ defined by the formula

$$u_j := (-1)^{j-1} \binom{\lambda - 1}{j-1} \quad \text{for} \quad 1 \leq j \leq \lambda,$$

and $u_j = 0$ otherwise. Note that

$$\|u\|_2 = \binom{2\lambda - 2}{\lambda - 1}^{1/2}.$$

By the variational form for the minimum singular value, we have

$$\sigma_{\min}(\Phi) \leq \frac{\|\Phi u\|_2}{\|u\|_2} = \binom{2\lambda - 2}{\lambda - 1}^{-1/2} \|\Phi u\|_2. \tag{B.1}$$

To estimate $\|\Phi u\|_2$, we identify u with the discrete measure

$$\mu := \sum_{j=1}^{\lambda} u_j \delta_{(j-1)\alpha/M}.$$

We also define a modulated Dirichlet kernel $D_M \in C^\infty(\mathbb{T})$ by the formula, $D_M(\omega) := \sum_{m=0}^{M} e^{2\pi i m \omega}$. We readily check that

$$\|\Phi u\|_2 = \sum_{m=0}^{M} |(\Phi u)_m|^2 = \Big(\sum_{m=0}^{M} |\widehat{\mu}(m)|^2 \Big)^{1/2} = \|\mu * D_M\|_{L^2(\mathbb{T})}. \tag{B.2}$$

We see that all $\omega \in \mathbb{T}$,

$$(\mu * D_M)(\omega) = \sum_{j=0}^{\lambda-1} (-1)^j \binom{\lambda-1}{j} D_M\Big(\omega - \frac{j\alpha}{M}\Big). \tag{B.3}$$

The right hand side is the $(\lambda-1)$-th order backwards finite difference of D_M. It is well-known that for each $\omega \in \mathbb{T}$, we have

$$(\mu * D_M)(\omega) = \Big(\frac{\alpha}{M}\Big)^{\lambda-1} D_M^{(\lambda-1)}(\omega) + R_{\lambda-1}(\omega), \tag{B.4}$$

where $D_M^{(\lambda-1)}$ denotes the $(\lambda-1)$-th derivative of D_M and the remainder term $R_{\lambda-1}$ in magnitude is pointwise $O((\alpha/M)^\lambda)$ as $\alpha \to 0$. In order to exactly determine how small we require α to be, we calculate the remainder term explicitly. By a Taylor expansion of D_M, for each $\omega \in \mathbb{T}$ and $0 \le j \le \lambda - 1$, there exists $\omega_j \in (\omega - j\alpha/M, \omega)$ such that

$$D_M(\omega - j\alpha) = \sum_{k=0}^{\lambda-1} D_M^{(k)}(\omega)\Big(\frac{\alpha}{M}\Big)^k \frac{(-1)^k j^k}{k!} + D_M^{(\lambda)}(\omega_j)\Big(\frac{\alpha}{M}\Big)^\lambda \frac{(-1)^\lambda j^\lambda}{\lambda!}.$$

Using this formula in equations (B.2) and (B.3), we see that

$$R_{\lambda-1}(\omega) = \sum_{j=0}^{\lambda-1} (-1)^{j+\lambda} \binom{\lambda-1}{j} D_M^{(\lambda)}(\omega_j)\Big(\frac{\alpha}{M}\Big)^\lambda \frac{j^\lambda}{\lambda!}.$$

We are ready to bound equation (B.4) in the $L^2(\mathbb{T})$ norm. By the Bernstein inequality for trigonometric polynomials, we have

$$\|D_M^{(\lambda-1)}\|_{L^2(\mathbb{T})} \le (2\pi M)^{\lambda-1} \|D_M\|_{L^2(\mathbb{T})} = \sqrt{M+1}\,(2\pi M)^{\lambda-1}.$$

By the same argument, we have

$$\begin{aligned}
\|R_{\lambda-1}\|_{L^2(\mathbb{T})} &\le \sum_{j=0}^{\lambda-1} \binom{\lambda-1}{j} \Big(\frac{\alpha}{M}\Big)^\lambda \frac{j^\lambda}{\lambda!} \|D_M^{(\lambda)}\|_{L^\infty(\mathbb{T})} \\
&\le C(\lambda)\alpha(2\pi\alpha)^{\lambda-1} \|D_M\|_{L^\infty(\mathbb{T})} \\
&\le C(\lambda)\alpha(2\pi\alpha)^{\lambda-1}(M+1).
\end{aligned}$$

Using these upper bounds together with (B.4), we have

$$\|\mu * D_M\|_{L^2(\mathbb{T})} \le \sqrt{M+1}\,(2\pi\alpha)^{\lambda-1}\Big(1 + C(\lambda)\alpha\sqrt{M+1}\Big).$$

This inequality and the assumed upper bound for α (2.9), we see that

$$\|\mu * D_M\|_{L^2(\mathbb{T})} \le 2\sqrt{M+1}\,(2\pi\alpha)^{\lambda-1}.$$

Combining this inequality with (B.1) and (B.2) completes the proof. □

B.2. Proof of Proposition 2.12

Proof. We first prove that $\mathcal{P}(\Omega, M, w)$ is non-empty for any $w \in \mathbb{C}^S$ and $S \leq M-1$. For each $1 \leq k \leq S$, we consider the Lagrange polynomials,

$$L_k(\omega) := \prod_{j \neq k} \frac{e^{2\pi i\omega} - e^{2\pi i\omega_j}}{e^{2\pi i\omega_k} - e^{2\pi i\omega_j}}.$$

We have $L_k(\omega_j) = \delta_{j,k}$ by definition, and after expanding L_k as a summation, we see that $L_k \in \mathcal{P}(S-1) \subseteq \mathcal{P}(M)$. This implies $\sum_{k=1}^S w_k L_k \in \mathcal{P}(\Omega, S-1, w)$, which proves the first part of the proposition.

Let v be any unit norm vector such that $\|\Phi v\|_2 = \sigma_{\min}(\Phi)$. The set of all trigonometric polynomials $f \in \mathcal{P}(M)$ can be written in the form

$$f(\omega) = \sum_{m=0}^{M-1} \widehat{f}(m) e^{2\pi i m\omega}.$$

Then $f \in \mathcal{P}(M, \Omega, v)$ if and only if $f \in \mathcal{P}(M)$ and it Fourier coefficients satisfy the under-determined system of equations,

$$v_j = \sum_{m=0}^{M-1} \widehat{f}(m) e^{2\pi i m\omega_j} \quad \text{for} \quad 1 \leq j \leq S.$$

Since $\mathcal{P}(\Omega, M, v)$ is non-empty, pick any $f \in \mathcal{P}(\Omega, M, v)$. Since $\|f\|_{L^2(\mathbb{T})} = \|\widehat{f}\|_{\ell^2(\mathbb{Z})}$, the functions $f \in \mathcal{P}(M)$ that satisfy this system of equations and have minimal $L^2(\mathbb{T})$ norm are the ones with Fourier coefficients given by the Moore-Penrose pseudo-inverse solution to the above system of equations. Namely,

$$\min_{f \in \mathcal{P}(M,\Omega,v)} \|f\|_{L^2(\mathbb{T})} = \min_{\Phi^* u = v} \|u\|_2 = \|(\Phi^*)^\dagger v\|_2 = \frac{1}{\sigma_{\min}(\Phi)}.$$

Rearranging this inequality completes the proof of the proposition. □

B.3. Proof of Proposition 2.13

Proof. Define the measure $\mu = \sum_{j=1}^S v_j \delta_{\omega_j}$, and note that $\widehat{\mu}(m) = (\Phi v)_m$. We have

$$\Big| \int_{\mathbb{T}} \overline{f}\, d\mu \Big| = \Big| \sum_{j=1}^S \overline{f(\omega_j)} v_j \Big| = \Big| \|v\|_2^2 + \sum_{j=1}^S \overline{v_j} \varepsilon_j \Big| \geq \|v\|_2^2 - \|v\|_2 \|\varepsilon\|_2 = 1 - \|\varepsilon\|_2.$$

On the other hand, using that $f \in \mathcal{P}(M)$, Cauchy-Schwartz, and Parseval,

$$\Big| \int_{\mathbb{T}} \overline{f}\, d\mu \Big| = \Big| \sum_{m=1}^{M-1} \overline{\widehat{f}(m)} \widehat{\mu}(m) \Big| \leq \|\widehat{f}\|_{\ell^2(\mathbb{Z})} \|\Phi v\|_2 = \|f\|_{L^2(\mathbb{T})} \|\Phi v\|_2.$$

Combining the previous two inequalities completes the proof. □

B.4. Proof of Proposition 3.4

For this proof, we make the following changes to the notation. We can identify every discrete measure μ whose support is contained in the grid with spacing $1/N$ and consists of S points with a S-sparse vector $x \in \mathbb{C}^N$. Under this identification, the Fourier transform of μ is identical to the discrete Fourier transform of x. Let $\mathbb{C}_S^N$ be the set of S-sparse vectors in $\mathbb{C}^N$, and $\mathcal{F}$ be the first $M+1$ rows of the $N \times N$ discrete Fourier transform matrix. With this notation at hand, the min-max error is

$$\mathcal{E}(M, N, S, \delta) = \inf_{\varphi \in \mathcal{A}} \sup_{y(x,\eta) \in \mathcal{Y}} \|\varphi_y - x\|_2.$$

Proof. We prove the upper bound first. Let φ be the function that maps each $y \in \mathcal{Y}$ to the sparsest vector $\varphi_y \in \mathbb{C}^N$ such that $\|\mathcal{F}\varphi_y - y\|_2 \leq \delta$. If there is not a unique choice of vector φ_y, just choose any one of them arbitrarily. Note that φ_y exists because x also satisfies the constraint that $\|\mathcal{F}x - y\|_2 \leq \delta$, and the choice of φ_y does not explicitly depend on x and η. Note that $\|\widetilde{x}\|_0 \leq \|\varphi_y\|_0 \leq S$ by definition of φ. Then we have

$$\mathcal{E}(M, N, S, \delta) \leq \sup_{y(x,\eta) \in \mathcal{Y}} \|\varphi_y - x\|_2.$$

For any $x \in \mathbb{C}_S^N$ and η with $\|\eta\|_2 \leq \delta$, we have $\varphi_y - x \in \mathbb{C}_{2S}^N$ and

$$\Theta(M, N, 2S) \leq \frac{\|\mathcal{F}(\varphi_y - x)\|_2}{\|\varphi_y - x\|_2} \leq \frac{\|\mathcal{F}\varphi_y - y\|_2 + \|\mathcal{F}x - y\|_2}{\|\varphi_y - x\|_2} \leq \frac{2\delta}{\|\varphi_y - x\|_2}.$$

Combining the previous two inequalities and rearranging completes the proof of the upper bound for the min-max error.

We focus our attention on the lower bound for the min-max error. By definition of the smallest singular value, there exists $v \in \mathbb{C}_{2S}^N$ of unit norm such that

$$\Theta(M, N, 2S) = \|\mathcal{F}v\|_2.$$

Pick any vectors $v_1, v_2 \in \mathbb{C}_S^N$ such that

$$\frac{\delta}{\Theta(M, N, 2S)}\, v = v_1 - v_2.$$

Suppose we are given the data

$$y = \mathcal{F}v_1 = \mathcal{F}v_2 + \mathcal{F}(v_1 - v_2).$$

Let $\eta := \mathcal{F}(v_1 - v_2) \in \mathbb{C}^{M+1}$. The previous three equations imply

$$\|\eta\|_2 = \|\mathcal{F}(v_1 - v_2)\|_2 = \frac{\delta}{\Theta(M, N, 2S)} \|\mathcal{F}v\|_2 \leq \delta.$$

This proves that y is both the noiseless first M Fourier coefficients of v_1 as well as the noisy first M Fourier coefficients of v_2 with noise $\mathcal{F}(v_1 - v_2)$ with noise η. Thus, we have $y \in \mathcal{Y}$ with $y = y(v_1, 0)$ and $y = y(v_2, \eta)$. Consequently, we have

$$\mathcal{E}(M, N, S, \delta) \geq \inf_{\varphi \in \mathcal{A}} \max_{k=1,2} \|f(y) - v_k\|.$$

Using that v has unit norm, for any $\varphi \in \mathcal{A}$, we have

$$\frac{\delta}{\Theta(M,N,2S)} = \|v_1 - v_2\|_2 \leq \|\varphi_y - v_1\|_2 + \|\varphi_y - v_2\|_2 \leq 2 \max_{k=1,2} \|\varphi_y - v_k\|_2.$$

This holds for all $f \in \mathcal{A}$, so combining the previous two inequalities completes the proof of the lower bound for the min-max error. □

Appendix C. Proof of lemmas

C.1. Proof of Lemma A.1

Proof. Fix a $\omega_j \in \Omega$, and so $\omega_j \in \Lambda_a$ for some $1 \leq a \leq A$. We explicitly construct each I_j, and it is more convenient to break the construction into two cases.

The simpler case is when $\lambda_a = 1$. Note that $B_a = \rho_j = 1$. Then we simply set

$$I_j(\omega) := e^{2\pi i M(\omega - \omega_j)} F_M(\omega - \omega_j),$$

where we recall that F_M is the Fejér kernel. We trivially have $I_j(\omega_k) = \delta_{j,k}$ for all $\omega_k \in \Lambda_a$ and $I_j \in \mathcal{P}(M)$. Using the point-wise bound for the Fejér kernel (A.2) and the cluster separation condition (2.2), we have

$$|I_j(\omega_k)| \leq \frac{1}{4(M+1)^2 |\omega_k - \omega_j|_{\mathbb{T}}^2} \leq \frac{1}{400S}.$$

Using the L^2 norm bound for the Fejér kernel (A.1), we see that

$$\|I_j\|_{L^2(\mathbb{T})} \leq \frac{1}{\sqrt{M+1}}.$$

This completes the proof of the lemma when $\lambda_a = 1$.

From here onwards, we assume that $\lambda \geq 2$. To define I_j, we must construct two axillary functions G_j and H_j. We define the Lagrange-like polynomial,

$$G_j(\omega) := \prod_{\omega_k \in \Lambda_a \setminus \{\omega_j\}} \frac{e^{2\pi i Q_j t} - e^{2\pi i Q_j \omega_k}}{e^{2\pi i Q_j \omega_j} - e^{2\pi i Q_j \omega_k}}, \quad \text{where} \quad Q_j := \left\lfloor \frac{M}{\lambda_a} \right\rfloor.$$

Note that Q_j is positive because $M/\lambda_a \geq M/S \geq 1$. This function is well-defined because its denominator is always non-zero: this follows from the observation that the inequalities, $Q_j \leq M/2$ and $|\omega_j - \omega_k|_{\mathbb{T}} < 1/M$, imply

$$|Q_j \omega_j - Q_j \omega_k|_{\mathbb{T}} = Q_j |\omega_j - \omega_k|_{\mathbb{T}}.$$

By construction, the function G_j satisfies the important property that

$$G_j(\omega_k) = \delta_{j,k}, \quad \text{for all} \quad \omega_k \in \Lambda_a. \tag{C.1}$$

We upper bound G_j in the sup-norm. We begin with the estimate

$$\|G_j\|_{L^\infty(\mathbb{T})} \leq \prod_{\omega_k \in \Lambda_a \setminus \{\omega_j\}} \frac{2}{|1 - e^{2\pi i Q_j(\omega_j - \omega_k)}|}.$$

Recall the trigonometric inequality,

$$2 - 2\cos(2\pi t) \geq (2\pi t)^2\Big(1 - \frac{\pi^2 t^2}{3}\Big) \quad \text{for } t \in [-1/2, 1/2],$$

which follows from a Taylor expansion of cosine. Using this inequality, we deduce the bound,

$$\|G_j\|_{L^\infty(\mathbb{T})} \leq \prod_{\omega_k \in \Lambda_a \setminus \{\omega_j\}} \frac{1}{\pi Q_j |\omega_j - \omega_k|_{\mathbb{T}}} \Big(1 - \frac{\pi^2 Q_j^2 |\omega_j - \omega_k|_{\mathbb{T}}^2}{3}\Big)^{-1/2}.$$

Since $Q_j = \lfloor M/\lambda_a \rfloor$ and $|\omega_j - \omega_k|_{\mathbb{T}} < 1/M$, we have

$$\|G_j\|_{L^\infty(\mathbb{T})} \leq \Big(1 - \frac{\pi^2}{3\lambda_a^2}\Big)^{-(\lambda_a - 1)/2} \prod_{\omega_k \in \Lambda_a \setminus \{\omega_j\}} \frac{1}{\pi Q_j |\omega_j - \omega_k|_{\mathbb{T}}} = \widetilde{B}_a \lambda_a^{\lambda_a - 1} \rho_j. \tag{C.2}$$

We next define the function H_j by the formula,

$$H_j(\omega) := \big(e^{2\pi i P_j(\omega - \omega_j)} F_{P_j}(\omega - \omega_j)\big)^{\lambda_a}, \quad \text{where} \quad P_j := \Big\lfloor \frac{M}{2\lambda_a^2} \Big\rfloor.$$

Recall that F_{P_j} denotes the Fejér kernel and note that P_j is positive because $M/(2\lambda_a^2) \geq M/(2S^2) \geq 1$. We need both a decay and norm bound for H_j. To obtain a norm bound, we use Hölder's inequality, that the Fejér kernel is point-wise upper bounded by 1, the norm bound for the Fejér kernel (A.1), and the inequality $P_j + 1 \geq M/(2\lambda_a^2)$, to obtain,

$$\|H_j\|_{L^2(\mathbb{T})} \leq \|F_{P_j}\|_{L^\infty(\mathbb{T})}^{\lambda_a - 1} \|F_{P_j}\|_{L^2(\mathbb{T})} \leq \frac{1}{\sqrt{P_j + 1}} \leq \Big(\frac{2\lambda_a^2}{M}\Big)^{1/2}. \tag{C.3}$$

To obtain a decay bound for H_j, we use the point-wise bound for the Fejér kernel (A.2) to deduce,

$$|H_j(\omega)| \leq \Big(\frac{1}{2(P_j + 1)|\omega - \omega_j|_{\mathbb{T}}}\Big)^{2\lambda_a} \leq \Big(\frac{\lambda_a^2}{M|\omega - \omega_j|_{\mathbb{T}}}\Big)^{2\lambda_a}, \quad \text{for all} \quad \omega \in \mathbb{T}.$$

We would like to specialize this to the case that $\omega = \omega_j$ for $\omega_j \notin \Lambda_a$. We need to make the following observations first. Observe that $1 \leq \lfloor t \rfloor / t \leq 2$ for any $t \geq 1$. Using this inequality and that $\lambda_a \geq 2$, we see that

$$(20\widetilde{B}_a)^{1/(2\lambda_a)} \leq 20^{1/(2\lambda_a)} \Big(1 - \frac{\pi^2}{3\lambda_a^2}\Big)^{-1/4 + 1/(4\lambda_a)} 2^{(\lambda_a - 1)/(2\lambda_a)} \leq 10.$$

This inequality and the cluster separation condition (2.2) imply

$$|\omega_k - \omega_j|_{\mathbb{T}} \geq \frac{10\lambda_a^2 (S\lambda_a^{\lambda_a - 1}\rho_j)^{1/(2\lambda_a)}}{M} \geq \frac{\lambda_a^2 (20\widetilde{B}_a S\lambda_a^{\lambda_a - 1}\rho_j)^{1/(2\lambda_a)}}{M} \quad \text{for all} \quad \omega_k \notin \Lambda_a.$$

Combining this with the previous upper bound on H_j shows that

$$|H_j(\omega_k)| \leq \frac{1}{20S\widetilde{B}_a \lambda_a^{\lambda_a - 1}\rho_j} \quad \text{for all} \quad \omega_k \notin \Lambda_a. \tag{C.4}$$

We define the function I_j by the formula

$$I_j(\omega) := G_j(\omega) H_j(\omega).$$

It follows immediately from the property (C.1) that

$$I_j(\omega_k) = \delta_{j,k} \quad \text{for all} \quad \omega_k \in \Lambda_a.$$

The negative frequencies of I_j are zero, while its largest non-negative frequency is bounded above by

$$2P_j\lambda_a + (\lambda_a - 1)Q_j \leq \frac{M}{\lambda_a} + (\lambda_a - 1)\Big(\frac{M}{\lambda_a}\Big) \leq M,$$

which proves that $I_j \in \mathcal{P}(M)$. We use Hölder's inequality, the sup-norm bound for G_j (C.2), and the norm bound for H_j (C.3) to see that

$$\|I_j\|_{L^2(\mathbb{T})} \leq \|G_j\|_{L^\infty(\mathbb{T})}\|H_j\|_{L^2(\mathbb{T})} \leq \widetilde{B}_a\lambda_a^{\lambda_a - 1}\rho_j\Big(\frac{2}{M}\Big)^{1/2}.$$

Finally, we use the sup-norm bound for G_j (C.2) and the bound for $|H_j(\omega_k)|$ (C.4) to see that

$$|I_j(\omega_k)| \leq \|G_j\|_{L^\infty(\mathbb{T})}|H_j(\omega_k)| \leq \frac{1}{20S} \quad \text{for all} \quad \omega_k \notin \Lambda_a. \quad \square$$

C.2. Proof of Lemma A.2

Proof. Fix integers M, N, S satisfying the assumptions of Lemma A.2. Fix a support set Ω, contained in the grid with spacing $1/N$ and of cardinality S. We do a two-scale analysis. For each $\omega_j \in \Omega$, we define the discrete sets and integers,

$$\Gamma_j := \Gamma_j(\Omega) = \Big\{\omega_k \in \Omega\colon |\omega_k - \omega_j|_{\mathbb{T}} < \frac{1}{M}\Big\} \quad \text{and} \quad \gamma_j := |\Gamma_j|,$$
$$\mathcal{T}_j := \mathcal{T}_j(\Omega) = \Big\{\omega_k \in \Omega\colon |\omega_k - \omega_j|_{\mathbb{T}} < \frac{S}{2M}\Big\} \quad \text{and} \quad \tau_j := |\mathcal{T}_j|.$$

To construct $H_j(\cdot, \Omega)$, we need to define two axillary functions, similar to the construction done in Lemma A.1. We define the integers

$$Q_{j,k} := Q_{j,k}(\Omega) := \begin{cases} \lfloor M/S \rfloor & \text{if } \omega_k \in \mathcal{T}_j \setminus \{\omega_j\}, \\ \lfloor 1/(2|\omega_j - \omega_k|_{\mathbb{T}}) \rfloor & \text{if } \omega_k \in \Omega \setminus \mathcal{T}_j. \end{cases}$$

We readily verify that we have the inequalities $1 \leq Q_{j,k} \leq M/S$ and

$$|Q_{j,k}\omega_j - Q_{j,k}\omega_k|_{\mathbb{T}} = Q_{j,k}|\omega_j - \omega_k|_{\mathbb{T}} \quad \text{for all} \quad \omega_j, \omega_k \in \Omega. \tag{C.5}$$

This observation implies that the Lagrange-like polynomial,

$$G_j(\omega) := G_j(\omega, \Omega) := \prod_{\omega_k \in \Omega \setminus \{\omega_j\}} \frac{e^{2\pi i Q_{j,k}\omega} - e^{2\pi i Q_{j,k}\omega_k}}{e^{2\pi i Q_{j,k}\omega_j} - e^{2\pi i Q_{j,k}\omega_k}},$$

has non-zero denominators, and is thus well-defined. By construction, we have the interpolation identity,

$$G_j(\omega_k) = \delta_{j,k} \quad \text{for all} \quad \omega_j, \omega_k \in \Omega.$$

We bound G_j in the sup-norm. We begin with the inequality,

$$\|G_j\|_{L^\infty(\mathbb{T})} \le \prod_{\omega_k\in\Omega\setminus\{\omega_j\}} \frac{2}{|1-e^{2\pi i Q_{j,k}(\omega_j-\omega_k)}|}. \tag{C.6}$$

Recall that we have the partition,

$$\Omega\setminus\{\omega_j\} = (\Gamma_j\setminus\{\omega_j\})\cup(\mathcal{T}_j\setminus\Gamma_j)\cup(\Omega\setminus\mathcal{T}_j).$$

Then we break (C.6) into three products according to this partition, and estimate each term at a time.

(a) We first consider the product over $\omega_k\in\Gamma_j\setminus\{\omega_j\}$. If $\Gamma_j\setminus\{\omega_j\}=\emptyset$, there is nothing to do. Hence, assume that $\gamma_j\ge 2$. By a Taylor expansion for cosine, we obtain the inequality,

$$2-2\cos(2\pi t)\ge (2\pi t)^2\Big(1-\frac{\pi^2t^2}{3}\Big)\quad \text{for } t\in[-1/2,1/2].$$

Using this lower bound, the observation that $Q_{j,k}=\lfloor M/S\rfloor\le M/S\le M/\gamma_j$ when $\omega_k\in\Gamma_j\setminus\{\omega_j\}$, and the assumption that $|\omega_j-\omega_k|<1/M$ for all $\omega_k\in\Gamma_j$, we obtain

$$\begin{aligned}
&\prod_{\omega_k\in\Gamma_j\setminus\{\omega_j\}} \frac{2}{|1-e^{2\pi i Q_{j,k}(\omega_j-\omega_k)}|}\\
&\le \prod_{\omega_k\in\Gamma_j\setminus\{\omega_j\}} \Big(1-\frac{\pi^2Q_{j,k}^2|\omega_j-\omega_k|_{\mathbb{T}}^2}{3}\Big)^{-1/2} \prod_{\omega_k\in\Gamma_j\setminus\{\omega_j\}} \frac{1}{\pi Q_{j,k}|\omega_j-\omega_k|_{\mathbb{T}}}\\
&\le \Big(1-\frac{\pi^2}{3\gamma_j^2}\Big)^{-(\gamma_j-1)/2}\Big\lfloor\frac{M}{S}\Big\rfloor^{-(\gamma_j-1)} \prod_{\omega_k\in\Gamma_j\setminus\{\omega_j\}} \frac{1}{\pi|\omega_j-\omega_k|_{\mathbb{T}}}\\
&\le \Big(\frac{12}{12-\pi^2}\Big)^{1/2}\Big\lfloor\frac{M}{S}\Big\rfloor^{-(\gamma_j-1)} \prod_{\omega_k\in\Gamma_j\setminus\{\omega_j\}} \frac{1}{\pi|\omega_j-\omega_k|_{\mathbb{T}}}.
\end{aligned}$$

For the last inequality, we made the observation that $(1-\pi^2/(3t^2))^{-(t-1)/2}$ is a decreasing function of t on the domain $t\ge 2$.

(b) We consider the product over $\omega_k\in\mathcal{T}_j\setminus\Gamma_j$, and note that $Q_{j,k}=\lfloor M/S\rfloor$ for this case. Recall the trigonometric inequality

$$|e^{2\pi i t}-1|\ge 4|t|_{\mathbb{T}},\quad \text{for all}\quad t\in\mathbb{R}. \tag{C.7}$$

We this trigonometric inequality and (C.5) to see that

$$\begin{aligned}
\prod_{\omega_k\in\mathcal{T}_j\setminus\Gamma_j} \frac{2}{|1-e^{2\pi i Q_{j,k}(\omega_j-\omega_k)}|} &\le \prod_{\omega_k\in\mathcal{T}_j\setminus\Gamma_j} \frac{1}{2Q_{j,k}|\omega_j-\omega_k|_{\mathbb{T}}}\\
&\le \Big\lfloor\frac{M}{S}\Big\rfloor^{-\tau_j+\gamma_j}\Big(\frac12\Big)^{\tau_j-\gamma_j} \prod_{\omega_k\in\mathcal{T}_j\setminus\Gamma_j} \frac{1}{|\omega_j-\omega_k|_{\mathbb{T}}}.
\end{aligned}$$

(c) For the product over $\omega_k\in\Omega\setminus\mathcal{T}_j$, note that $Q_{j,k}|\omega_j-\omega_k|_{\mathbb{T}}\ge 1/4$. Using this and the trigonometric inequality (C.7) again, we see that

$$\prod_{\omega_k\in\Omega\setminus\mathcal{T}_j} \frac{2}{|1-e^{2\pi i Q_{j,k}(\omega_j-\omega_k)}|} \le \prod_{\omega_k\in\Omega\setminus\mathcal{T}_j} \frac{1}{2Q_{j,k}|\omega_j-\omega_k|_{\mathbb{T}}} \le 2^{S-\tau_j}.$$

Combining the above three inequities with inequality (C.6) and simplifying, we obtain an upper bound

$$\|G_j\|_{L^\infty(\mathbb{T})} \le \Big(\frac{12}{12-\pi^2}\Big)^{1/2} \Big\lfloor \frac{M}{S} \Big\rfloor^{-\tau_j+1} \Big(\frac{1}{\pi}\Big)^{\gamma_j-1} 2^{S-2\tau_j+\gamma_j} \prod_{\omega_k\in\mathcal{T}_j\setminus\{\omega_j\}} \frac{1}{|\omega_j-\omega_k|_{\mathbb{T}}}. \tag{C.8}$$

Let $P = \lfloor M/(2S)\rfloor$ and note that $P \ge 1$ because $M \ge 2S$. Let F_P be the Fejér kernel, and by the $L^2(\mathbb{T})$ bound for the Fejér kernel and the observation that $P+1 \ge M/(2S)$, we have

$$\|F_P\|_{L^2(\mathbb{T})} \le \Big(\frac{1}{P+1}\Big)^{1/2} \le \Big(\frac{2S}{M}\Big)^{1/2}. \tag{C.9}$$

Finally, we define H_j by the formula,

$$H_j(\omega) := H_j(\omega,\Omega) := e^{2\pi i P(\omega-\omega_j)} F_P(\omega-\omega_j) G_j(\omega).$$

We still have the interpolation property that

$$H_j(\omega_k) = \delta_{j,k} \quad \text{for all} \quad \omega_j,\omega_k \in \Omega.$$

By construction, the negative frequencies of H_j are zero while its largest positive frequency is bounded above by

$$2P + \sum_{k\ne j} Q_{j,k} \le \frac{M}{S} + \sum_{k\ne j} \frac{M}{S} = \frac{M}{S} + \frac{M(S-1)}{S} \le M.$$

This proves that $H_j \in \mathcal{P}(M)$.

It remains to upper bound $\sum_{j=1}^S \|H_j\|^2_{L^2(\mathbb{T})}$. By Hölder's inequality and the inequalities,

$$\Big(\sum_{j=1}^S \|H_j\|^2_{L^2(\mathbb{T})}\Big)^{1/2} \le \Big(\sum_{j=1}^S \|F_P\|^2_{L^2(\mathbb{T})}\|G_j\|^2_{L^\infty(\mathbb{T})}\Big)^{1/2} \le \Big(\frac{24}{12-\pi^2}\Big)^{1/2}\Big(\frac{S}{M}\Big)^{1/2} E(\Omega)^{1/2},$$

where the constant $E(\Omega)$ is defined as

$$E(\Omega) := \sum_{j=1}^S \Big\lfloor \frac{M}{S}\Big\rfloor^{-2\tau_j+2} \Big(\frac{1}{\pi^2}\Big)^{\gamma_j-1} 4^{S-2\tau_j+\gamma_j} \prod_{\omega_k\in\mathcal{T}_j\setminus\{\omega_j\}} \frac{1}{|\omega_j-\omega_k|^2_{\mathbb{T}}}. \tag{C.10}$$

To complete the proof of the lemma, we need to obtain the appropriate bound on $E(\Omega)$ uniformly in Ω. This is handled separately in Lemma C.1, which is stated below and proved in Appendix C.3. □

Lemma C.1. *Suppose the assumptions of Lemma A.2 hold and let $E(\Omega)$ be the quantity defined in (C.10). Then*

$$E(\Omega) \le \Big\lfloor \frac{M}{S}\Big\rfloor^{-2S+2} N^{2S-2} \Big(\frac{1}{\pi}\Big)^{2S-2} \sum_{j=1}^S \prod_{k\ne j} \frac{1}{(j-k)^2}.$$

C.3. Proof of Lemma C.1

Before we prove the lemma, we motivate the argument that we are about to use. We view $E(\Omega)$ as a function defined on all $\binom{N}{S}$ possible sets Ω supported on the grid with spacing $1/N$ and of cardinality S. To upper bound $E(\Omega)$ uniformly in Ω, one method is to determine which Ω attain(s) the maximum. The maximizer is clearly not unique, since $E(\Omega)$ is invariant under cyclic shifts of Ω by $1/N$. However, we shall argue that the maximizer is attained by shifts of $\Omega_* = \{n/N\}_{n=0}^{S-1}$. Note that

$$E(\Omega_*) = \Big\lfloor \frac{M}{S} \Big\rfloor^{-2S+2} N^{2S-2} \Big(\frac{1}{\pi}\Big)^{2S-2} \sum_{j=1}^{S} \prod_{k\neq j} \frac{1}{(j-k)^2}. \tag{C.11}$$

Thus, the lemma is complete once we prove that $E(\Omega) \leq E(\Omega_*)$. While it seems intuitive that $E(\Omega) \leq E(\Omega_*)$ for all Ω, it is not straightforward to prove. When Ω is contained in a small interval, the product over $\mathcal{T}_j$ in the definition of $E(\Omega)$ given in (C.10) is large, but that is offset by the remaining terms, which are small. The major difficulty is that $E(\Omega)$ is highly dependent on the configuration of Ω. If we perturb just one of the $\omega_j \in \Omega$ and keep the rest fixed, it is possible for all S terms in the summation in the definition of $E(\Omega)$ to change. This makes continuity and perturbation arguments difficult to carry out. To deal with this difficultly, we proceed with the following extension argument.

Proof. We extend E to a function of $D = S(S-1)$ variables in the following way. We write $w \in \mathbb{R}^D$ to denote the D variables $\{w_{j,k}\}_{1\leq j,k\leq S, j\neq k}$. We do not impose that $\{w_{j,k}\}_{j\neq k}$ are unique, that $w_{j,k} = w_{k,j}$, or that they lie on some grid. They are just D independent real variables for now. We define the sets and integers,

$$A_j(w) := \Big\{ w_{j,k} \colon w_{j,k} < \frac{S}{2M} \Big\} \quad \text{and} \quad a_j(w) := |A_j(w)|,$$
$$B_j(w) := \Big\{ w_{j,k} \colon w_{j,k} < \frac{1}{M} \Big\} \quad \text{and} \quad b_j(w) := |B_j(w)|.$$

We define the function $F \colon \mathbb{R}^D \to \mathbb{R}$ by the formula,

$$F(w) := \sum_{j=1}^{S} \Big\lfloor \frac{M}{S} \Big\rfloor^{-2a_j(w)+2} \Big(\frac{1}{\pi^2}\Big)^{b_j(w)-1} 4^{S-2a_j(w)+b_j(w)} \prod_{w_{j,k}\in A_j(w)} \frac{1}{w_{j,k}^2}. \tag{C.12}$$

We restrict F to the domain $[1/N, 1/2]^D \cap H$, where

$$H := \bigcap_{k=1}^{S} \Big\{ w \in \mathbb{R}^D \colon \sum_{j\neq k} w_{j,k} \geq \frac{c(S)}{N} \Big\},$$

and the constant $c(S)$ is defined as

$$c(S) := \begin{cases} 2\Big(1+2+\cdots+\frac{S-1}{2}\Big) & \text{if } S \text{ is odd,} \\ 2\Big(1+2+\cdots+\frac{S-2}{2}\Big) + \frac{S}{2} & \text{if } S \text{ is even.} \end{cases}$$

We argue that F is an extension of E. Note that any Ω can be mapped to a $w(\Omega) \in \mathbb{R}^D$ via the relationship $(w(\Omega))_{j,k} = |\omega_j - \omega_k|_{\mathbb{T}}$ for all $j \neq k$. Under this mapping, we have $a_j(w) = \tau_j$ and $b_j(w) = \gamma_j$, which shows that

$$F(w(\Omega)) = E(\Omega).$$

Moreover, $w(\Omega)$ is clearly contained in $[1/N, 1/2]^D$. For each $1 \le k \le S$, we have

$$\sum_{j \neq k} (w(\Omega))_{j,k} = \sum_{j \neq k} |\omega_j - \omega_k|_{\mathbb{T}} \ge \frac{c(S)}{N}.$$

This inequality implies that $w(\Omega)$ is contained in the set $[1/N, 1/2]^D \cap H$. Thus, F is indeed an extension of E, and for all Ω, we have

$$E(\Omega) = F(w(\Omega)) \le \sup_{w \in [1/N, 1/2]^D \cap H} F(w). \tag{C.13}$$

We remark that there is a clear advantage of working with F instead of E. If one coordinate of w is perturbed while the rest of the $D-1$ coordinates of w remain fixed, then only one of the S terms in the summation in (C.12) is perturbed.

Observe that $[1/N, 1/2]^D \cap H$ is compact because it is the intersection of a closed cube with S closed half-spaces. Clearly F is continuous on the domain $[1/N, 1/2]^D \cap H$, so the supremum of F is attained at some point in this set. We first simplify matters and prove that

$$\max_{w \in [1/N, 1/2]^D \cap H} F(w) = \max_{w \in [1/N, 1/M]^D \cap H} F(w), \tag{C.14}$$

which is done via the following two reductions.

(a) Our first claim is that

$$\max_{w \in [1/N, 1/2]^D \cap H} F(w) = \max_{w \in [1/N, S/(2M)]^D \cap H} F(w).$$

Suppose for the purpose of yielding a contradiction, the maximum of F is not attained at any point in $[1/N, S/(2M)]^D \cap H$. This is equivalent to the claim that, for any maximizer w of F, there exist indices (m, n) such that $a_m(w) \le S-1$ and $w_{m,n} > S/(2M)$. We define the vector $v \in [1/N, 1/2]^D \cap H$ by the relationship

$$v_{j,k} := \begin{cases} S/N & \text{if } (j,k) = (m,n), \\ w_{j,k} & \text{otherwise.} \end{cases}$$

Since v and w agree except at one coordinate, we readily calculate that

$$\begin{aligned} &F(w) - F(v) \\ &= \Big\lfloor \frac{M}{S} \Big\rfloor^{-2a_m(w)+2} \Big(\frac{1}{\pi^2}\Big)^{b_m(w)-1} 4^{S-2a_m(w)+b_m(w)} \Big(\prod_{w_{j,k} \in A_j(w)} \frac{1}{w_{m,k}^2} \Big) \Big(1 - \Big\lfloor \frac{M}{S} \Big\rfloor^{-2} \frac{1}{4\pi^2 v_{m,n}^2}\Big). \end{aligned}$$

The assumption that $N \ge \pi M S$ and $S \ge 2$ imply

$$\frac{1}{4\pi^2 v_{m,n}^2} = \frac{N^2}{4\pi^2 S^2} \ge \Big(\frac{M}{2}\Big)^2 \ge \Big\lfloor \frac{M}{S} \Big\rfloor^2.$$

This proves that $F(w) \le F(v)$, which is a contradiction.

(b) Our second claim is that

$$\max_{w\in[1/N,S/(2M)]^D\cap H} F(w) = \max_{w\in[1/N,1/M]^D\cap H} F(w).$$

Suppose for the purpose of yielding a contradiction, the maximum of F is not attained at any point in $[1/N,1/M]^D\cap H$. This is equivalent to the claim that, for any maximizer $w\in[1/N,S/(2M)]^D\cap H$, there exist indices (m,n) such that $b_m(w)\leq S-1$ and $w_{m,n}\geq 1/M$. We define the vector $v\in[1/N,1/M]^D\cap H$ by the relationship,

$$v_{j,k} = \begin{cases} S/N & \text{if } (j,k)=(m,n) \\ w_{j,k} & \text{otherwise.} \end{cases}$$

Since v and w agree except at one coordinate, we see that

$$F(w)-F(v) = \Big\lfloor \frac{M}{S}\Big\rfloor^{-2S+2}\Big(\frac{1}{\pi^2}\Big)^{b_m(w)-1}4^{-S+b_m(w)}\Big(\prod_{k\in A_m(w)\setminus\{n\}}\frac{1}{w_{m,k}^2}\Big)\Big(\frac{1}{w_{m,n}^2}-\frac{4}{\pi^2 v_{m,n}^2}\Big).$$

The assumption that $N\geq \pi MS$ implies

$$\frac{4}{\pi^2 v_{m,n}^2} = \frac{4N^2}{\pi^2 S^2} \geq 4M^2 \geq \frac{1}{w_{m,n}^2}.$$

This shows that $F(w)\leq F(v)$, which is a contradiction.

Thus, we have established (C.14), and combining this fact with (C.13) yields,

$$E(\Omega) = F(w(\Omega)) \leq \max_{w\in[1/N,1/M]^D\cap H} F(w). \tag{C.15}$$

When $w\in[1/N,1/M]^D\cap H$, the function F reduces to

$$F(w) = \Big\lfloor \frac{M}{S}\Big\rfloor^{-2S+2}\Big(\frac{1}{\pi^2}\Big)^{S-1}\sum_{j=1}^{S}\prod_{k\in B_j(w)}\frac{1}{w_{j,k}^2}.$$

Since F is a smooth function of w, a straightforward calculation shows that each partial derivative of F, with respect to the canonical basis of $\mathbb{R}^D$, is strictly negative on $[1/N,1/M]^D\cap H$. Thus, the maximum of F is attained on the boundary of $[1/N,1/M]^D\cap H$. In fact, H is the intersection of S half-spaces and the boundary of the k-th half-space is the hyperplane

$$H_k := \Big\{w\in\mathbb{R}^D\colon \sum_{j\neq k} w_{j,k} = \frac{c(S)}{N}\Big\}.$$

Since each partial derivative of F is strictly negative on $[1/N,1/M]^D\cap H$, we see that the maximum of F must be attained on one of these hyperplanes. We observe that $w(\Omega)$ lies on a H_k if and only if Ω consists of S consecutive indices. This proves that for all Ω, we have

$$E(\Omega) = F(w(\Omega)) \leq F(w(\Omega_*)) = E(\Omega_*).$$

This combined with the formula for $E(\Omega_*)$ given in (C.11) completes the proof of the lemma. □

C.4. Upper bounds for discrete quantities

Lemma C.2. *For any integer $n \geq 2$, we have*

$$\Big(\sum_{j=1}^{n}\prod_{k=1,\ k\neq j}^{n}\frac{1}{(j-k)^2}\Big)^{1/2} \leq 2\pi e\sqrt{n}\left(\frac{n}{2}-1\right)^{-n}e^{n}.$$

Proof. Notice that

$$\min_{1\leq j\leq n}\prod_{k=1,\ k\neq j}^{n}|j-k| \geq \begin{cases} (\frac{n}{2}-1)!(\frac{n}{2})! & \text{if } n \text{ is even} \\ (\frac{n-1}{2})!(\frac{n-1}{2})! & \text{if } n \text{ is odd.} \end{cases}$$

By further using the well-known inequality, $k! \geq \sqrt{2\pi}\,k^{k+\frac{1}{2}}e^{-k}$ for any integer $k \geq 1$,

$$\min_{1\leq j\leq n}\prod_{k=1,\ k\neq j}^{n}|j-k| \geq \begin{cases} 2\pi(\frac{n}{2}-1)^{\frac{n}{2}-\frac{1}{2}}(\frac{n}{2})^{\frac{n}{2}+\frac{1}{2}}e^{-n+1} & \text{if } n \text{ is even} \\ 2\pi(\frac{n-1}{2})^{n}e^{-n+1} & \text{if } n \text{ is odd.} \end{cases}$$

Hence, for any $n \geq 2$, we have

$$\min_{1\leq j\leq n}\prod_{k=1,\ k\neq j}^{n}|j-k| \geq 2\pi e\Big(\frac{n}{2}-1\Big)^{n}e^{-n}.$$

Thus,

$$\Big(\sum_{j=1}^{n}\prod_{k=1,\ k\neq j}^{n}\frac{1}{(j-k)^2}\Big)^{1/2} \leq 2\pi e\Big(\sum_{j=1}^{n}\Big(\frac{n}{2}-1\Big)^{-2n}e^{2n}\Big)^{1/2} = 2\pi e\sqrt{n}\left(\frac{n}{2}-1\right)^{-n}e^{n}. \quad \square$$

References

[1] Albert C. Fannjiang, Thomas Strohmer, Pengchong Yan, Compressed remote sensing of sparse objects, SIAM J. Imaging Sci. 3 (3) (2010) 595–618.
[2] Albert C. Fannjiang, Compressive sensing theory for optical systems described by a continuous model, arXiv preprint, arXiv:1507.00794, 2015.
[3] Albert C. Fannjiang, Compressive inverse scattering: I. high-frequency simo/miso and mimo measurements, Inverse Probl. 26 (3) (2010) 035008.
[4] Hamid Krim, Viberg Mats, Two decades of array signal processing research: the parametric approach, IEEE Signal Process. Mag. 13 (4) (1996) 67–94.
[5] Ralph Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag. 34 (3) (1986) 276–280.
[6] Petre Stoica, Randolph L. Moses, Introduction to Spectral Analysis, Vol. 1, Prentice Hall, 1997.
[7] Arnold Jan Den Dekker, A. Van den Bos. Resolution, A survey, J. Opt. Soc. Am. 14 (3) (1997) 547–557.
[8] Gaspard Riche de Prony, Essai expérimentale et analytique, J. Éc. Polytech. 1 (22) (1795) 24–76.
[9] Dmitry Batenkov, Yosef Yomdin, On the accuracy of solving confluent Prony systems, SIAM J. Appl. Math. 73 (1) (2013) 134–154.
[10] Gene Golub, Victor Pereyra, Separable nonlinear least squares: the variable projection method and its applications, Inverse Probl. 19 (2) (2003) R1.
[11] Gregory Beylkin, Lucas Monzón, On approximation of functions by exponential sums, Appl. Comput. Harmon. Anal. 19 (1) (2005) 17–48.
[12] Daniel Potts, Manfred Tasche, Parameter estimation for exponential sums by approximate Prony method, Signal Process. 90 (5) (2010) 1631–1642.
[13] Richard Roy, Thomas Kailath, ESPRIT – estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust. Speech Signal Process. 37 (7) (1989) 984–995.
[14] Yingbo Hua, Tapan K. Sarkar, Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise, IEEE Trans. Acoust. Speech Signal Process. 38 (5) (1990) 814–824.
[15] Emmanuel J. Candès, Carlos Fernandez-Granda, Super-resolution from noisy data, J. Fourier Anal. Appl. 19 (6) (2013) 1229–1254.

[16] Carlos Fernandez-Granda, Support detection in super-resolution, in: Proceedings of the 10th International Conference on Sampling Theory and Applications, 2013, pp. 145–148.
[17] Jean-Marc Azaïs, Yohann De Castro, Fabrice Gamboa, Spike detection from inaccurate samplings, Appl. Comput. Harmon. Anal. 38 (2) (2015) 177–195.
[18] Vincent Duval, Gabriel Peyré, Exact support recovery for sparse spikes deconvolution, Found. Comput. Math. 15 (5) (2015) 1315–1355.
[19] Weilin Li, Elementary L^∞ error estimates for super-resolution de-noising, arXiv preprint, arXiv:1702.03021, 2017.
[20] Marco F. Duarte, Richard G. Baraniuk, Spectral compressive sensing, Appl. Comput. Harmon. Anal. 35 (1) (2013) 111–129.
[21] Albert C. Fannjiang, Wenjing Liao, Coherence-pattern guided compressive sensing with unresolved grids, SIAM J. Imaging Sci. 5 (1) (2012) 179–202.
[22] Kristian Bredies, Hanna Katriina Pikkarainen, Inverse problems in spaces of measures, ESAIM Control Optim. Calc. Var. 19 (1) (2013) 190–218.
[23] Nicholas Boyd, Geoffrey Schiebinger, Benjamin Recht, The alternating descent conditional gradient method for sparse inverse problems, SIAM J. Optim. 27 (2) (2017) 616–639.
[24] Quentin Denoyelle, Vincent Duval, Gabriel Peyré, Emmanuel Soubies, The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy, Inverse Probl. 36 (1) (2019) 014001.
[25] Albert C. Fannjiang, The MUSIC algorithm for sparse objects: a compressed sensing analysis, Inverse Probl. 27 (3) (2011) 035013.
[26] Wenjing Liao, Albert Fannjiang, MUSIC for single-snapshot spectral estimation: stability and super-resolution, Appl. Comput. Harmon. Anal. 40 (1) (2016) 33–67.
[27] Wenjing Liao, Music for multidimensional spectral estimation: stability and super-resolution, IEEE Trans. Signal Process. 63 (23) (2015) 6395–6406.
[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, 2015.
[29] Albert C. Fannjiang, Wenjing Liao, Mismatch and resolution in compressive imaging, in: Wavelets and Sparsity XIV, Proceedings of SPIE, 2011.
[30] Yuejie Chi, Louis L. Scharf, Pezeshki Ali, A. Robert Calderbank, Sensitivity to basis mismatch in compressed sensing, IEEE Trans. Signal Process. 59 (5) (2011) 2182–2195.
[31] Emmanuel J. Candès, Justin Romberg, Terence Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory 52 (2) (2006) 489–509.
[32] David L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306.
[33] Veniamin I. Morgenshtern, Emmanuel J. Candes, Super-resolution of positive sources: the discrete setup, SIAM J. Imaging Sci. 9 (1) (2016) 412–444.
[34] Quentin Denoyelle, Vincent Duval, Gabriel Peyré, Support recovery for sparse super-resolution of positive measures, J. Fourier Anal. Appl. 23 (5) (2017) 1153–1194.
[35] Veniamin I. Morgenshtern, Super-resolution of positive sources on an arbitrarily fine grid, arXiv preprint, arXiv:2005.06756, 2020.
[36] John J. Benedetto, Weilin Li, Super-resolution by means of Beurling minimal extrapolation, Appl. Comput. Harmon. Anal. 48 (1) (2020) 218–241.
[37] David L. Donoho, Superresolution via sparsity constraints, SIAM J. Math. Anal. 23 (5) (1992) 1309–1331.
[38] Laurent Demanet, Nam Nguyen, The recoverability limit for superresolution via sparsity, arXiv preprint, arXiv:1502.01385, 2015.
[39] Dmitry Batenkov, Gil Goldman, Yosef Yomdin, Super-resolution of near-colliding point sources, arXiv preprint, arXiv:1904.09186, 2019.
[40] J.W. Odendaal, E. Barnard, C.W.I. Pistorius, Two-dimensional superresolution radar imaging using the music algorithm, IEEE Trans. Antennas Propag. 42 (10) (1994) 1386–1391.
[41] Petre Stoica, Virginija Šimonyte, Torsten Söderström, On the resolution performance of spectral analysis, Signal Process. 44 (2) (1995) 153–161.
[42] Harry B. Lee, The Cramér–Rao bound on frequency estimates of signals closely spaced in frequency, IEEE Trans. Signal Process. 40 (6) (1992) 1507–1517.
[43] Harry B. Lee, Eigenvalues and eigenvectors of covariance matrices for signals closely spaced in frequency, IEEE Trans. Signal Process. 40 (10) (1992) 2518–2535.
[44] Paul Turán, On rational polynomials, Acta Univ. Szeged., Sect. Sci. Math. (1946) 106–113.
[45] Alon Noga, Yuval Peres, Uniform dilations, Geom. Funct. Anal. 2 (1) (1992) 1–28.
[46] Sergei V. Konyagin, Imre Z. Ruzsa, Wilhelm Schlag, On uniformly distributed dilates of finite integer sequences, J. Number Theory 82 (2) (2000) 165–187.
[47] Walter Gautschi, Gabriele Inglese, Lower bounds for the condition number of Vandermonde matrices, Numer. Math. 52 (3) (1987) 241–250.
[48] Bernhard Beckermann, The condition number of real Vandermonde, Krylov and positive definite Hankel matrices, Numer. Math. 85 (4) (2000) 553–577.
[49] Alfredo Eisinberg, Giuseppe Franzé, Nicola Salerno, Rectangular Vandermonde matrices on Chebyshev nodes, Linear Algebra Appl. 338 (1–3) (2001) 27–36.
[50] Alfredo Eisinberg, Paolo Pugliese, Nicola Salerno, Vandermonde matrices on integer nodes: the rectangular case, Numer. Math. 87 (4) (2001) 663–674.
[51] Øyvind Ryan, Mérouane Debbah, Asymptotic behavior of random Vandermonde matrices with entries on the unit circle, IEEE Trans. Inf. Theory 55 (7) (2009) 3115–3147.
[52] Gabriel H. Tucci, Philip A. Whiting, Eigenvalue results for large scale random Vandermonde matrices with unit complex entries, IEEE Trans. Inf. Theory 57 (6) (2011) 3938–3954.

[53] Gabriel H. Tucci, Philip A. Whiting, Asymptotic behavior of the maximum and minimum singular value of random Vandermonde matrices, J. Theor. Probab. 27 (3) (2014) 826–862.
[54] Fermín Bazán, Conditioning of rectangular Vandermonde matrices with nodes in the unit disk, SIAM J. Matrix Anal. Appl. 21 (2) (2000) 679–693.
[55] Céline Aubel, Helmut Bölcskei, Vandermonde matrices with nodes in the unit disk and the large sieve, Appl. Comput. Harmon. Anal. (2017).
[56] Paulo J.S.G. Ferreira, Super-resolution, the recovery of missing samples and Vandermonde matrices on the unit circle, in: Proceedings of the Workshop on Sampling Theory and Applications, Loen, Norway, 1999.
[57] Lihu Berman, Arie Feuer, On perfect conditioning of Vandermonde matrices on the unit circle, Electron. J. Linear Algebra 16 (1) (2007) 13.
[58] Gautschi Walter, On inverses of Vandermonde and confluent Vandermonde matrices, Numer. Math. 4 (1) (1962) 117–123.
[59] Dmitry Batenkov, Laurent Demanet, Gil Goldman, Yosef Yomdin, Conditioning of partial nonuniform Fourier matrices with clustered nodes, SIAM J. Matrix Anal. Appl. 41 (1) (2020) 199–220.
[60] Stefan Kunis, Dominik Nagel, On the condition number of Vandermonde matrices with pairs of nearly-colliding nodes, Numer. Algorithms (2020) 1–24.
[61] Benedikt Diederichs, Well-posedness of sparse frequency estimation, arXiv preprint, arXiv:1905.08005, 2019.
[62] Stefan Kunis, Dominik Nagel, On the smallest singular value of multivariate Vandermonde matrices with clustered nodes, Linear Algebra Appl. 604 (2020) 1–20.
[63] Dmitry Batenkov, Benedikt Diederichs, Gil Goldman, Yosef Yomdin, The spectral properties of Vandermonde matrices with clustered nodes, Linear Algebra Appl. 609 (2021) 37–72.
[64] Alex H. Barnett, How exponentially ill-conditioned are contiguous submatrices of the Fourier matrix?, arXiv preprint, arXiv:2004.09643, 2020.
[65] Emmanuel J. Candes, The restricted isometry property and its implications for compressed sensing, C. R. Math. 346 (9–10) (2008) 589–592.
[66] Mark Rudelson, Roman Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, Commun. Pure Appl. Math. 61 (8) (2008) 1025–1045.
[67] Ping Liu, Hai Zhang, Resolution limit for line spectral estimation: theory and algorithm, arXiv preprint, arXiv:2003.02917, 2020.
[68] Per-Åke Wedin, Perturbation bounds in connection with singular value decomposition, BIT Numer. Math. 12 (1) (1972) 99–111.
[69] Mark Meckes, et al., On the spectral norm of a random Toeplitz matrix, Electron. Commun. Probab. 12 (2007) 315–325.
[70] Radosław Adamczak, A few remarks on the operator norm of random Toeplitz matrices, J. Theor. Probab. 23 (1) (2010) 85–108.
[71] Joel A. Tropp, User-friendly tail bounds for sums of random matrices, Found. Comput. Math. 12 (4) (2012) 389–434.
[72] Petre Stoica, Arye Nehorai Music, Maximum likelihood, and Cramer–Rao bound, IEEE Trans. Acoust. Speech Signal Process. 37 (5) (1989) 720–741.
[73] Bjorn Ottersten, Mats Viberg, Thomas Kailath, Performance analysis of the total least squares esprit algorithm, IEEE Trans. Signal Process. 39 (5) (1991) 1122–1135.
[74] Albert C. Fannjiang, Compressive spectral estimation with single-snapshot ESPRIT: stability and resolution, arXiv preprint, arXiv:1607.01827, 2016.
[75] Weilin Li, Wenjing Liao, Albert Fannjiang, Super-resolution limit of the esprit algorithm, IEEE Trans. Inf. Theory 66 (7) (2020) 4593–4608.