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

Weilin Li\(^*\) \quad Wenjing Liao\(^†\)

October 18, 2018

Abstract

Super-resolution refers to the process 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. The recovery process is highly sensitive to noise whenever the distance \(\Delta\) between the two closest point sources is less than \(1/M\). This paper studies the fundamental difficulty of super-resolution and the performance guarantees of a subspace method called MUSIC in the regime that \(\Delta < 1/M\).

The most important quantity in our theory is the minimum singular value of the Vandermonde matrix whose nodes are specified by the source locations. Under the assumption that the nodes are closely spaced within several well-separated clumps, we derive a sharp and non-asymptotic lower bound for this quantity. Our estimate is given as a weighted \(\ell^2\) sum, where each term only depends on the configuration of each individual clump. This implies that, as the noise increases, the super-resolution capability of MUSIC degrades according to a power law where the exponent depends on the cardinality of the largest clump. Numerical experiments validate our theoretical bounds for the minimum singular value and the resolution limit of MUSIC.

When there are \(S\) point sources located on a grid with spacing \(1/N\), the fundamental difficulty of super-resolution can be quantitatively characterized by a min-max error, which is the reconstruction error incurred by the best possible algorithm in the worst-case scenario. We show that the min-max error is closely related to the minimum singular value of Vandermonde matrices, and we provide a non-asymptotic and sharp estimate for the min-max error, where the dominant term is \((N/M)^{2S-1}\).

Keywords: Super-resolution, Vandermonde matrix with nodes on the unit circle, Fourier matrix, minimum singular value, subspace methods, MUSIC, min-max error, sparse recovery, polynomial interpolation, uncertainty principles

2010 Math subject classification: 42A10, 42A15, 94A08, 94A15, 94A20

1 Introduction

1.1 Problem statement

This paper studies the inverse problem of recovering the locations and amplitudes of a collection of point sources from its noisy low-frequency Fourier coefficients. Suppose there

\(^*\)Courant Institute of Mathematical Sciences, New York University. Email: weilinli@cims.nyu.edu

\(^†\)School of Mathematics, Georgia Institute of Technology. Email: wliao60@gatech.edu
are $S$ point sources with amplitudes $x = \{x_j\}_{j=1}^S \in \mathbb{C}^S$ located on an unknown discrete set $\Omega = \{\omega_j\}_{j=1}^S$ contained in the periodic interval $T = [0, 1)$. The collection of point sources can be represented by a discrete measure,

$$\mu(\omega) = \sum_{j=1}^S x_j \delta_{\omega_j}(\omega).$$

(1.1)

Here, $\delta_{\omega_j}$ denotes the Dirac measure supported in $\omega_j$ and $\Omega$ is the support of $\mu$, denoted $\text{supp}(\mu)$. 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 $\mu$:

$$y_k = \hat{\mu}(m) + \eta_k = \sum_{j=1}^S x_j e^{-2\pi i k \omega_j} + \eta_k, \quad k = 0, 1, \ldots, M,$$

(1.2)

where $\hat{\mu}$ is the Fourier transform of $\mu$ and $\eta_k$ represents some unknown noise at the $k$-th sensor. The goal is to accurately recover $\mu$, which consists of the location of the sources $\Omega$ and the amplitudes $x \in \mathbb{C}^S$, from the given noisy low-frequency Fourier data $y = \{y_k\}_{k=0}^M \in \mathbb{C}^{M+1}$.

The connection between this recovery problem and Vandermonde matrices is that $y$ satisfies the following linear system,

$$y = \Phi x + \eta,$$

(1.3)

where $\Phi$ is the $(M + 1) \times S$ Fourier or Vandermonde matrix (with nodes on the unit circle),

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

The terminology comes from the fact that $\Phi$ is a generalization of the discrete Fourier transform matrix and $e^{2\pi i \omega_j}$ lies on the complex unit circle. While it is convenient to reformulate the problem this way, we caution the reader that we do not have access to the sensing matrix $\Phi$ because it depends on the unknown $\Omega$.

This inverse problem arises in many interesting applications in imaging and signal processing, including:

(a) Inverse source and inverse scattering \cite{29,25,23}: in remote sensing, one often needs to detect a few point sources from the far-field measurements at a uniform array of sensors (see Figure 1 (a)). Since wave propagation in homogeneous media can be modeled as the Fourier transform by paraxial approximation to the Green’s function of the Helmholtz equation \cite{29}, it is reasonable to assume that sensors collect the noisy Fourier coefficients of the point sources with a cut-off frequency. In inverse scattering, one sends probe waves and collects scattered measurements as the observable data, with the goal of recovering the profile of the scatters. If the scatters are point-like objects and Born approximation (also known as Rayleigh-Gans scattering in optics) is used to model the scattered measurements, the inverse scattering problem can be simplified to (1.3) as well, see \cite{23}.
(a) Inverse source problem in imaging and DOA estimation in signal processing

(b) Direction-Of-Arrival (DOA) estimation [36, 48]: in signal processing, direction of arrival denotes the direction from which a propagating wave arrives at an array of sensors (also called antenna elements, see Figure 1 (b)). The goal is to find the direction relative to the array where the wave propagates from. It has a wide range of applications including radar, sonar and computed tomography. By letting $\omega_j = \Delta/\lambda \cos \phi_j$, where $\Delta$ is the spacing between sensors, $\lambda$ is the wave length, and $\phi_j$ is the direction of the $j$-th wave, we can formulate the DOA estimation as the inverse problem in (1.3).

(c) Spectral analysis [49]: spectral analysis considers the problem of determining the spectral content (i.e., the distribution of power over frequency) of a time series from a finite set of measurements. It has applications to many diverse fields, such as economics, astronomy, medicine, seismology, etc. For example, in speech analysis, spectral models of voice signals are useful in better understanding the speech production process, and are also used for both speech synthesis (or compression) and speech recognition. When the signals can be well described by a superposition of sines and cosines, classical spectral analysis is identical to the inverse problem in (1.3).

1.2 Background

The recovery problem is typically separated into two steps: estimation of the support set $\Omega$ and recovery of the coefficients $x$. Many methods first estimate $\Omega$ and then approximate $x$ by the least squares solution. In the noiseless setting, where $\eta = 0$, the classical Prony’s method [44] can recover $\mu$ exactly. However, this algorithm is unstable to noise, and this observation has partly motivated the development of alternative methods that are robust to noise.

When there is noise, the sensitivity of recovery to noise depends on $\Omega$. A key quantity that describes this sensitivity is the distance between the two closest points in $\Omega$. This is...
called the *minimum separation*, defined as 

\[ \Delta = \Delta(\Omega) = \min_{1 \leq j < k \leq S} |\omega_j - \omega_k|_T, \]

where \( |\cdot|_T \) is the metric on the torus \( T \). As a manifestation of the Heisenberg uncertainty principle, recovery is sensitive to noise whenever \( \Delta < 1/M \). In the imaging community, \( 1/M \) is typically called the *Rayleigh Length* (RL), and it is regarded as the minimum separation between two point sources that a standard imaging system can resolve [15]. The super-resolution factor (SRF) is \( M/\Delta \), which is the maximum number of points in \( \Omega \) that is contained in \( 1 \text{ RL} \).

Prior mathematical work on the super-resolution problem can be divided into three primary categories.

(a) The *well-separated case* is when \( \Delta \geq 1/M \), and in which case, we say that \( \Omega \) is well-separated. The majority of prior work deals with this case, and we have at our disposal several polynomial-time algorithms that provably recover an accurate approximation of \( \mu \). These methods include total variation minimization (TV-min) [11, 30, 4, 20, 38], greedy algorithms [19, 28], and subspace methods [34, 24, 40, 39, 41, 45, 48]. These important results address the discretization error and basis mismatch issues [27, 28, 13] that arise in sparse recovery and compressed sensing [12, 18]. However, they are not necessarily applicable to super-resolution imaging if they cannot deal with separation below the Rayleigh length.

(b) The *super-resolution regime* is when \( \Delta < 1/M \). This situation is called super-resolution because \( \mu \) consists of Dirac masses that are separated below the classical resolution limit. There are two main approaches to this problem. Optimization-based methods such those found in [43, 10] require the restrictive assumption that \( \mu \) is positive. Likewise, although the optimization method found in [8] allows for complex \( \mu \), it requires that the sign of \( \mu \) equals a dual polynomial on the support of \( \mu \).

Alternatively, *subspace methods* exploit a low-rank factorization of the data and can recover complex measures, but there are many unanswered questions related to its stability to noise that we would like to address. The focus of this paper is a particular subspace method called MUltiple SIgnal Classification (MUSIC) [48].

(c) Prior works [17, 14] addressed super-resolution from an information theoretic view. They considered the situation where the point sources are located on a grid on \( \mathbb{R} \) with spacing \( 1/N \) and the given information consists of noisy continuous Fourier measurements. They both derived lower and upper bounds for a min-max error. These results are asymptotic as the grid spacing needs to be sufficiently small and the constants in the bounds are not explicitly given.

1.3 Motivation

The material in this paper is motivated by two important questions: What are the fundamental limits of super-resolution and what is the resolution limit of subspace methods? Our
Figure 2: Let $M = 50$ and consider the three sets, $\Omega_1 = \{0, 0.01, 0.02, 0.03, 0.04\}$, $\Omega_2 = \{0, 0.01, 0.02, 0.4, 0.5\}$, and $\Omega_3 = \{0, 0.01, 0.3, 0.4, 0.5\}$. For each $0.2 \leq \varepsilon \leq 1$, let $\varepsilon \Omega_j$ be the set obtained by multiplying each entry of $\Omega_j$ by $\varepsilon$. Each $\varepsilon \Omega_j$ has the same cardinality of $S = 5$ and minimum separation of $\Delta_\varepsilon = \varepsilon/100$. However, the functions $\varepsilon \mapsto \sigma_{\text{min}}(\Phi(\varepsilon \Omega_j, M))$ decrease exponentially at different rates.

(a) To measure the fundamental difficulty of super-resolution, one can define a min-max error, which is the reconstruction error incurred by the best possible algorithm in the worst case scenario. Donoho proposed a min-max error for the recovery of point sources on a grid of $\mathbb{R}$ with continuous Fourier measurements, and established lower and upper bounds on a min-max error [17]. These estimates were improved by Demanet and Nguyen in the case that the point sources are sparse [14]. A common theme of their work is that the min-max error can be controlled in terms of the invertibility of a Fourier-like operator.

Our work is motivated by a wide variety of applications in imaging and signal processing where discrete Fourier coefficients are measured by sensors and the measure is supported in a finite interval as opposed to the real line. Assuming that the measure is on a grid of the torus with spacing below $1/N$, we define a min-max error for this model. By using the techniques developed in [17, 14], we connect the min-max error for our model to $\sigma_{\text{min}}(\Phi(\Omega, M))$ for the worst subset $\Omega$ of the grid.

(b) Motivated by imaging applications, there is much interest in developing super-resolution methods. An important problem is to understand their super-resolution limit, which we define to be the relationship between the geometry of the support set $\Omega$ and the noise level for which such methods can guarantee a stable recovery of all point sources. It has been experimentally observed that a collection of algorithms called subspace methods are effective at solving the super-resolution problem. Specific examples include Multiple Signal Classification (MUSIC) [48], Estimation of Signal Parameters via Rotational
Invariance Technique (ESPRIT) [45], and the Matrix Pencil Method (MPM) [34]. Even though these algorithms were introduced several decades ago, they are still widely used today due to their superior numerical performance.

While the theoretical works [40, 41, 26] studied the performance of subspace methods in the well-separated case, their super-resolution capabilities are unknown. The resolution limit of MUSIC was discovered by numerical experiments in [40], but a rigorous justification is lacking. One goal of this paper is to prove the resolution limit of MUSIC. As a consequence of Wedin’s theorem, the sensitivity of the MUSIC algorithm to noise obeys, in a very informal manner,

\[
\text{Sensitivity} \leq \frac{\text{Constant}}{\sigma_{\min}^2 (\Phi)} \cdot Q(\eta),
\]

where \( Q(\eta) \) is a quantity depending on noise. Thus, MUSIC can accurately estimate the support of the measure provided that the noise term is sufficiently small compared to the noise amplification factor which depends crucially on \( \sigma_{\min} (\Phi) \).

Having highlighted the key role that \( \sigma_{\min} (\Phi) \) plays in understanding both the min-max error and the resolution limit of MUSIC, we need to obtain an accurate lower bound for \( \sigma_{\min} (\Phi) \). In the well-separated case, accurate estimates for \( \sigma_{\min} (\Phi) \) and \( \sigma_{\max} (\Phi) \) are known. By using properties of the Beurling-Selberg majorant function, see [42, 53] for an overview, Moitra [41] proved that

\[
\sigma_{\max} (\Phi) \leq (M + \Delta^{-1})^{1/2} \quad \text{and} \quad \sigma_{\min} (\Phi) \geq (M - \Delta^{-1})^{1/2} \quad \text{if} \quad \Delta > \frac{1}{M}.
\]

Such inequalities provide a stability analysis for subspace methods in the well-separated case. Similar approaches were considered in [40, 24].

In the super-resolution regime, \( \sigma_{\min} (\Phi) \) is extremely sensitive to the “geometry” or configuration of \( \Omega \) whenever \( \Delta < 1/M \). To support this assertion, Figure 2 provides examples of three sets that have the same minimum separation and cardinalities, but the minimum singular values of their associated Vandermonde matrices have drastically different behaviors. This simple numerical experiment demonstrates that it is impossible to accurately describe \( \sigma_{\min} (\Phi) \) solely in terms of \( \Delta \), and that a more sophisticated description of the “geometry” of \( \Omega \) is required. This observation motivates us to define a model where \( \Omega \) consists of well-separated subsets, where each subset contains several points that can be closely spaced. This situation occurs naturally in applications where we would like to resolve point sources that are clustered in far apart sets.

### 1.4 Contributions

This paper encapsulates three main topics: accurate lower bounds for the minimum singular value of Vandermonde matrices under geometric assumptions, improvements to the min-max error of super-resolution by sparsity constraints, and application of these results to the MUSIC algorithm.
1.4.1 Minimum singular value of Vandermonde matrices

One of the major difficulties with estimating $\sigma_{\text{min}}(\Phi)$, as well as developing a rich theory of super-resolution, is incorporating geometric information about $\Omega$. An important open question in the super-resolution theory is to analyze $\sigma_{\text{min}}(\Phi)$ when $\Omega$ is heterogenous. Consider the case when $\Omega$ is of the following form.

**Definition 1** (Localized clumps). A set $\Omega \subseteq \mathbb{T}$ consists of $A$ **localized clumps** if $\Omega$ can be written as the disjoint union of $A$ sets,

$$\Omega = \bigcup_{a=1}^{A} \Lambda_a,$$

where each **clump** $\Lambda_a$ is contained in an open interval of length $1/M$ and

$$\text{dist}(\Lambda_m, \Lambda_n) = \min_{\omega_j \in \Lambda_m, \omega_k \in \Lambda_n} |\omega_j - \omega_k|_T > \frac{1}{M}.$$  

Let $\lambda_a = |\Lambda_a|$.

To accurately analyze the behavior of $\sigma_{\text{min}}(\Phi)$ under the localized clumps model, we must consider both the intra-clump and inter-clump distances. We shall see that the following notion quantifies the local geometry of $\Omega$.

**Definition 2** (Complexity). The **complexity** at $\omega_j \in \Omega$ is the quantity,

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

If $\Omega$ consists of $A$ localized clumps that are sufficiently far apart from each other, then we intuitively expect $\sigma_{\text{min}}(\Phi)$ to be an $\ell^2$ aggregate of $A$ terms, where each term only depends on the local geometry. Theorem 1 precisely quantifies this intuition. Assume that the clumps $\{\Lambda_a\}_{a=1}^{A}$ are sufficiently far apart from each other depending on the complexities of $\Omega$. Then there exist explicit constants $\{B_a\}_{a=1}^{A}$ such that

$$\sigma_{\text{min}}(\Phi) \geq \sqrt{M} \left( \sum_{a=1}^{A} \sum_{\omega_j \in \Lambda_a} (B_a \lambda_a^{-1} \rho_j)^2 \right)^{-1/2}. \quad (1.6)$$

This inequality is quite general because it holds for any $\Omega$ consisting of localized clumps that are sufficiently far apart. One might wonder what this bound reduces to under more restrictive assumptions.

**Definition 3** (Localized clumps with separation). A set $\Omega \subseteq \mathbb{T}$ consists of $A$ **localized clumps with separation parameter** $\alpha$ if $\Omega$ satisfies the localized clumps model and $\Delta \geq \alpha/M$. Note that SRF = $1/\alpha$. 

7
Theorem 2 states that if $\Omega$ consists of $A$ localized clumps with separation parameter $\alpha$ and they are separated by at least $20\alpha^{-1/2}M^{-1} \max_{1 \leq a \leq A} \lambda_a^{5/2}$, then there exist explicit constants $\{C_a\}_{a=1}^A$ such that

$$\sigma_{\min}(\Phi) \geq \sqrt{M} \left( \sum_{a=1}^A (C_a \text{SRF}^{\lambda_a^{-1}})^2 \right)^{-1/2}.$$  

(1.7)

This inequality holds non-asymptotically in $\alpha = 1/\text{SRF}$. Notice that the dominant term in this inequality is $\text{SRF}^{-\lambda+1}$, where $\lambda$ is the cardinality of the largest clump. This is in contrast to the exponent that shall appear in min-max rate, and we shall see later on why this is of great importance.

We verify numerically that the exponent on SRF is optimal, which means that $\sigma_{\min}(\Phi)$ depends on the local cardinality of the clumps and not on the total number of point sources. In fact, Proposition 3 proves this assertion under stronger assumptions on $\Omega$. Notice that the constant $C_a$ in (1.7) scales like $\lambda_a^{\lambda_a}$ and might seem wildly inaccurate at first, the numerical experiments verify our theory that the true constant depending on $\lambda_a$ should decay extremely quickly. These simulations are shown in Figure 4.

1.4.2 Worst case analysis and min-max error

In some imaging applications, we have little information about the configuration of $\Omega$. Suppose $\Omega$ has cardinality $S$ and is a subset of $\{n/N\}_{n=0}^{N-1}$, which we refer to as the grid of width $1/N$. In the super-resolution literature, this situation 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 we have $\text{SRF} = N/M$. We can define the $\ell^2$ min-max error for this model as the error incurred by the best recovery algorithm on the worst case signal and noise in the usual way.

**Definition 4** (Min-max error). Fix positive integers $M, N, S$ such that $S \leq M \leq N$ and let $\delta > 0$. Let $\mathcal{Y} = \mathcal{Y}(M, N, S, \delta)$ denote the collection of all $y = y(\mu, \eta) \in \mathbb{C}^{M+1}$ of the form $y_m = \mu(m) + \eta_m$ for $0 \leq m \leq M$, where $\mu$ consists of $S$ Dirac masses supported on the grid of width $1/N$, and the noise $\eta \in \mathbb{C}^{M+1}$ satisfies $\|\eta\|_2 \leq \delta$. Let $\mathcal{A} = \mathcal{A}(M, N, S, \delta)$ be the set of functions $\varphi$ that map each $y \in \mathcal{Y}$ to a discrete measure $\varphi_y$ supported on the grid of width $1/N$. Then, the $\ell^2$ min-max error for the on-the-grid model is

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

In this definition, we interpret a function $\varphi \in \mathcal{A}$ as an algorithm that maps a given signal $y = y(\mu, \eta) \in \mathcal{Y}$ to a measure $\varphi_y$ that approximates $\mu$. 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 limitation of super-resolution under sparsity constraints, and no algorithm can perform better than the min-max rate.
To estimate the min-max error, we follow the approach of [14], and connect it to the worst case minimum singular value of Fourier matrices and then estimate the latter quantity. These steps are carried out in Proposition [4] Proposition [3] and Theorem [3]. By using these results, Theorem [4] provides explicit constants $A(M,S)$ and $B(M,S)$ such that

$$A(M,S) \, \text{SRF}^{2S-1} \, \delta \leq \epsilon(M,N,S,\delta) \leq B(M,S) \, \text{SRF}^{2S-1} \, \delta.$$ 

The dominant factor in both inequalities is $\text{SRF}^{2S-1}$. Hence, without any prior geometric assumptions on $\Omega$, no algorithm can accurately estimate the worst case measure $\mu$ and noise $\eta$, unless $\|\eta\|_2$ is significantly smaller than $\text{SRF}^{2S-1}$.

1.4.3 Super-resolution of MUSIC

Most subspace methods form a Hankel matrix from the measurements $y$, and exploit a Vandermonde decomposition of the Hankel matrix. This paper focuses on the MUSIC algorithm. It amounts to finding the noise space $W$ of the Hankel matrix, forming a noise-space correlation function (or its reciprocal which is called the imaging function), and identifying the $S$ smallest local minima of the noise-space correlation function (or the $S$ largest peaks of the imaging function) as the support set.

When there is no noise, the noise-space correlation function vanishes and the imaging function peaks exactly at the source locations. MUSIC can exactly recover $\Omega$ provided that the number of measurements is at least twice the number of sources: $M + 1 \geq 2S$. When there is noise, the Hankel matrix is contaminated by noise, and its corresponding noise space is perturbed to $\hat{W}$. If the noise-to-signal ratio is low, then $W$ and $\hat{W}$ are similar, and it is possible to obtain an accurate estimate of $\Omega$. Earlier works on subspace methods [40, 26, 41] proved stability of MUSIC, ESPRIT and MPM when the point sources are separated by at least $2/M$ or $2 \text{RL}$.

However, it is an open problem to understand when subspace methods are successful in the super-resolution regime. The numerical experiments in [40] demonstrate a phase transition phenomenon on the relation between the noise level and the separation between the point sources for which MUSIC guarantees a successful recovery. A main contribution of this paper is to provide a rigorous justification on this phase transition phenomenon.

In MUSIC, point sources are localized through the smallest local minima of the noise-space correlation function. We show that the perturbation of this function by noise satisfies:

$$\text{Perturbation of the noise-space correlation} \leq \frac{\text{Noise}}{x_{\min} \sigma_{\min}^2(\Phi)}.$$ 

With this stability bound and our lower bound for $\sigma_{\min}(\Phi)$, we prove that, if $\Omega$ satisfies the localized clumps model with separation parameter $\alpha$ (see Definition [3]), and if noise is i.i.d. Gaussian, $\eta \sim \mathcal{N}(0, \sigma^2 I)$, then the noise-to-signal ratio that MUSIC can tolerate obeys the following scaling law:

$$\frac{\sigma}{x_{\min}} \propto \sqrt{\frac{M}{\log M}} \left( \sum_{a=1}^{A} c_a^2 \alpha^{-2(\lambda_a-1)} \right)^{-1}.$$ 

9
where \( c_a = c_a(\lambda_a, M) \) is given explicitly. In the special case where each \( \Lambda_a \) contains \( \lambda \) equally spaced object with spacing \( \alpha/M \) (see Figure 3), the equation above becomes

\[
\frac{\sigma}{\sigma_{\min}} \propto \sqrt{\frac{M}{\log M}} \alpha^{2\lambda - 2} = \sqrt{\frac{M}{\log M}} \left( \frac{1}{\text{SRF}} \right)^{2\lambda - 2}.
\]

Our result shows that the resolution limit of MUSIC is exponential on \( 1/\text{SRF} \), and importantly, the exponent depends on the cardinality of the localized clumps instead of the total number of point sources. This estimate is verified by numerical experiments in Figure 8.

1.5 Related work

Many works on super-resolution have been mentioned in Section 1.2 and 1.3. Here we will discuss the ones which are most closely related with our main results.

1.5.1 Minimum singular of Vandermonde matrices

The invertibility of a Vandermonde matrix is highly dependent on the configuration of its nodes. For this reason, many papers considered special configurations of the nodes and past works are typically categorized according to the location of their nodes.

This paper is concerned with Vandermonde matrices with complex nodes on the unit circle. For such matrices, it is possible to arrange the nodes in such a way that the matrix has perfect conditioning, see [9] for a characterization. In the same spirit, if the nodes are sufficiently separated, then the Vandermonde matrix is well-conditioned. There are various results of this nature, see [31, 40, 41], which only apply to the case that \( \Delta > 1/M \). When \( \Delta < 1/M \) and the matrix is not too tall, then certain arrangements of the nodes can cause it to be ill-conditioned, see [41].

A significant portion of our work considers the case when the nodes belong to well-separated clumps. Although \( \sigma_{\min}(\Phi) \) is small under such assumptions, one of our main contributions is to provide an accurate estimate for this quantity. In the process of writing the second draft of this paper, the authors of [5], independent of us, derived a lower bound for the \( \sigma_{\min}(\Phi) \) under similar geometric assumptions. We shall make a detailed comparison between their work and our Theorems 1 and 2 in Remark 4.

Other papers such as [6, 3] studied the more general case where the nodes are within the closed unit disk of the complex plane. When specialized to the unit circle, the result in [3] requires that \( \Delta > 1/M \) and their lower bound for \( \sigma_{\min}(\Phi) \) is similar to the one in [41]. If \( M + 1 \) is an integer multiple of \( S \), then [6, Theorem 1] lower bounds \( \sigma_{\min}(\Phi) \) in terms of the spectral norm of the inverse of a square Vandermonde matrix; an upper bound for this latter quantity can be found in [32, Theorem 1]. Combining these results, they yield the inequality,

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

If we consider the case that \( \Omega \) satisfies either the localized clumps or grid model, this inequality is significantly worse than the ones presented in our theorems.
Finally, there has been work on estimating the minimum singular value of Vandermonde matrices for other configurations. When the nodes are real, the matrix is usually ill-conditioned, see [33, 7, 21, 22]. Randomly selected nodes were treated in [47, 50, 51], but this approach does shed light on how $\sigma_{\text{min}}(\Phi)$ depends on the configuration of $\Omega$.

1.5.2 Super-resolution limit

The min-max formulation of the super-resolution problem can be traced back to [17, 14]. The authors of those papers defined and studied the min-max error for measures supported on a grid of the real line $\mathbb{R}$ with continuous measurements. The main difference is that [14] considered measures with $S$ Dirac masses, whereas [17] allowed for a broader class of discrete measures, including those with an infinite number of Dirac masses. The former obtained the exact dependence of the min-max error on the super-resolution factor.

In contrast to these works, we studied a min-max error for discrete measures with $S$ Diracs supported on a grid of the torus with discrete measurements. Our model is more similar to the one in [14]. Even though the min-max error for the discrete case, see Theorem 4, and the min-max error for the continuous models [17, 14] are strikingly similar, their results do not imply ours or vice versa. For our model, we obtain non-asymptotic and explicit estimates for the min-max error, while the inequalities in [17, 14] require sufficiently small grid spacing and the constants are not explicitly given.

1.5.3 Stability of subspace methods

Subspace methods are well known in imaging and signal processing due to their superior numerical performance. Earlier works addressed the stability of MUSIC [40], ESPRIT [26, 2], and MPM [41] when all point sources are separated by $1/M$. This paper focuses on the MUSIC algorithm and proves its resolution capability when the minimum separation is below $1/M$, i.e. $\Delta < 1/M$. This case is more interesting since MUSIC is well known for its super-resolution phenomenon.

1.6 Outline

The remainder of this paper is organized as follows. Section 2 includes our main results on the minimum singular value of Vandermonde matrices and the min-max error. It also describes our general approach for deriving these estimates and contains several essential propositions. We also include numerical experiments highlighting the accuracy of our estimates. Section 3 fully explains the MUSIC algorithm, and it includes new stability results for MUSIC in the super-resolution regime. We include numerical simulations, and they validate our theoretical results. Appendices A, B, and C contains the proofs for all the theorems, propositions, and lemmas, respectively.
2 Minimum singular value of Vandermonde matrices

2.1 Duality and interpolation

Our method for estimating $\sigma_{\text{min}}(\Phi)$ is through a dual characterization. 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 $\omega \in \mathbb{T}$,

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

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

**Definition 5** (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 $1 \leq j \leq S$.

The polynomial interpolation set is non-empty whenever $S \leq M - 1$. This is an immediate consequence of the existence of the Lagrange interpolating polynomials. We have the following duality between the minimum singular value of Fourier matrices and the polynomial interpolation set.

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

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

While we suspect that this simple proposition had been previously discovered, we are unable to find a reference. Perhaps we could not find the proposition in the literature because, for the following reasons, it is not immediately clear that this proposition is useful for estimating $\sigma_{\text{min}}(\Phi)$.

(a) 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 $\Omega$ because we know that $\sigma_{\text{min}}(\Phi)$ is highly sensitive to the configuration of $\Omega$.

(b) We do not know anything about $v$, a right singular vector corresponding to $\sigma_{\text{min}}(\Phi)$.

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. We have a relaxed version of exact duality, which will provide us with an extra bit of flexibility.

**Proposition 2** (Robust duality). Fix positive integers $M$ and $S$ such that $S \leq M$. For any set $\Omega = \{\omega_j\}_{j=1}^{S} \subseteq \mathbb{T}$, unit norm vector $u \in \mathbb{C}^S$, and $\varepsilon \in \mathbb{C}^S$ with $\|\varepsilon\|_2 < 1$, if there exists $f \in \mathcal{P}(M)$ such that $f(\omega_j) = u_j + \varepsilon_j$ for each $1 \leq j \leq S$, then

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

12
The proofs of Propositions 1 and 2 can be found in Appendix B.1 and B.2 respectively. In order to use these results to derive a lower bound for $\sigma_{\min}(\Phi)$ for a given $\Omega$, for each unit norm $v \in \mathbb{C}^S$, we construct a $f_v \in P(\Omega, M, v)$ and then bound $\|f_v\|_{L^2(T)}$ uniformly in $v$. This process must be done carefully; otherwise we would obtain a lousy lower bound for $\sigma_{\min}(\Phi)$. This construction is technical and our approach is inspired by uncertainty principles for trigonometric polynomials [52] and uniform dilation problems on the torus [1, 35].

2.2 Upper bound on the minimum singular value

Before we proceed to obtain lower bounds for $\sigma_{\min}(\Phi)$ for various $\Omega$, it might it be helpful to obtain an upper bound on $\sigma_{\min}(\Phi)$ for specific examples to gain some intuition for what we expect. The following proposition carries out this calculation using a method similar to one found in [17]. Its proof can be found in Appendix B.3.

Proposition 3. Fix positive integers $M, S, \lambda$ such that $\lambda \leq S \leq M - 1$. Let $\omega \in \mathbb{T}$ and $\alpha > 0$ such that

$$\alpha \leq \frac{1}{C(\lambda)\sqrt{M + 1}}.$$

where $C(\lambda) = 2\pi \sum_{j=0}^{\lambda-1} \binom{\lambda - 1}{j} \frac{j^\lambda}{\lambda!}$. (2.1)

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 $\Omega$ and $M$. Then

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

Here, $\sqrt{M + 1}$ is the natural scaling factor since each column of $\Phi$ has $\ell^2$ norm $\sqrt{M + 1}$. This proposition shows that if $\Omega$ contains a set $\Lambda$, which consists of $\lambda$ points equispaced by $\alpha/M$ for a sufficiently small $\alpha$, then $\sigma_{\min}(\Phi)$ should depend on the super-resolution factor $\text{SRF} = 1/\alpha$ and the local cardinality of $\Omega$. This motivates us to derive a tight lower bound of the same order.

2.3 Localized clumps and the geometric model

We consider the case where $\Omega$ satisfies the localized clumps model (Definition 1), and we derive a lower bound for $\sigma_{\min}(\Phi)$ in terms of a weighted $\ell^2$ aggregate of the contribution from each clump. The following theorem is proved in Appendix A.1.

Theorem 1. Fix positive integers $M$ and $S$ with $M \geq 2S^2$. Suppose $\Omega = \{\omega_j\}_{j=1}^S \subseteq \mathbb{T}$ consists of $A$ localized clumps $\{\Lambda_a\}_{a=1}^A$. If $A > 1$, assume that

$$\min_{m \neq n} \text{dist}(\Lambda_m, \Lambda_n) \geq \max_{1 \leq a \leq A} \max_{\omega_j \in \Lambda_a} \frac{10\lambda_a^{5/2} (S\rho_j)^{1/(2\lambda_a)}}{M}. \quad (2.2)$$


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

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

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

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

**Remark 1.** Although the quantity $B_a$ depends on both $\lambda_a$ and $M$, it is insensitive to the geometry of each $\Lambda_a$ because it only depends on $\lambda_a$, the cardinality of the clump $\Lambda_a$. Further, we can think of $B_a$ as a small universal constant because rounding becomes increasingly negligible as $M/\lambda_a$ increases, and 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 very quickly as $n$ increases. Thus, in the regime where each $\lambda_a$ is of moderate size and $M/\lambda_a$ is large, we can think of $B_a$ as being approximately $20\sqrt{2}/19 \approx 1.4886$.

**Remark 2.** Although this is not the main point of Theorem 1, we can apply it to the well-separated case. Assume that $\Delta \geq 10S^{1/2}/M$. Then each clump $\Lambda_a$ 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 the theorem 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 $\Delta$ is about $\sqrt{S}$ times larger than $1/M$. This result is weaker than the one obtained in [41], which was derived using tools that specialized to the well-separated case. Note that $\sqrt{M}$ is approximately the largest $\sigma_{\min}(\Phi)$ can be because $\sigma_{\max}(\Phi) \leq \|\Phi\|_F = \sqrt{MS}$, where $\|\cdot\|_F$ is the Frobenius norm.

Theorem 1 provides us with a lower bound for $\sigma_{\min}(\Phi)$ in terms of the complexities of $\Omega$. One might wonder what the bound reduces to in a more concrete situation. Suppose that $\Omega$ consists of $A$ localized clumps, but additionally, we assume that the distances between points in $\Omega$ is at least $\alpha/M$ for some $0 < \alpha < 1$. Note that SRF = $1/\alpha$. Then we can upper bound the complexities of each $\omega_j$ and determine the sufficient inter-clump separation. We prove the following theorem in Appendix A.2.

**Theorem 2.** Fix positive integers $M$ and $S$ with $M \geq S^2$. Suppose $\Omega = \{\omega_j\}_{j=1}^{S} \subseteq \mathbb{T}$ consists of $A$ localized clumps $\Lambda_a \subseteq \mathbb{T}$ and $\Delta(\Omega) \geq \alpha/M$, where

$$\max_{1 \leq a \leq A} (\lambda_a - 1) < \frac{1}{\alpha}. \quad (2.4)$$

If $A > 1$, assume that

$$\min_{m \neq n} dist(\Lambda_m, \Lambda_n) \geq \max_{1 \leq a \leq A} \frac{20S^{1/2} \lambda_a^{5/2}}{\alpha^{1/2} M}. \quad (2.5)$$
For each $1 \leq a \leq A$, let $B_a = B_a(\lambda_a, M)$ be the constant defined in Theorem 1 and

$$C_a = C_a(\lambda_a, M) = B_a \left( \frac{\lambda_a}{\pi} \right)^{\lambda_a - 1} \left( \sum_{j=1}^{\lambda_a} \prod_{k=1, k \neq j}^{\lambda_a} \frac{1}{(j - k)^2} \right)^{1/2}. \quad (2.6)$$

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

$$\sigma_{\min}(\Phi) \geq \sqrt{M} \left( \sum_{a=1}^{A} \left( C_a \alpha^{-\lambda_a + 1} \right)^2 \right)^{-1/2}. \quad (2.7)$$

**Remark 3.** We would like to compare the assumptions and statements of Theorems 1 and 2. The latter is more concrete since it bounds $\sigma_{\min}(\Phi)$ in terms of $\alpha = 1/\text{SRF}$, but it is less accurate. Suppose each clump $\Lambda_a$ consists of $\lambda_a$ points equispaced by $\alpha/M$. Then the lower bounds (2.3) and (2.7) are identical. For all other configurations of $\Omega$, the estimate (2.3) is more accurate than (2.7). The separation condition (2.2) is always weaker than (2.5). As mentioned in Remark 1, the constant $B_a$ weakly depends on $\lambda_a$. This is not the case for $C_a$, which for large $\lambda_a$, scales like $\lambda_a^\lambda - 1$. This discrepancy arises because $\rho_j$ is significantly different when $\lambda_a$ is large.

The main contribution of this theorem is the exponent on $\text{SRF} = 1/\alpha$, which depends on the cardinalities of the clumps $\Lambda_a$ as opposed to the total number of points $S$. As an example, let us look at a special case of $\Omega$, where each clump $\Lambda_a$ contains $\lambda$ points consecutively spaced by $\alpha/M$ and the distance between clumps is $\beta/M$ where $\beta$ is chosen appropriate so that (2.2) holds. See Figure 3 for an illustration.

![Figure 3: An example of $\Omega$ that consists of $A$ localized clumps. Each $\Lambda_a$ contains $\lambda$ points $(\lambda = 3$ shown in the figure) spaced consecutively by $\alpha/M$. The distance between clumps is $\beta/M$ for some sufficiently large $\beta \geq 1.$](image)

In this case, Theorem 2 implies

$$\sigma_{\min}(\Phi(\Omega, M)) \geq C(\lambda) A^{-1/2} \sqrt{M} \cdot \left( \frac{\alpha^{\lambda - 1}}{(1/\text{SRF})^{\lambda - 1}} \right)^{\lambda - 1}, \quad (2.8)$$

where the constant $C(\lambda)$ depends only on $\lambda$ and its explicit form is given by (2.6). 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 $\lambda$ is the cardinality of each clump instead of the total number of points $S$, which matches our intuition that the conditioning of $\Phi(\Omega, M)$ should only depend on how complicated each individual clump is.
Remark 4. In the process of revising the first draft of this manuscript, the authors of \([5]\), independent of our work, used different techniques to derive lower bounds for \(\sigma_{\min}(\Phi)\) when \(\Omega\) consists of clumps, see \([5, \text{Definition 1.1}]\) for their model. We point out the differences between their \([5, \text{Corollary 1.1}]\) and our Theorem 2.

(a) They assume that \(\Omega\) 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.

(b) They require the aspect ratio \(M/S\) of the Vandermonde matrix \(\Phi\) to be at least \(4S^2\), whereas we only need at least \(2S\). They also require an upper bound on \(M/S\), which prohibits their Vandermonde matrix from being too tall.

(c) If \(\lambda\) is the cardinality of the largest clump in \(\Omega\), then they obtained the inequality \(\sigma_{\min}(\Phi) \geq C_S \sqrt{M \cdot \text{SRF}}^{-\lambda+1}\) for some \(C_S\) depending only on \(S\) which scales like \(S^{-2S}\). Our implicit constant is more complicated, but it scales like \(A^{-1/2} \lambda^{-\lambda}\).

2.4 Min-max error and worst case analysis

To estimate the min-max error \(E(M, N, S, \delta)\), we introduce the following quantity, which can be interpreted as the worst case minimum singular value of Fourier matrices.

Definition 6 (Lower restricted isometry constant). Fix positive integers \(M, N, S\) such that \(S \leq M \leq N\). The lower restricted isometry constant of order \(S\) is the quantity

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

The minimum is taken over all \(\Omega\) supported in the grid of width \(1/N\) and of cardinality \(S\).

Remark 5. This quantity is related to the lower bound of the \(S\)-restricted isometry property (RIP) from compressive sensing \([10]\), but with a major difference. Indeed, it is known that if we randomly select \(M\) rows of the \(N \times N\) Discrete Fourier Transform (DFT) matrix for appropriately chosen \(M\), then every \(M \times S\) sub-matrix is well-conditioned, see \([46]\). However, \(\Phi(\Omega, M)\) uses the first \(M\) rows of the DFT matrix, so it is possible for \(\Phi(\Omega, M)\) to 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 \([14]\). We borrow their ideas and their proof carries over to this discrete setting with minor modifications. To keep this paper self-contained, we prove this proposition in full detail in Appendix \([\text{B.4}]\).

Proposition 4 (Min-max error and lower restricted isometry constant). Fix positive integers \(M, N, S\) such that \(2S \leq M \leq N\), and let \(\delta > 0\). Then,

\[
\frac{\delta}{2\Theta(M, N, 2S)} \leq E(M, N, S, \delta) \leq \frac{2\delta}{\Theta(M, N, 2S)}.
\]
One of the key advantages of the on-the-grid model lies in Proposition \[4\]. With this result at hand, our main focus is on the derivation of a lower bound for $\Theta(M, N, S)$. From our earlier experience with the localized clumps models, we suspect that the minimum in the definition of $\Theta(M, N, S)$ occurs when all the points in $\Omega$ are consecutively spaced by $1/N$. If we could prove this, then we could simply apply Theorem \[2\] for the single clump case and we would be done. Without a result of this nature, we face a major technical difficulty: since we have no control over which $\Omega$ attains the minimum in $\Theta(M, N, S)$, in order to appeal to our duality results, we must construct trigonometric polynomials for all $\binom{N}{S}$ possible choices of $\Omega$ and then uniformly bound over all these possibilities. There are exponentially many possible such $\Omega$, 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.** Fix positive integers $M, N, S$ such that $S \geq 2$, $M \geq 2S$, and $N \geq \pi MS$. We define the constant,

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

Then we have

$$\Theta(M, N, S) \geq C(M, S) \sqrt{M} \left(\frac{M}{N}\right)^{S-1}. \quad (2.9)$$

**Remark 6.** To see where these numbers come from, observe that the assumption that $N \geq \pi MS$ implies that $\Omega_*$ 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

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

We are ready to derive lower and upper bounds for the min-max error. The proof of the following theorem can be found in Appendix \[A.4\].

**Theorem 4.** 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 \[3\]. Then, we have the upper bound,

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

(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 \[3\]. Then, we have the lower bound,

$$\mathcal{E}(M, N, S, \delta) \geq \frac{\delta}{4} \left(\frac{4S - 2}{2S - 1}\right)^{1/2} \frac{1}{\sqrt{M+1}} \left(\frac{N}{2\pi M}\right)^{2S-1}.$$
Figure 4: Plots of $\zeta$, $\xi$, and their ratio $\zeta/\xi$, as functions of SRF = $1/\alpha$, when the other parameters are fixed. For all the curves, we set $M = 2^{15}$. The set $\Omega$ consists of $A$ clumps, where each clump contains $\lambda$ points equispaced by $\alpha/M$. We consider the following range of parameters: $1 \leq A \leq 3$, $2 \leq \lambda \leq 5$, and $\lambda \leq \text{SRF} \leq 8$. The clumps are separated by $\beta/M$, where $\beta = 20S^{1/2}\lambda^{5/2}\alpha^{-1/2}$.

2.5 Numerical accuracy of the lower bounds

2.5.1 Accuracy of Theorems 1 and 2

To numerically evaluate the accuracy of the lower bounds in Theorems 1 and 2, we consider the case where $\Omega$ consists of $A$ clumps, each clump contains $\lambda$ equispaced points by $\alpha/M$, and the clump separation is $\beta/M$, see Figure 3 for an illustration. The parameter $\beta$ is chosen sufficiently large so that the separation condition (2.5) holds. As discussed in Remark 3, both theorems provide the identical lower bound for $\sigma_{\min}(\Omega)$ in this situation. For convenience, let $\zeta(A, \lambda, \alpha) = \sigma_{\min}(\Phi)$ and let $\xi(A, \lambda, \alpha, M)$ be our estimate in the right hand side of the inequality (2.8). The theorems state that

$$\zeta(A, \lambda, \alpha, M) \geq \xi(A, \lambda, \alpha, M).$$

The figures in the top row in Figure 4 plot $\xi$ and $\zeta$ as functions of SRF = $1/\alpha$, for each $A, \lambda, \alpha$. Observe that the slopes for $\xi$ and $\zeta$ are identical, which validates our prediction that there exists a constant $c(\lambda) > 0$ such that

$$\sigma_{\min}(\Phi) = c(\lambda)\sqrt[M]{M}\alpha^{-1}. \quad (2.10)$$

In other words, $\sigma_{\min}(\Phi)$ should only depend on the cardinality of each clump and not on
the total number of points $S$. We also see that $\zeta$ appears to be independent of the total number of clumps $A$, whereas $\xi$ has a mild $\sqrt{A}$ dependence.

To determine the accuracy of Theorems 1 and 2 we can examine the ratio,

$$\frac{\zeta(A, \lambda, \alpha, M)}{\xi(A, \lambda, \alpha, M)}.$$  

This is shown on the bottom row of Figure 4. The ratio is almost horizontal as a function of the SRF, which again, shows that our lower bound captures the true dependence of $\zeta$ on SRF. Further, the ratio is bounded above by a reasonable constant for our various choices of parameters, which indicates that our the estimate provided by the theorems is accurate.

Notice that the ratio grows as $\lambda$ increases (when $\lambda \geq 3$) and that the constant $C(\lambda)$ in (2.8) scales like $\lambda^{-\lambda+1}$, which is super exponentially decaying in $\lambda$. The numerical situations suggest that $c(\lambda)$ in (2.10) should decay quickly in $\lambda$, which is consistent with our estimate.

2.5.2 Accuracy of Theorem 3

Let $\theta(M, N, S)$ denote the right hand side of (2.9), the lower bound for $\Theta(M, N, S)$ given in Theorem 3. Note that $\theta$ is only defined with $N/M \geq \pi S$. To numerically evaluate the accuracy of the theorem, we make two observations.

(a) While we would like to compare $\theta$ directly with $\Theta$, this is not possible because we would need to enumerate through all possible $\Omega$, for numerous values of $M, N, S$. Instead, we compare $\theta$ with the quantity,

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

Recall that $\Omega^*$ denotes the set with $S$ consecutively spaced points separated by distance $1/N$. Note that $\theta_s$ serves as a useful substitute for $\Theta$ because of the inequality

$$1 \leq \frac{\Theta(M, N, S)}{\theta(M, N, S)} \leq \frac{\theta_s(M, N, S)}{\theta(M, N, S)}.$$  

(b) While both $\theta$ and $\theta_s$ depend on three parameters, Theorem 3 and Proposition 3 suggest that they should only depend on two parameters, the super-resolution factor 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.

Figure 5 displays the values of $\theta$ and $\theta_s$ as functions of SRF as well as their ratios. The left figure suggests that, if SRF $\geq 2$, then for some unknown $c_*(M, S) > 0$,

$$\theta_s(M, N, S) = c_*(M, S) \sqrt{M} \left( \frac{M}{N} \right)^{S-1}.$$  

The right figure displays the ratio between $\theta_s$ and $\theta$. As a consequence of inequalities (2.11), this experiment also indirectly provides us information about the ratio of $\Theta$ and $\theta$. The lines in the right figure are horizontal, which confirms our theory that, if SRF $\geq \pi S$, then there exists an unknown constant $c(M, S) > 0$ such that

$$\Theta(M, N, S) = c(M, S) \sqrt{M} \left( \frac{M}{N} \right)^{S-1}.$$
Figure 5: The left figure displays the behavior of $\theta(M, N, S)$ and $\theta_*(M, N, S)$ as functions of SRF = $N/M$, where $S = 2, 3, 4, 5$ and $\pi S \leq$ SRF $\leq 20$ and $2 \leq$ SRF $\leq 20$, respectively. The right figure displays the ratio of $\theta_*(M, N, S)$ and $\theta(M, N, S)$, which quantifies the accuracy of the lower bound in Theorem 3.

The figure also provides us with information about the behavior of $c_*(M, S)$ and $c(M, S)$. As seen in the figure, $\theta_*/\theta$ grows in $S$. Observe that the implicit constant $C(M, S)$ in (2.9) scales like $S^{-S+1/2}$, which is super-exponentially decaying in $S$. Thus, this experiment shows that $c(M, S)$ and $c_*(M, S)$ also decay rapidly in $S$.

3 MUSIC and its super-resolution limit

Many interests in imaging center on inventing super-resolution algorithms and understanding the resolution limit of these algorithms. In signal processing, a class of subspace methods, including MUSIC [48], has been widely used due to their superior numerical performance. It was well known that MUSIC has super-resolution phenomenon, i.e. the capability of resolving point sources separated below RL. The resolution limit of MUSIC was discovered by numerical experiments in [40], but has never been rigorously proved. By resolution limit we mean the relation between the geometry of the support set $\Omega$ and the noise level for which the recovery of all point sources is possible. A main contribution of this paper is to prove the resolution limit of MUSIC.
Algorithm 1 MUltiple SIgnal Classification (MUSIC)

Input: $y \in \mathbb{C}^{M+1}$, sparsity $S$, $L$
1: Form Hankel matrix $\mathcal{H}(y) \in \mathbb{C}^{(L+1) \times (M-L+1)}$
2: Compute the SVD of $\mathcal{H}(y)$:

$$\mathcal{H}(y) = \begin{bmatrix} \hat{U} & \hat{W} \end{bmatrix}_{(L+1) \times (L+S)} \begin{bmatrix} \text{diag}(\hat{\sigma}_1, \ldots, \hat{\sigma}_S, \hat{\sigma}_{S+1}, \ldots) \end{bmatrix}_{(L+1) \times (M-L+1)} \begin{bmatrix} \hat{V}_1 & \hat{V}_2 \end{bmatrix}_{(M-L+1) \times (M-L+1-S)}^*$$

where $\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \ldots \geq \hat{\sigma}_S \geq \hat{\sigma}_{S+1} \geq \ldots$ are the singular values of $\mathcal{H}(y)$.
3: Compute the imaging function $\hat{J}(\omega) = \| \phi_L(\omega) \|_2 / \| \hat{W}^* \phi_L(\omega) \|_2$, $\omega \in [0, 1)$.

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

3.1 Hankel matrix and Vandermonde decomposition

Most subspace methods are built upon a Hankel matrix and its Vandermonde decomposition. Fixing a positive integer $L \leq M$, we form the Hankel matrix of $y$:

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

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

$$\mathcal{H}(y^0) = \Phi_L X \Phi_M^T,$$

where $X = \text{diag}(x) \in \mathbb{R}^{S \times S}$. We always assume that the number of measurements satisfies $M + 1 \geq 2S$ and $L$ is chosen such that $L \geq S$ and $M - L + 1 \geq S$, in which case $\Phi_L$ and $\Phi_M$ have full column rank, and $\mathcal{H}(y^0)$ has rank $S$. We will next introduce the MUSIC algorithm and prove its super-resolution limit.

3.2 The MUSIC algorithm

The MUSIC algorithm was proposed by Schmidt [48]. It amounts to finding the noise space of the Hankel matrix, forming a noise-space correlation function (or its reciprocal which is called the imaging function), and identifying the $S$ smallest local minima of the noise-space correlation function (or the $S$ peaks of the imaging function) as the support set.

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

$$\mathcal{H}(y^0) = \begin{bmatrix} U & W \end{bmatrix}_{L \times S} \begin{bmatrix} \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_S, 0, \ldots, 0) \end{bmatrix}_{L \times (M-L+1)} \begin{bmatrix} V_1 & V_2 \end{bmatrix}_{(M-L+1) \times (M-L+1-S)}^*$$

(3.2)
where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_S$ are the non-zero singular values of $\mathcal{H}(y^0)$. Here the column spaces of $U$ and $W$ are exactly equal to $\text{Range}(\mathcal{H}(y^0))$ and $\text{Range}(\mathcal{H}(y^0))^\perp$ respectively, which are called the signal space and the noise space.

For any $\omega \in [0, 1)$ and positive integer $L$, we define the steering vector of length $L + 1$ at $\omega$ 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 $\Phi_L = [\phi_L(\omega_1) \ldots \phi_L(\omega_S)]$. MUSIC was proposed based on the following observation on the Vandermonde structure of $\Phi_L$: if $L \geq S$ and $M - L + 1 \geq S$, then

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

Figure 6: Plots of the noise-space correlation functions $R(\omega)$, $\hat{R}(\omega)$ and the imaging functions $J(\omega)$, $\hat{J}(\omega)$. Three sources with complex coefficients of unit magnitude are separated by 0.5 RL. The true source locations are represented by the red dots in all plots. In the noiseless case, $R(\omega)$ vanishes and $J(\omega)$ peaks exactly at the true support. When $\sigma = 0.001$, the imaging function $\hat{J}(\omega)$ still peaks around the true support, which is not the case when the noise level increases to $\sigma = 0.01$.

We define a noise-space correlation function $R(\omega)$ and let the imaging function be its reciprocal (see Table 1 for definitions). The following lemma is a basis for the MUSIC algorithm in the noiseless case.

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

$$
\omega \in \{\omega_j\}_{j=1}^S \iff R(\omega) = 0 \iff J(\omega) = \infty.
$$
### Table 1: Noise-space correlation functions and imaging functions in MUSIC

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

In the noiseless case, source locations can be exactly identified through the zeros of the noise-space correlation function \( R(\omega) \) or the peaks of the imaging function \( J(\omega) \) (see Figure 6 for an example), as long as the number of measurements is at least twice the number of point sources to be recovered.

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

\[
H(y) = H(y^0) + H(\eta)
\]

whose SVD is given by (3.1) in Algorithm 1. The noise-space correlation function and the imaging function are perturbed to \( \hat{R}(\omega) \) and \( \hat{J}(\omega) \) respectively (see Table 1 for the definitions).

### 3.3 Stability of MUSIC

Figure 6 shows that, the imaging function \( \hat{J}(\omega) \) still peaks around the true sources, as long as the noise-to-signal ratio is low enough. However, MUSIC can fail when the noise-to-signal ratio increases. Stability of MUSIC depends on the perturbation of the noise-space correlation function from \( R(\omega) \) to \( \hat{R}(\omega) \) which we measure by \( \|\hat{R} - R\|_\infty := \max_{\omega \in [0,1]} |\hat{R}(\omega) - R(\omega)| \). Thanks to the classical perturbation theory on singular subspaces by Wedin [55, 54, 37, Theorem 3.4], \( \|\hat{R} - R\|_\infty \) can be estimated as follows:

**Proposition 5.** Suppose \( M+1 \geq 2S \) and \( L \) is chosen such that \( L \geq S \) and \( M-L+1 \geq S \). Suppose \( 2\|H(\eta)\|_2 < x_{\min}\sigma_{\min}(\Phi_L)\sigma_{\min}(\Phi_{M-L}) \). Then

\[
\|\hat{R} - R\|_\infty \leq \frac{2\|H(\eta)\|_2}{x_{\min}\sigma_{\min}(\Phi_L)\sigma_{\min}(\Phi_{M-L})}.
\]

**Remark 7.** With Wedin’s perturbation bound, Proposition 5 improves Theorem 3 in [40], which shows \( \|\hat{R} - R\|_\infty \leq \text{Constant} \frac{x_{\max}\sigma_{\max}(\Phi_L)\sigma_{\max}(\Phi_{M-L})}{x_{\min}\sigma_{\min}(\Phi_L)\sigma_{\min}(\Phi_{M-L})} \|H(\eta)\|_2 \).

If \( \eta \) is independent gaussian noise, i.e., \( \eta \sim \mathcal{N}(0,\sigma^2 I) \), the spectral norm of \( H(\eta) \) satisfies the following concentration inequality [39, Theorem 4]:

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

\[
\mathbb{E}\|H(\eta)\|_2 \leq \sigma \sqrt{2 \max(L+1,M-L+1) \log(M+2)},
\]

\[
\mathbb{P}\{\|H(\eta)\|_2 \geq t\} \leq (M+2) \exp \left(-\frac{t^2}{2\sigma^2 \max(L+1,M-L+1)}\right), \forall t > 0.
\]
The crucial quantities that determine the stability of MUSIC are precisely $\sigma_{\text{min}}(\Phi_L)$ and $\sigma_{\text{min}}(\Phi_{M-L})$. It is the best to set $L = \lfloor M/2 \rfloor$ to balance them. Theorem 2 gives an accurate estimate for the minimum singular value of $\sigma_{\text{min}}(\Phi_L)$ under the localized-clump model, which constitutes a theoretical foundation to explain the super-resolution limit of MUSIC. Combining Proposition 5, Lemma 2 and Theorem 2 gives rise to the following:

**Theorem 5.** Let $M$ be an even integer satisfying $M \geq 2S^2$, $\alpha > 0$, $\varepsilon > 0$ and $\nu > 1$. Suppose $\Omega = \{\omega_j\}_{j=1}^S \subseteq \mathbb{T}$ consists of $A$ localized clumps $\{\Lambda_a\}_{a=1}^A$, and if $A > 1$, assume that $\Delta(\Omega) \geq \alpha/M$ with $\alpha$ satisfying (2.4) and

$$
\min_{m \neq n} \text{dist}(\Lambda_m, \Lambda_n) \geq \max_{1 \leq a \leq A} \frac{20S^{1/2}\lambda_a^{5/2}}{\alpha^{1/2}M}.
$$

For each $1 \leq a \leq A$, let $c_a = C_a(\Omega, M/2)$. Suppose $\eta \sim \mathcal{N}(0, \sigma^2 I)$ and

$$
\frac{\sigma}{x_{\text{min}}} < \frac{M}{32\sqrt{\nu(M + 2)\log(M + 2)}} \left( \sum_{a=1}^A c_a^2 \alpha^{-2(\lambda_a-1)} \right)^{-1} \varepsilon.
$$

Let $\hat{R}$ and $R$ be the noise-space correlation functions in MUSIC with $L = M/2$ in the noisy and noiseless case, respectively. Then,

$$
\|\hat{R} - R\|_\infty \leq \varepsilon.
$$

with probability no less than $1 - (M + 2)^{-(\nu-1)}$.

Theorem 5 is proved in Appendix A.5. It is stated under the assumption that $M$ is even only for simplicity of the constants, and a similar result holds for odd $M$ by setting $L = \lfloor M/2 \rfloor$. Theorem 5 shows that the noise-to-signal ratio that MUSIC can tolerate to guarantee a fixed level perturbation of the noise-space correlation function obeys

$$
\frac{\sigma}{x_{\text{min}}} \propto \sqrt{\frac{M}{\log M}} \left( \sum_{a=1}^A c_a^2 \alpha^{-2(\lambda_a-1)} \right)^{-1}.
$$

As an example, let us look at the special case where each $\Lambda_a$ contains $\lambda$ equally spaced object with spacing $\alpha/M$ (see Figure 3). In this case, (3.5) becomes

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

which shows that the resolution limit of MUSIC is exponential in $1/\text{SRF}$. The key contribution of this result is that, the exponent only depends on the cardinality of localized clumps instead of the total sparsity $S$. These estimates are verified by numerical experiments below. When noise is i.i.d. Gaussian, increasing $M$ helps to improve the resolution limit of subspace methods.
3.4 Numerical simulations on the super-resolution limit of MUSIC

In our experiments, the true support Ω contains A = 1, 2, 3, 4 clusters of λ equally spaced sources separated by \( \Delta = \alpha/M \) where \( 1/\alpha \) is the SRF (super-resolution factor) of Ω (see Figure 7 (a) for an example). Clumps are separated at least by \( 10/\alpha \) where \( \alpha \geq \text{Constant} / \lambda \). All coefficients have unit magnitudes and random phases. We set \( M = 100 \) and let \( \Delta \) vary so that the SRF varies from 1 to 10. Noise is gaussian: \( \eta \sim \mathcal{N}(0, \sigma^2 I) \). We run MUSIC with varying SRF and \( \sigma \) for 10 trials. The support error is measured by the bottleneck distance between Ω and \( \hat{\Omega} \):

\[
\text{dist}_B(\Omega, \hat{\Omega}) = \inf_{\text{bijection } \psi} \sup_{\hat{\omega} \in \hat{\Omega}} |\hat{\omega} - \psi(\omega)|_T.
\]

Figure 7 (b) displays the average \( \log_2[\text{dist}_B(\Omega, \hat{\Omega})/\Delta] \) over 10 trials with respect to \( \log_{10} \text{SRF} \) (x-axis) and \( \log_{10} \sigma \) (y-axis) when Ω contains 2 clusters 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 \( \sigma \) is below a certain threshold.

Figure 7: Figure 7 (b) displays the average \( \log_2[\text{dist}_B(\Omega, \hat{\Omega})/\Delta] \) over 10 trials with respect to \( \log_{10} \text{SRF} \) (x-axis) and \( \log_{10} \sigma \) (y-axis) when Ω contains 2 clusters of 3 equally spaced point sources.

| Clump | \( A \) | \( \lambda = 2 \) | \( \lambda = 3 \) | \( \lambda = 4 \) | \( \lambda = 5 \) | 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 | 7.2999 | 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 | \( 2\lambda - 2 \) |

Table 2: Numerical simulations of \( q(\Omega) \) in (3.7) on the super-resolution limit of MUSIC.

In Figure 8 we display the phase transition curves at which \( \text{dist}_B(\Omega, \hat{\Omega}) \approx \Delta/2 \) with respect to \( \log_{10} \text{SRF} \) (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.
Figure 8: Phase transition curves: (a-d) shows the average $\log_2[\text{dist}_B(\Omega, \hat{\Omega})/\Delta]$ over 10 trials with respect to $\log_{10}\text{SRF}$ (x-axis) and $\log_{10}\sigma$ (y-axis) when $\Omega$ 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 of each curve by a straight line.

lines, manifesting that the noise level $\sigma$ that MUSIC can tolerate satisfies

$$\sigma \propto \text{SRF}^{-q(\Omega)}.$$  \hfill (3.7)

A least squares fitting of the curves by straight lines gives rise to the exponent $q(\Omega)$ numerically, summarized in Table 2. It is almost consistent with our theory that the exponent $q(\Omega)$ in the phase transition curves depends on the cardinality of the clumps instead of the total number of sources.

### 3.5 Stability of other subspace methods

Our estimates for $\sigma_{\min}(\Phi)$ apply to subspace methods in general. The stabilities of ESPRIT and MPM were studied in [26, 2, 41] when all point sources are well separated. It has been
proved that, roughly speaking, the recovered source locations by ESPRIT and MPM have the following error

\[
\text{Error} \leq \text{Constant} \cdot \left( \frac{x_{\text{max}}}{x_{\text{min}}} \right)^k \cdot \frac{\sigma_{\text{max}}^2(\Phi)}{\sigma_{\text{min}}^2(\Phi)} \cdot \frac{\|H(\eta)\|_2}{x_{\text{min}} \sigma_{\text{min}}^2(\Phi)}, \tag{3.8}
\]

based on \[2\, \text{Theorem 1}\], where \(k = 0\) for ESPRIT and \(k = 1\) for MPM.

Our sharp estimates of \(\sigma_{\text{min}}(\Phi)\) can be used to explain the super-resolution of ESPRIT and MPM as well. Suppose \(\Omega\) consists of \(A\) clumps and each clump \(\Lambda_a\) contains \(\lambda\) equally spaced object with spacing \(\alpha/M\). Consider i.i.d. gaussian noise: \(\eta \sim \mathcal{N}(0, \sigma^2 I)\). In this case, combining (3.8) and Theorem 2 implies that, the noise level that ESPRIT and MPM can tolerate obeys

\[
\frac{\sigma}{x_{\text{min}}} \propto \sqrt{\frac{M}{\log M} \left( \frac{x_{\text{min}}}{x_{\text{max}}} \right)^k \alpha^{4\lambda-4} \left( \frac{1}{\text{SRF}} \right)^{4\lambda-4}.} \tag{3.9}
\]

Notice that the exponent is worse than the one by MUSIC in (3.6). It is well known that ESPRIT improves over MUSIC in terms of accuracy and computational cost. We conjecture that the stability bounds in (3.9) is not optimal, and will consider the improvement in future work.

4 Conclusion

Without any additional assumptions on the unknown measure \(\mu\), super-resolution is an ill-posed inverse problem. Past works showed that the assumption \(\Delta > C/M\), for a sufficiently large universal constant \(C\), regularizes the super-resolution problem. When \(\Delta < 1/M\), a different kind of regularization is needed. As seen in our estimate for the min-max error, sparsity is not strong enough to regularize the problem.

In this paper, we regularized this ill-posed inverse problem by imposing a geometric constraint on \(\Omega\). More formally, if \(\Omega\) consists of well separated clumps and the complexity of each clump is low, then super-resolution is possible by a class of subspace methods provided that the noise level is sufficiently small compared to \(\sigma_{\text{min}}(\Phi)\). By accurately estimating \(\sigma_{\text{min}}(\Phi)\) under the geometric constraint on \(\Omega\), we derived the resolution limit of a subspace method called MUSIC, to explain when MUSIC succeed or fail. This is the first result to rigorously confirm prior numerical evidence that MUSIC can succeed in the regime \(\Delta < 1/M\).

A Proof of theorems

A.1 Proof of Theorem 1

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 \(\text{Fejér kernel} F_P \in C^\infty(\mathbb{T})\) by the formula,

\[
F_P(\omega) = \frac{1}{P+1} \sum_{m=-P}^{P} \left( 1 - \frac{|m|}{P+1} \right) e^{2\pi i m \omega} = \frac{1}{(P+1)^2} \left( \frac{\sin(\pi(P+1)\omega)}{\sin(\pi \omega)} \right)^2.
\]
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

$$
\|F_P\|_{L^2(\mathbb{T})} = \frac{1}{(P+1)^2} \left( (P+1)^2 + 2 \sum_{m=1}^{P} m^2 \right)^{1/2} \leq \frac{1}{(P+1)^{1/2}}.
$$

(A.1)

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

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

(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 1 relies on the quantitative properties of a set of polynomials $\{I_j\}_{j=1}^S$, with $I_j$ depending on $\Omega$ 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 $\omega_j$ and

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

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 3.** Suppose the assumptions of Theorem 1 hold. For each $1 \leq a \leq A$ 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) $|I_j(\omega_k)| \leq 1/(20S)$ for all $\omega_k \notin \Lambda_a$.

(c) $\|I_j\|_{L^2(\mathbb{T})} \leq (2/M)^{1/2}B_a \lambda_a^{-1} \rho_j$.

**Proof of Theorem 1.** Let $\{I_j\}_{j=1}^S$ be the polynomials constructed in Lemma 3. 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 $1 \leq k \leq S$, we define the quantity

$$
\varepsilon_k = I(\omega_k) - v_k.
$$
Since $I_j(\omega_j) = 1$, we have

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

Fix a $\omega_k \in \Omega$. Then $\omega_k \in \Lambda_a$ for some $1 \leq a \leq A$. 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 3, we deduce

$$|\varepsilon_k| \leq \left( \sum_{j \neq k} |I_j(\omega_k)|^2 \right)^{1/2} = \left( \sum_{\omega_j \notin \Lambda_a} |I_j(\omega_k)|^2 \right)^{1/2} \leq \frac{1}{20\sqrt{S}}.$$

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

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

The conditions of robust duality, Proposition 2, are satisfied, so we have

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

To complete the proof, we need to upper bound $\|I(\cdot, v)\|_{L^2(T)}$ uniformly in $v$. We use Cauchy-Schwartz, that $v$ has unit norm, and the norm bound for $I_j$ given in Lemma 3 to obtain the upper bound,

$$\|I(\cdot, v)\|_{L^2(T)} \leq \left( \sum_{j=1}^S \|I_j\|_{L^2(T)}^2 \right)^{1/2} \leq \left( \frac{2}{M} \right)^{1/2} \left( \sum_{j=1}^S (B_a^2 \lambda_a^{-1}\rho_j)^2 \right)^{1/2}.$$

Combining the previous two inequalities completes the proof.

A.2 Proof of Theorem 2

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

$$\rho_j = \prod_{\omega_k \in \Lambda_a \setminus \{\omega_j\}} \frac{1}{\pi M |\omega_k - \omega_j|_T} \leq \left( \frac{1}{\pi \alpha} \right)^{\lambda_a^{-1}}.$$

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 (2.5) implies (2.2). Hence, the assumptions of Theorem 1 are satisfied, and we have

$$\sigma_{\min}(\Phi) \geq \sqrt{M} \left( \sum_{a=1}^A \sum_{\omega_j \in \Lambda_a} (B_a \lambda_a^{-1}\rho_j)^2 \right)^{-1/2}.$$
We can write the right hand side in terms of $\alpha$. Observe that if $\tilde{\Lambda}_a = \{\tilde{\omega}_j\}_{j=1}^{\lambda_a}$ contains $\lambda_a$ points that are equispaced by $\alpha/M$ and $\tilde{\rho}_j$ is the complexity of $\tilde{\omega}_j$, then

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

Thus we have the inequality,

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

Combining the above inequalities completes the proof. \qed

A.3 Proof of Theorem 3

The crux of the proof is to construct, for each subset $\Omega$ of the grid of width $1/N$ and cardinality $S$, a family of polynomials $\{H_j(\cdot, \Omega)\}_{j=1}^S$ with small $L^2(T)$ 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 4. Suppose the assumptions of Theorem 3 hold and let $C(M, S)$ be the constant defined in the theorem. For each subset $\Omega$ of the grid of width $1/N$ and of cardinality $S$, there exist a family of polynomials $\{H_j(\cdot, \Omega)\}_{j=1}^S \subseteq \mathcal{P}(M)$ such that

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

Moreover, we have the upper bound,

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

Proof of Theorem 3 By definition of the lower restricted isometry constant, there exists a set $\Omega$ of cardinality $S$ and supported on the grid of width $1/N$ 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 4. 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 4, we see that \( H \in \mathcal{P}(M, \Omega, u) \). By exact duality, Proposition 1, we have

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

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

\[
\|H\|_{L^2(T)} \leq \left( \sum_{j=1}^S \|H_j\|_{L^2(T)}^2 \right)^{1/2}.
\]

Combining the previous inequalities and using the upper bound given in Lemma 4 completes the proof of the theorem.

\[\square\]

### A.4 Proof of Theorem 4

**Proof.** The upper bound for the min-max error is a direct consequence of Proposition 4 and Theorem 3. To obtain a lower bound for the min-max error, we first apply Proposition 3 to the case that \( \Omega \) 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 3 are satisfied, and thus,

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

Combining this with Proposition 4 establishes a lower bound for the min-max error.

\[\square\]

### A.5 Proof of Theorem 5

**Proof.** By Theorem 2, we have \( \sigma_{\min}(\Phi_{M/2}) \geq \sqrt{M/2} \left( \sum_{a=1}^A c_a^2 \alpha^{-2(\lambda_a - 1)} \right)^{-1/2} \). A sufficient condition for \( \|\hat{R} - R\|_{\infty} \leq \varepsilon \) is

\[
\|\mathcal{H}(\eta)\|_2 \leq M x_{\min} \sigma_{\min}^2(\Phi_{M/2}) \varepsilon / 2
\]

by Proposition 5, which is guaranteed when

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

Lemma 2 implies that \( \text{[A.3]} \) holds with probability no less than \( 1 - (M + 2)e^{-\nu} \) as long as \( t = M x_{\min} \left( \sum_{a=1}^A c_a^2 \alpha^{-2(\lambda_a - 1)} \right)^{-1} \varepsilon / 4 \) and \( \frac{t^2}{2\alpha^2(M + 2)} \geq \nu \log(M + 2) \), which is given by (3.4).

\[\square\]
B Proof of propositions

B.1 Proof of Proposition 1

Proof. The set of all trigonometric polynomials \( f \in \mathcal{P}(M) \) can be written in the form
\[
f(\omega) = \sum_{m=0}^{M-1} \hat{f}(m)e^{2\pi i m \omega}.
\]
Then \( f \in \mathcal{P}(M, \Omega, v) \) if and only if \( f \in \mathcal{P}(M) \) and its Fourier coefficients satisfy the under-determined system of equations,
\[
v_j = \sum_{m=0}^{M-1} \hat{f}(m)e^{2\pi i m \omega_j} \quad \text{for} \quad 1 \leq j \leq S.
\]
By our earlier observation about the Lagrange polynomials, there exists a \( f \in \mathcal{P}(\Omega, M, u) \).
Since \( \|f\|_{L^2(T)} = \|\hat{f}\|_{\ell^2(\mathbb{Z})} \), the functions \( f \in \mathcal{P}(M) \) that satisfy this system of equations and have minimal \( L^2(T) \) norm are the ones with Fourier coefficients given by the Moore-Penrose inverse solution to the above system of equations. Namely,
\[
\min_{f \in \mathcal{P}(M, \Omega, v)} \|f\|_{L^2(T)} = \min_{\Phi \ast u = v} \|u\|_2 = \| (\Phi^\ast)^\dagger v \|_2 = \frac{1}{\sigma_{\min}(\Phi)}.
\]
Rearranging this inequality completes the proof of the proposition.

B.2 Proof of Proposition 2

Proof. Define the measure \( \mu = \sum_{j=1}^{S} v_j \delta_{\omega_j} \), and note that \( \hat{\mu}(m) = (\Phi v)_m \). We have
\[
\left| \int_T \overline{f} d\mu \right| = \left| \sum_{j=1}^{S} f(\omega_j)v_j \right| = \left| \|v\|_2^2 + \sum_{j=1}^{S} v_j \varepsilon_j \right| \geq \|v\|_2^2 - \|v\|_2 \|\varepsilon\|_2 = 1 - \|\varepsilon\|_2.
\]
On the other hand, using that \( f \in \mathcal{P}(M) \), Cauchy-Schwartz, and Parseval,
\[
\left| \int_T \overline{f} d\mu \right| = \left| \sum_{m=1}^{M-1} \overline{\hat{f}(m)} \hat{\mu}(m) \right| \leq \|\hat{f}\|_{\ell^2(\mathbb{Z})} \|\Phi v\|_2 = \|f\|_{L^2(T)} \|\Phi v\|_2.
\]
Combining the previous two inequalities completes the proof.

B.3 Proof of Proposition 3

Proof. The argument relies on the variational form for the minimum singular value,
\[
\sigma_{\min}(\Phi) = \min_{u \in \ell^2(S), u \neq 0} \frac{\|\Phi u\|_2}{\|u\|_2}.
\]
To obtain an upper bound, it suffices to consider a specific \( u \), and our choice is inspired by Donoho [17]. 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
\[
\|u\|_2 = 2^{\lambda - 2}\lambda - 1 \cdot \lambda - 1^{1/2}.
\]
By the variational form for the minimum singular value, we have
\[
\sigma_{\min}(\Phi) \leq \frac{\|\Phi u\|_2}{\|u\|_2} = \left(\frac{2\lambda - 2}{\lambda - 1}\right)^{-1/2} \|\Phi u\|_2.
\]
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 the 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
\[
\|\Phi u\|_2 = \sum_{m=0}^{M} |(\Phi u)_m|^2 = \left(\sum_{m=0}^{M} |\hat{\mu}(m)|^2 \right)^{1/2} = \|\mu * D_M\|_{L^2(\mathbb{T})}.
\]
We see that all \( \omega \in \mathbb{T} \),
\[
(\mu * D_M)(\omega) = \sum_{j=0}^{\lambda-1} (-1)^j \binom{\lambda-1}{j} D_M(\omega - j\alpha/M).
\]
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
\[
(\mu * D_M)(\omega) = \left(\frac{\alpha}{M}\right)^{\lambda-1} D_M^{(\lambda-1)}(\omega) + R_{\lambda-1}(\omega),
\]
where \( D_M^{(\lambda-1)} \) denotes the \((\lambda - 1)\)-th derivative of \( D_M \) and the remainder term \( R_{\lambda-1} \) in magnitude is point-wise \( O((\alpha/M)^{\lambda}) \) as \( \alpha \to 0 \). In order to exactly determine how small we require \( \alpha \) to be, we calculate the remainder term explicitly. By a Taylor expansion of \( D_M \), for each \( \omega \in \mathbb{T} \) and \( 0 \leq j \leq \lambda - 1 \), there exists \( \omega_j \in (\omega - j\alpha/M, \omega) \) such that
\[
D_M(\omega - j\alpha) = \sum_{k=0}^{\lambda-1} D_M^{(k)}(\omega) \left(\frac{\alpha}{M}\right)^{\lambda-1} D_M^{(\lambda)}(\omega) \left(\frac{\alpha}{M}\right)^{\lambda} \frac{(-1)^{\lambda} j^\lambda}{\lambda!}.
\]
Using this formula in equations \((B.2)\) and \((B.3)\), we see that

\[
R_{\lambda-1}(\omega) = \sum_{j=0}^{\lambda-1} (-1)^j + \lambda \binom{\lambda - 1}{j} D_M^{(\lambda)}(\omega_j) \left( \frac{\alpha}{M} \right)^j \left( \frac{\alpha}{M} \right)^{\lambda-j}.
\]

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

\[
\|D^{(\lambda-1)}_M\|_{L^2(T)} \leq (2\pi M)^{\lambda-1} \|D_M\|_{L^2(T)} = \sqrt{M+1} (2\pi M)^{\lambda-1}.
\]

By the same argument, we have

\[
\|R_{\lambda-1}\|_{L^2(T)} \leq \sum_{j=0}^{\lambda-1} \lambda \binom{\lambda - 1}{j} \left( \frac{\alpha}{M} \right)^j \left( \frac{\alpha}{M} \right)^{\lambda-j} \|D_M^{(\lambda)}\|_{L^\infty(T)}
\]

\[
\leq C(\lambda) \alpha (2\pi M)^{\lambda-1} \|D_M\|_{L^\infty(T)}
\]

\[
\leq C(\lambda) \alpha (2\pi M)^{\lambda-1} (M + 1).
\]

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

\[
\|\mu * D_M\|_{L^2(T)} \leq \sqrt{M+1} (2\pi M)^{\lambda-1} \left( 1 + C(\lambda) \alpha \sqrt{M+1} \right).
\]

This inequality and the assumed upper bound for \(\alpha \) \((2.1)\), we see that

\[
\|\mu * D_M\|_{L^2(T)} \leq 2\sqrt{M+1} (2\pi M)^{\lambda-1}.
\]

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

### B.4 Proof of Proposition 4

For this proof, we make the following notational changes. We can identify every discrete measure \(\mu\) whose support is contained in the grid of width \(1/N\) and consists of \(S\) points with a \(S\)-sparse vector \(x \in \mathbb{C}^N\). Under this identification, the Fourier transform of \(\mu\) is identical to the discrete Fourier transform of \(x\). Let \(\mathbb{C}^N_S\) 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

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

**Proof.** We prove the upper bound first. Let \(\varphi\) 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 \(\varphi_y\), just choose any one of them arbitrarily. Note that \(\varphi_y\) exists because \(x\) also satisfies the constraint that \(\|\mathcal{F}x - y\|_2 \leq \delta\), and the choice of \(\varphi_y\) does not explicitly depend on \(x\) and \(\eta\). Note that \(\|\tilde{x}\|_0 \leq \|\varphi_y\|_0 \leq S\) by definition of \(\varphi\). Then we have

\[
\mathcal{E}(M, N, S, \delta) \leq \sup_{(x, \eta) \in \mathcal{Y}} \|\varphi_y - x\|_2.
\]
For any \( x \in \mathbb{C}^N_S \) and \( \eta \) with \( \| \eta \|_2 \leq \delta \), we have \( \varphi_y - x \in \mathbb{C}^N_{2S} \) and
\[
\Theta(M, N, 2S) \leq \frac{\| F(\varphi_y - x) \|_2}{\| \varphi_y - x \|_2} \leq \frac{\| F\varphi_y - y \|_2 + \| Fx - y \|_2}{\| \varphi_y - x \|_2} \leq \frac{2\delta}{\| \varphi_y - x \|_2}.
\]
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 \( v \in \mathbb{C}^N_{2S} \) of unit norm such that
\[
\Theta(M, N, 2S) = \| Fv \|_2.
\]
Pick any vectors \( v_1, v_2 \in \mathbb{C}^N_S \) such that
\[
\frac{\delta}{\Theta(M, N, 2S)} v = v_1 - v_2.
\]
Suppose we are given the data
\[
y = Fv_1 = Fv_2 + F(v_1 - v_2).
\]
Let \( \eta = F(v_1 - v_2) \in \mathbb{C}^{M+1} \). The previous three equations imply
\[
\| \eta \|_2 = \| F(v_1 - v_2) \|_2 = \frac{\delta}{\Theta(M, N, 2S)} \| Fv \|_2 \leq \delta.
\]
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 \( F(v_1 - v_2) \) with noise \( \eta \). Thus, we have \( y \in \mathcal{Y} \) with \( y = y(v_1, 0) \) and \( y = y(v_2, \eta) \). Consequently, we have
\[
E(M, N, S, \delta) \geq \inf_{\varphi \in \mathcal{A}} \max_{k=1,2} \| f(y) - v_k \|.
\]
Using that \( v \) has unit norm, for any \( \varphi \in \mathcal{A} \), we have
\[
\frac{\delta}{\Theta(M, N, 2S)} = \| v_1 - v_2 \|_2 \leq \| \varphi_y - v_1 \|_2 + \| \varphi_y - v_2 \|_2 \leq \max_{k=1,2} \| \varphi_y - v_k \|_2.
\]
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.

\[\square\]

C Proof of lemmas

C.1 Proof of Lemma

Lemma 3

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

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

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

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

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 auxiliary functions $G_j$ and $H_j$. We define the Lagrange-like polynomial,

$$G_j(\omega) = \prod_{\omega_k \in \Lambda_a \setminus \{\omega_j\}} \frac{e^{2\pi i Q_j t} - e^{2\pi i Q_j \omega_j}}{e^{2\pi i Q_j \omega_j} - e^{2\pi i Q_j \omega_k}},$$

where $Q_j = \left\lfloor \frac{M}{\lambda_a} \right\rfloor$.

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 $|\omega_j - \omega_k|_T < 1/M$, imply

$$|Q_j \omega_j - Q_j \omega_k|_T = Q_j |\omega_j - \omega_k|_T.$$  

By construction, the function $G_j$ satisfies the important property that

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

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

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

Recall the trigonometric inequality,

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

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

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

Since $Q_j \leq M/\lambda_a$ and $|\omega_j - \omega_k|_T < 1/M$, we have

$$\|G_j\|_{L^\infty(T)} \leq \left(1 - \frac{\pi^2}{3\lambda_a^2}\right)^{-(\lambda_a - 1)/2} \prod_{\omega_k \in \Lambda_a \setminus \{\omega_j\}} \frac{1}{\pi Q_j |\omega_j - \omega_k|_T} = B_a \lambda_a^{\lambda_a - 1} \rho_j. \quad (C.2)$$
We next define the function $H_j$ by the formula,

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

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,

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

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

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

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 $1 \leq |t|/t \leq 2$ for any $t \geq 1$. Using this inequality and that $\lambda_a \geq 2$, we see that

$$(20B_a)^{1/(2\lambda_a)} \leq 20^{1/(2\lambda_a)} \left(1 - \frac{\pi^2}{3\lambda_a^2}\right)^{-1/4+1/(4\lambda_a)} 2^{(\lambda_a - 1)/(2\lambda_a)} \leq 10.$$  

This inequality and the cluster separation condition ([2.2]) imply

$$|\omega_k - \omega_j|_T \geq \frac{10\lambda_a^2(S\lambda_a^{-1})^{1/(2\lambda_a)}}{M} \geq \frac{\lambda_a^2(20B_a S\lambda_a^{-1} \rho_j)^{1/(2\lambda_a)}}{M} \quad \text{for all} \quad \omega_k \notin \Lambda_a.$$  

Combining this with the previous upper bound on $H_j$ shows that

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

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

$$2P_j \lambda_a + (\lambda_a - 1)Q_j \leq \frac{M}{\lambda_a} + (\lambda_a - 1)\left(\frac{M}{\lambda_a}\right) \leq M.$$
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
\[
\|I_j\|_{L^2(T)} \leq \|G_j\|_{L^\infty(T)}\|H_j\|_{L^2(T)} \leq B_a \lambda_a^{-1} \rho_j \left(\frac{2}{M}\right)^{1/2}.
\]
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
\[
|I_j(\omega_k)| \leq \|G_j\|_{L^\infty(T)}|H_j(\omega_k)| \leq \frac{1}{20S} \text{ for all } \omega_k \not\in \Lambda_a.
\]

\[\square\]

C.2 Proof of Lemma 4

**Proof.** Fix integers $M, N, S$ satisfying the assumptions of Lemma 4. Fix a support set $\Omega$, contained in the grid of width $1/N$ and of cardinality $S$. We do a two-scale analysis. For each $\omega_j \in \Omega$, we define the discrete sets and integers,
\[
\Gamma_j = \Gamma_j(\Omega) = \{\omega_k \in \Omega: |\omega_k - \omega_j|_T < \frac{1}{M}\} \quad \text{and} \quad \gamma_j = |\Gamma_j|,
\]
\[
\mathcal{T}_j = \mathcal{T}_j(\Omega) = \{\omega_k \in \Omega: |\omega_k - \omega_j|_T < \frac{S}{2M}\} \quad \text{and} \quad \tau_j = |\mathcal{T}_j|.
\]

To construct $H_j(\cdot, \Omega)$, we need to define two auxiliary functions, similar to the construction done in Lemma 3. We define the integers
\[
Q_{j,k} = Q_{j,k}(\Omega) = \begin{cases} 
\lfloor M/S \rfloor & \text{if } \omega_k \in \mathcal{T}_j \setminus \{\omega_j\}, \\
\lfloor 1/(2|\omega_j - \omega_k|_T) \rfloor & \text{if } \omega_k \in \Omega \setminus \mathcal{T}_j.
\end{cases}
\]
We readily verify that we have the inequalities $1 \leq Q_{j,k} \leq M/S$ and
\[
|Q_{j,k}\omega_j - Q_{j,k}\omega_k|_T = Q_{j,k}|\omega_j - \omega_k|_T \quad \text{for all } \omega_j, \omega_k \in \Omega. \tag{C.5}
\]
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,
\[
G_j(\omega_k) = \delta_{j,k} \quad \text{for all } \omega_j, \omega_k \in \Omega.
\]
We bound $G_j$ in the sup-norm. We begin with the inequality,
\[
\|G_j\|_{L^\infty(T)} \leq \prod_{\omega_k \in \Omega \setminus \{\omega_j\}} \frac{2}{|1 - e^{2\pi i Q_{j,k}(\omega_j - \omega_k)}|}. \tag{C.6}
\]
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 team 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 $\gamma_j \geq 2$. By a Taylor expansion for cosine, we obtain the inequality,

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

Using this lower bound, the observation that $Q_{j,k} = \lfloor M/S \rfloor \leq M/S \leq M/\gamma_j$ 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

$$\prod_{\omega_k \in \Gamma_j \setminus \{\omega_j\}} \frac{2}{1 - e^{2\pi i Q_{j,k}(|\omega_j - \omega_k|)}} \leq \prod_{\omega_k \in \Gamma_j \setminus \{\omega_j\}} \left(1 - \frac{\pi^2 Q_{j,k}^2 |\omega_j - \omega_k|^2}{3}\right)^{-1/2} \prod_{\omega_k \in \Gamma_j \setminus \{\omega_j\}} \frac{1}{\pi Q_{j,k} |\omega_j - \omega_k|^T} \leq \left(1 - \frac{\pi^2}{3 \gamma_j^2}\right)^{-(\gamma_j - 1)/2} \left[\frac{M}{S}\right]^{-(\gamma_j - 1)} \prod_{\omega_k \in \Gamma_j \setminus \{\omega_j\}} \frac{1}{\pi |\omega_j - \omega_k|^T}.$$

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 $t \geq 2$.

(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

$$|e^{2\pi it} - 1| \geq 4|t|_T, \quad \text{for all } t \in \mathbb{R}. \quad (C.7)$$

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

$$\prod_{\omega_k \in \mathcal{T}_j \setminus \Gamma_j} \frac{2}{1 - e^{2\pi i Q_{j,k}(|\omega_j - \omega_k|)}} \leq \prod_{\omega_k \in \mathcal{T}_j \setminus \Gamma_j} \frac{1}{2Q_{j,k} |\omega_j - \omega_k|^T} \leq \left[\frac{M}{S}\right]^{-\tau_j + \gamma_j} \left(\frac{1}{2}\right)^{-\gamma_j} \prod_{\omega_k \in \mathcal{T}_j \setminus \Gamma_j} \frac{1}{\pi |\omega_j - \omega_k|^T}.$$

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

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

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

$$\|G_j\|_{L^\infty(\mathcal{T})} \leq \left(\frac{12}{12 - \pi^2}\right)^{1/2} \left[\frac{M}{S}\right]^{-\gamma_j + 1} \left(\frac{1}{\pi}\right)^{\gamma_j - 1} 2^{S - 2\gamma_j} \prod_{\omega_k \in \mathcal{T}_j \setminus \{\omega_j\}} \frac{1}{|\omega_j - \omega_k|^T}. \quad (C.8)$$
Let $P = \lfloor M/(2S) \rfloor$ and note that $P \geq 1$ because $M \geq 2S$. Let $F_P$ be the Fejér kernel, and by the $L^2(\mathbb{T})$ bound for the Fejér kernel and the observation that $P + 1 \geq M/(2S)$, we have

$$\|F_P\|_{L^2(\mathbb{T})} \leq \left(\frac{1}{P+1}\right)^{1/2} \leq \left(\frac{2S}{M}\right)^{1/2}. \quad (C.9)$$

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

$$H_j(\omega_k) = \delta_{j,k}$$

for all $\omega_j, \omega_k \in \Omega$.

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

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

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

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

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

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

$$E(\Omega) = \sum_{j=1}^S \left[ \frac{M}{S} \right]^{-2\tau_j+2} \left( \frac{1}{\pi^2} \right)^{\gamma_j-1} 4^{S-2\tau_j} \left( j \right)^{2\tau_j-1} \prod_{\omega_k \in T_j \setminus \{\omega_j\}} \frac{1}{|\omega_j - \omega_k|^2 T}. \quad (C.10)$$

To complete the proof of the lemma, we need to obtain the appropriate bound on $E(\Omega)$ uniformly in $\Omega$. This is handled separately in Lemma 5, which is stated below and proved in Appendix C.3.

**Lemma 5.** Suppose the assumptions of Lemma 4 hold and let $E(\Omega)$ be the quantity defined in (C.10). Then

$$E(\Omega) \leq \left[ \frac{M}{S} \right]^{-2S+2} N^{2S-2} \left( \frac{1}{\pi} \right)^{2S-2} \sum_{j=1}^S \prod_{k \neq j} \frac{1}{(j-k)^2}.$$ 

**C.3 Proof of Lemma 5**

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 $\Omega$ supported on the grid of width $1/N$ and of cardinality $S$. To upper bound $E(\Omega)$ uniformly in $\Omega$, one method is to determine which $\Omega$ attain(s) the maximum. The maximizer is clearly not unique, since $E(\Omega)$ is invariant.
under cyclic shifts of $\Omega$ by $1/N$. However, we shall argue that the maximizer is attained by $\Omega^*$, which denotes any support set consisting of $S$ consecutive points separated by $1/N$.

Note that
\[
E(\Omega^*) = \left[ \frac{M}{S} \right]^{-2S+2} N^{2S-2} \left( \frac{1}{\pi} \right)^{2S-2} \sum_{j=1}^{S} \prod_{k \neq j} \frac{1}{(j-k)^2}.
\] (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 $\Omega$, it is not straightforward to prove. When $\Omega$ is contained in a small interval, the product over $T_j$ 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 $\Omega$. 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 difficulty, we proceed with the following extension argument.

**Proof.** We extend $E$ to a function of $D = (S-1)^2$ 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,

\[
A_j(w) = \left\{ w_{j,k} : w_{j,k} < \frac{S}{2M} \right\} \quad \text{and} \quad a_j(w) = |A_j(w)|,
\]

\[
B_j(w) = \left\{ w_{j,k} : w_{j,k} < \frac{1}{M} \right\} \quad \text{and} \quad b_j(w) = |B_j(w)|.
\]

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

\[
F(w) = \sum_{j=1}^{S} \left[ \frac{M}{S} \right]^{-2S+2} \left( \frac{1}{\pi} \right)^{2S-2} \sum_{j=1}^{S} \prod_{k \neq j} \frac{1}{(j-k)^2} \cdot \prod_{w_{j,k} \in A_j(w)} \frac{1}{w_{j,k}^2}.
\] (C.12)

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

\[
H = \bigcap_{k=1}^{S} \left\{ w \in \mathbb{R}^D : \sum_{j \neq k} w_{j,k} \geq \frac{c(S)}{N} \right\},
\]

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

\[
c(S) = \begin{cases} 
2 \left( 1 + 2 + \cdots + \frac{S-1}{2} \right) & \text{if } S \text{ is odd}, \\
2 \left( 1 + 2 + \cdots + \frac{S-2}{2} \right) + \frac{S}{2} & \text{if } S \text{ is even}.
\end{cases}
\]

We argue that $F$ is an extension of $E$. Note that any $\Omega$ can be mapped to a $w(\Omega) \in \mathbb{R}^D$ via the relationship $(w(\Omega))_{j,k} = |\omega_j - \omega_k|_T$ 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 \( 1 \leq k \leq S \), we have
\[
\sum_{j \neq k} (w(\Omega))_{j,k} = \sum_{j \neq k} |\omega_j - \omega_k| \geq \frac{c(S)}{N}.
\]
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 \( \Omega \), we have
\[
E(\Omega) = F(w(\Omega)) \leq \sup_{w \in [1/N, 1/2]^D \cap H} F(w). \tag{C.13}
\]

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
\[
\max_{w \in [1/N, 1/2]^D \cap H} F(w) = \max_{w \in [1/N, S/(2M)]^D \cap H} F(w), \tag{C.14}
\]
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, for any maximizer \( w \) of \( F \), there exist indices \((m, n)\) such that \( a_m(w) \leq S-1 \) 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
\[
F(w) - F(v) = \left( \frac{M}{S} \right)^{-2a_m(w)+2} b_m(w)^{-1} 4^{-S-2a_m(w)+b_m(w)} \left( \prod_{w_{j,k} \in A_j(w)} \frac{1}{w_{m,k}^2} \right) \left( 1 - \frac{M}{S} \right)^{-2} \frac{1}{4\pi^2 v_{m,n}^2}.
\]
The assumption that \( N \geq \pi MS \) and \( S \geq 2 \) imply
\[
\frac{1}{4\pi^2 v_{m,n}^2} = \frac{N^2}{4\pi^2 S^2} \geq \left( \frac{M}{2} \right)^2 \geq \left( \frac{M}{S} \right)^2.
\]
This proves that \( F(w) \leq F(v) \), 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, 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
\[
F(w) - F(v) = |M|^{-2S+2} \left( \frac{1}{\pi^2} \right)^{b_{m}(w)-1} 4^{-S+b_{m}(w)} \left( \prod_{k \in A_{m}(w) \setminus \{n\}} \frac{1}{w_{m,k}^2} \right) \left( \frac{1}{w_{m,n}^2} - \frac{4}{\pi^2 v_{m,n}^2} \right).
\]
The assumption that \( N \geq \pi MS \) implies
\[
\frac{4}{\pi^2 v_{m,n}^2} = 4N^2 \geq 4M^2 \geq \frac{1}{w_{m,n}^2}.
\]
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,
\[
E(\Omega) = F(w(\Omega)) \leq \max_{w \in [1/N, 1/M]^D \cap H} F(w).
\]
When \( w \in [1/N, 1/M]^D \cap H \), the function \( F \) reduces to
\[
F(w) = |M|^{-2S+2} \left( \frac{1}{\pi^2} \right)^{S-1} \sum_{j=1}^{S} \prod_{k \in B_{j}(w)} \frac{1}{w_{j,k}^2}.
\]
Since \( F \) is a smooth function of \( w \), a straightforward calculation shows that each partial derivative of \( F \), with respect to the canonical basis on \( \mathbb{R}^d \), 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
\[
H_k = \left\{ w \in \mathbb{R}^D : \sum_{j \neq k} w_{j,k} = \frac{c(S)}{N} \right\}.
\]
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 \( \Omega \) consists of \( S \) consecutive indices. This proves that for all \( \Omega \), 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. \[\square\]
5 Acknowledgements

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.

References

[1] Noga Alon and Yuval Peres. Uniform dilations. Geometric and Functional Analysis, 2(1):1–28, 1992.

[2] Céline Aubel and Helmut Bölskei. Deterministic performance analysis of subspace methods for cisoid parameter estimation. In Information Theory (ISIT), 2016 IEEE International Symposium on, pages 1551–1555. IEEE, 2016.

[3] Céline Aubel and Helmut Bölskei. Vandermonde matrices with nodes in the unit disk and the large sieve. Applied and Computational Harmonic Analysis, 2017.

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

[5] Dmitry Batenkov, Laurent Demanet, Gil Goldman, and Yosef Yomdin. Stability of partial Fourier matrices with clustered nodes. arXiv preprint arXiv:1809.00658, 2018.

[6] Fermín Bazán. Conditioning of rectangular Vandermonde matrices with nodes in the unit disk. SIAM Journal on Matrix Analysis and Applications, 21(2):679–693, 2000.

[7] Bernhard Beckermann. The condition number of real Vandermonde, Krylov and positive definite Hankel matrices. Numerische Mathematik, 85(4):553–577, 2000.

[8] John J. Benedetto and Weilin Li. Super-resolution by means of Beurling minimal extrapolation. Applied and Computational Harmonic Analysis, 2018.

[9] Lihu Berman and Arie Feuer. On perfect conditioning of Vandermonde matrices on the unit circle. Electronic Journal of Linear Algebra, 16(1):13, 2007.

[10] Emmanuel J. Candès. The restricted isometry property and its implications for compressed sensing. Comptes Rendus Mathematique, 346(9-10):589–592, 2008.

[11] Emmanuel J. Candès and Carlos Fernandez-Granda. Super-resolution from noisy data. Journal of Fourier Analysis and Applications, 19(6):1229–1254, 2013.
[12] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[13] Yuejie Chi, Louis L. Scharf, Pezeshki Ali, and A. Robert Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Transactions on Signal Processing*, 59(5):2182–2195, 2011.

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

[15] Arnold Jan Den Dekker and A. Van den Bos. Resolution: A survey. *Journal of Optical Society of America*, 14(3):547–557, 1997.

[16] Quentin Denoyelle, Vincent Duval, and Gabriel Peyré. Support recovery for sparse super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, 23(5):1153–1194, 2017.

[17] David L. Donoho. Superresolution via sparsity constraints. *SIAM Journal on Mathematical Analysis*, 23(5):1309–1331, 1992.

[18] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[19] Marco F. Duarte and Richard G. Baraniuk. Spectral compressive sensing. *Applied and Computational Harmonic Analysis*, 35(1):111–129, 2013.

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

[21] Alfredo Eisinberg, Giuseppe Franzé, and Nicola Salerno. Rectangular Vandermonde matrices on Chebyshev nodes. *Linear Algebra and its Applications*, 338(1-3):27–36, 2001.

[22] Alfredo Eisinberg, Paolo Pugliese, and Nicola Salerno. Vandermonde matrices on integer nodes: the rectangular case. *Numerische Mathematik*, 87(4):663–674, 2001.

[23] Albert C. Fannjiang. Compressive inverse scattering: I. high-frequency simo/miso and mimo measurements. *Inverse Problems*, 26(3):035008, 2010.

[24] Albert C. Fannjiang. The MUSIC algorithm for sparse objects: a compressed sensing analysis. *Inverse Problems*, 27(3):035013, 2011.

[25] Albert C. Fannjiang. Compressive sensing theory for optical systems described by a continuous model. *arXiv preprint arXiv:1507.00794*, 2015.

[26] Albert C. Fannjiang. Compressive spectral estimation with single-snapshot ESPRIT: Stability and resolution. *arXiv preprint arXiv:1607.01827*, 2016.

[27] Albert C. Fannjiang and Wenjing Liao. Mismatch and resolution in compressive imaging. *Wavelets and Sparsity XIV, Proceedings of SPIE*, 2011.
[28] Albert C. Fannjiang and Wenjing Liao. Coherence-pattern guided compressive sensing with unresolved grids. *SIAM Journal on Imaging Sciences*, 5(1):179–202, 2012.

[29] Albert C. Fannjiang, Thomas Strohmer, and Pengchong Yan. Compressed remote sensing of sparse objects. *SIAM Journal on Imaging Sciences*, 3(3):595–618, 2010.

[30] Carlos Fernandez-Granda. Support detection in super-resolution. In *Proceedings of the 10th International Conference on Sampling Theory and Applications*, pages 145–148, 2013.

[31] Paulo Ferreira, J.S.G. 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.

[32] Walter Gautschi. On inverses of Vandermonde and confluent Vandermonde matrices. *Numerische Mathematik*, 4(1):117–123, 1962.

[33] Walter Gautschi and Gabriele Inglese. Lower bounds for the condition number of Vandermonde matrices. *Numerische Mathematik*, 52(3):241–250, 1987.

[34] Yingbo Hua and Tapan K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(5):814–824, 1990.

[35] Sergei V. Konyagin, Imre Z. Ruzsa, and Wilhelm Schlag. On uniformly distributed dilates of finite integer sequences. *Journal of Number Theory*, 82(2):165–187, 2000.

[36] Hamid Krim and Viberg Mats. Two decades of array signal processing research: the parametric approach. *IEEE Signal Processing Magazine*, 13(4):67–94, 1996.

[37] Ren-Cang Li. Relative perturbation theory: II. Eigenspace and singular subspace variations. *SIAM Journal on Matrix Analysis and Applications*, 20(2):471–492, 1998.

[38] Weilin Li. Elementary $L^\infty$ error estimates for super-resolution de-noising. *arXiv preprint arXiv:1702.03021*, 2017.

[39] Wenjing Liao. Music for multidimensional spectral estimation: stability and super-resolution. *IEEE Transactions on Signal Processing*, 63(23):6395–6406, 2015.

[40] Wenjing Liao and Albert Fannjiang. MUSIC for single-snapshot spectral estimation: Stability and super-resolution. *Applied and Computational Harmonic Analysis*, 40(1):33–67, 2016.

[41] Ankur Moitra. Super-resolution, extremal functions and the condition number of Vandermonde matrices. *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, 2015.

[42] Hugh L. Montgomery and Robert C. Vaughan. Hilbert’s inequality. *Journal of the London Mathematical Society*, 2(1):73–82, 1974.
[43] Veniamin I. Morgenshtern and Emmanuel J. Candes. Super-resolution of positive sources: The discrete setup. *SIAM Journal on Imaging Sciences*, 9(1):412–444, 2016.

[44] Par R. Prony. Essai expérimental et analytique. *Journal de L’École Polytechnique*, 1(22):24–76, 1795.

[45] Richard Roy and Thomas Kailath. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(7):984–995, 1989.

[46] Mark Rudelson and Roman Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008.

[47] Øyvind Ryan and Mérouane Debbah. Asymptotic behavior of random Vandermonde matrices with entries on the unit circle. *IEEE Transactions on Information Theory*, 55(7):3115–3147, 2009.

[48] Ralph Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas andPropagation*, 34(3):276–280, 1986.

[49] Petre Stoica and Randolph L. Moses. *Introduction to Spectral Analysis*, volume 1. Prentice Hall, 1997.

[50] Gabriel H. Tucci and Philip A. Whiting. Eigenvalue results for large scale random Vandermonde matrices with unit complex entries. *IEEE Transactions on Information Theory*, 57(6):3938–3954, 2011.

[51] Gabriel H. Tucci and Philip A. Whiting. Asymptotic behavior of the maximum and minimum singular value of random Vandermonde matrices. *Journal of Theoretical Probability*, 27(3):826–862, 2014.

[52] Paul Turán. On rational polynomials. *Acta Universitatis Szegediensis, Sect. Sci. Math*, pages 106–113, 1946.

[53] Jeffrey D. Vaaler. Some extremal functions in Fourier analysis. *Bulletin of the American Mathematical Society*, 12(2):183–216, 1985.

[54] Gilbert W. and Ji-Guang Sun. *Matrix Perturbation Theory*. Academic Press Boston, 1990.

[55] Per-Åke Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111, 1972.