MUSIC for Single-Snapshot Spectral Estimation: Stability and Super-resolution

Wenjing Liao * Albert Fannjiang †

November 9, 2014

Abstract

This paper studies the problem of line spectral estimation in the continuum of a bounded interval with *one snapshot* of array measurement. The single-snapshot measurement data is turned into a Hankel data matrix which admits the Vandermonde decomposition and is suitable for the MUSIC algorithm. The MUSIC algorithm amounts to finding the null space (the noise space) of the Hankel matrix, forming the noise-space correlation function and identifying the s smallest local minima of the noise-space correlation as the frequency set.

In the noise-free case *exact* reconstruction is guaranteed for any arbitrary set of frequencies as long as the number of measurement data is at least twice the number of distinct frequencies to be recovered. In the presence of noise the stability analysis shows that the perturbation of the noise-space correlation is proportional to the spectral norm of the noise matrix as long as the latter is smaller than the smallest (nonzero) singular value of the *noiseless* Hankel data matrix. Under the assumption that the true frequencies are separated by at least twice the Rayleigh Length (RL), the stability of the noise-space correlation is proved by means of novel discrete Ingham inequalities which provide bounds on the largest and smallest nonzero singular values of the noiseless Hankel data matrix.

The numerical performance of MUSIC is tested in comparison with other algorithms such as BLO-OMP and SDP (TV-min). While BLO-OMP is the stablest algorithm for frequencies separated above 4 RL, MUSIC becomes the best performing one for frequencies separated between 2 RL and 3 RL. Also, MUSIC is more efficient than other methods. MUSIC truly shines when the frequency separation drops to 1 RL or below when all other methods fail. Indeed, the resolution length of MUSIC decreases to zero as noise decreases to zero as a power law with an exponent much smaller than an upper bound established by Donoho.

Keywords: MUSIC algorithm, single-snapshot spectral estimation, stability, super-resolution, discrete Ingham inequalities.

1 Introduction

The field of Compressive Sensing (CS) [16] has provided us with a new technology of reconstructing a signal from a small number of linear measurements. With a few exceptions, signals considered in

*Statistical and Applied Mathematical Sciences Institute (SAMSI) and Department of Mathematics, Duke University, Durham, NC. Wenjing Liao is grateful to support from NSF DMS 0847388 and SAMSI under grant NSF DMS-1127914. Email: wjliao@math.duke.edu.

†Department of Mathematics, University of California, Davis, CA. Email: fannjiang@math.ucdavis.edu.

the compressive sensing community are assumed to be sparse under a discrete, finite-dimensional dictionary.

However, signals arising in applications such as radar [9], sonar and remote sensing [19] are represented by few parameters on a continuous domain. These signals are usually not sparse under any discrete dictionary but can be approximately sparsely represented by indicator functions on a discrete domain. An approximation error, called gridding error [21, 17] or basis mismatch [22, 26, 10] exists, manifesting the gap between the continuous world and the discrete world. This issue is well illustrated by the spectral estimation problem [46] as follows.

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

$$\{e^{-2\pi i\omega_j t} : \omega_j \in \mathbb{R},\ j = 1, \dots, s\}.$$

Consider the noisy signal model

$$y^\varepsilon(t) = y(t) + \varepsilon(t), \quad y(t) = \sum_{j=1}^{s} x_j e^{-2\pi i\omega_j t} \tag{1}$$

where $\varepsilon(t)$ is the external noise.

The task of spectral estimation is to find out the frequency support set $\mathcal{S} = \{\omega_1, ..., \omega_s\}$ and the corresponding amplitudes $x = [x_1, ..., x_s]^T$ from a finite data sampled at, say, $t = 0, 1, 2, \cdots, M \in \mathbb{N}$. Because the signal $y(t)$ depends nonlinearly on $\mathcal{S}$, the main difficulty of spectral estimation lies in identifying $\mathcal{S}$. The amplitudes x can be recovered by solving least squares once $\mathcal{S}$ is found.

More explicitly, denote (with a slight abuse of notation) $y = [y_k]_{k=0}^M$, $\varepsilon = [\varepsilon_k]_{k=0}^M$ and $y^\varepsilon = y + \varepsilon \in \mathbb{C}^{M+1}$, with $y_k = y(k)$, $y_k^\varepsilon = y^\varepsilon(k)$ and $\varepsilon_k = \varepsilon(k)$. Let

$$\phi^M(\omega) = [1\ e^{-2\pi i\omega}\ e^{-2\pi i2\omega}\ \dots\ e^{-2\pi iM\omega}]^T \in \mathbb{C}^{M+1} \tag{2}$$

be the imaging vector of size $M + 1$ at the frequency ω and define

$$\Phi^M = [\phi^M(\omega_1)\ \phi^M(\omega_2)\ \dots\ \phi^M(\omega_s)] \in \mathbb{C}^{(M+1)\times s}.$$

The single-snapshot formulation of spectral estimation takes the form

$$y^\varepsilon = \Phi^M x + \varepsilon. \tag{3}$$

Again the main difficulty is in the (nonlinear) dependence of Φ^M on the unknown frequencies in $\mathcal{S}$. In addition, with the sampling times $t = 0, 1, 2, \cdots, M \in \mathbb{N}$, one can only hope to determine frequencies on the torus $\mathbb{T} = [0, 1)$ with the natural metric

$$d(\omega_j, \omega_l) = \min_{n\in\mathbb{Z}} |\omega_j + n - \omega_l|.$$

One can attempt to linearize (3) by expanding the matrix Φ^M via setting up a grid

$$\mathcal{G} = \left\{ \frac{0}{N}, \frac{1}{N}, \dots, \frac{N-1}{N} \right\} \subset [0, 1), \tag{4}$$

where N is some large integer, and writing the spectral estimation problem in the form a linear inversion problem

$$y^\varepsilon = Ax + \varepsilon \tag{5}$$

where

$$A := \left[\phi^M\left(\frac{0}{N}\right) \quad \phi^M\left(\frac{1}{N}\right) \quad \dots \quad \phi^M\left(\frac{N-1}{N}\right)\right] \in \mathbb{C}^{(M+1)\times N}$$

Discretizing $[0, 1)$ as in (4) amounts to rounding frequencies on the continuum to the nearest grid points in $\mathcal{G}$, giving rise to a gridding error which is roughly proportional to the grid spacing. On the other hand, as N increases, correlation among adjacent columns of A also increases dramatically [21].

A key unit of frequency separation is the Rayleigh Length, roughly the minimum resolvable separation of two objects with equal intensities in classical resolution theory [12, 15]. Mathematically, the Rayleigh Length (RL) is the distance between the center and the first zero of the Dirichlet kernel

$$D(\omega) = \int_{-M/2}^{M/2} e^{2\pi i t\omega} dt = \frac{\sin \pi\omega M}{\pi\omega}.$$

Hence 1 RL $= 1/M$.

The ratio $F = N/M$ between RL and the grid spacing is called the refinement factor in [21] and super-resolution factor in [6]. The higher F is, the more coherent the measurement matrix A becomes.

1.1 Single-snapshot MUSIC

In this paper, to circumvent the gridding problem, we reformulate the spectral estimation problem (3) in the form of *multiple measurement vectors* that is suitable for the application of the MUltiple Signal Classification (MUSIC) algorithm [43, 42], widely used in signal processing[48, 30, 34] and array imaging [8, 13, 31].

Most state-of-the-art spectral estimation methods ([46] and references therein) assume many snapshots of array measurement as well as statistical assumptions on measurement noise. In contrast, we pursue below a deterministic approach to spectral estimation with a single snapshot of array measurement in common with [11].

Fixing a positive integer $1 \le L < M$, we form the Hankel matrix

$$H = \text{Hankel}(y) = \begin{bmatrix} y_0 & y_1 & \cdots & y_{M-L} \\ y_1 & y_2 & \cdots & y_{M-L+1} \\ \vdots & \vdots & \vdots & \vdots \\ y_L & y_{L+1} & \cdots & y_M \end{bmatrix}. \tag{6}$$

Since its first appearance in Prony's method [40] the Hankel data matrix (6) plays an important role in modern methods such as the state space method [35, 41] and the matrix pencil method [27].

It is straightforward to verify that Hankel(y) with $y = \Phi^M x$ admits the Vandermonde decomposition

$$H = \Phi^L X (\Phi^{M-L})^T, \quad X = \text{diag}(x_1, \dots, x_s) \tag{7}$$

with the Vandermonde matrix

$$\Phi^L = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{-2\pi i\omega_1} & e^{-2\pi i\omega_2} & \dots & e^{-2\pi i\omega_s} \\ (e^{-2\pi i\omega_1})^2 & (e^{-2\pi i\omega_2})^2 & \dots & (e^{-2\pi i\omega_s})^2 \\ \vdots & \vdots & \vdots & \vdots \\ (e^{-2\pi i\omega_1})^L & (e^{-2\pi i\omega_2})^L & \dots & (e^{-2\pi i\omega_s})^L \end{bmatrix}.$$

Here we use a special property of Fourier measurements: a time translation corresponds to a frequency phase modulation.

Let $H^\varepsilon = \text{Hankel}(y^\varepsilon)$ and $E = \text{Hankel}(\varepsilon)$. The multiple measurement vector formulation of spectral estimation takes the form

$$H^\varepsilon = H + E = \Phi^L X (\Phi^{M-L})^T + E. \tag{8}$$

The crux of MUSIC is this: In the noiseless case with $L \geq s$ and $M - L + 1 \geq s$ the ranges of H and Φ^L coincide and are a proper subspace (the signal space) of $\mathbb{C}^{L+1}$. Let the noise space be the orthogonal complement of the signal space in $\mathbb{C}^{L+1}$. Then $\mathcal{S}$ can be identified as the zero set of the orthogonal projection of the imaging vector $\phi^L(\omega)$ of size $L+1$ onto the noise space.

More specifically, let the Singular Value Decomposition (SVD) of H be written as

$$H = [\ \underbrace{U_1}_{(L+1)\times s}\ \ \underbrace{U_2}_{(L+1)\times(L+1-s)}\ ]\ \underbrace{\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_s, 0, \dots, 0)}_{(L+1)\times(M-L+1)}\ [\ \underbrace{V_1}_{(M-L+1)\times s}\ \ \underbrace{V_2}_{(M-L+1)\times(M-L+1-s)}\ ]^\star$$

with the singular values $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \sigma_s > 0$. The signal and noise spaces are exactly the column spaces of U_1 and U_2 respectively.

The orthogonal projection $\mathcal{P}_2$ onto the noise space is given by $\mathcal{P}_2 \mathrm{w} = U_2(U_2^\star \mathrm{w})$, $\forall \mathrm{w} \in \mathbb{C}^{L+1}$. Under mild assumptions one can prove that $\omega \in \mathcal{S}$ if and only if $\mathcal{P}_2 \phi^L(\omega) = \mathbf{0}$. Hence $\mathcal{S}$ can be identified as the zeros of the noise-space correlation function

$$R(\omega) = \frac{\|\mathcal{P}_2 \phi^L(\omega)\|_2}{\|\phi^L(\omega)\|_2} = \frac{\|U_2^\star \phi^L(\omega)\|_2}{\|\phi^L(\omega)\|_2},$$

or the peaks of the imaging function

$$J(\omega) = \frac{\|\phi^L(\omega)\|_2}{\|\mathcal{P}_2 \phi^L(\omega)\|_2} = \frac{\|\phi^L(\omega)\|_2}{\|U_2^\star \phi^L(\omega)\|_2}.$$

The following fact is the basis for noiseless MUSIC (See Appendix A for proof).

Theorem 1. *Suppose $\omega_k \neq \omega_l$ $\forall k \neq l$. If*

$$L \geq s, \quad M - L + 1 \geq s, \tag{9}$$

then

$$\omega \in \mathcal{S} \Longleftrightarrow R(\omega) = 0 \Longleftrightarrow J(\omega) = \infty.$$

Remark 1. *Condition (9) says that the number of measurement data $(M+1) \geq 2s$ suffices to guarantee exact reconstruction by the MUSIC algorithm.*

For the noisy data matrix H^ε let the SVD be written as

$$H^\varepsilon = [\underbrace{U_1^\varepsilon}_{(L+1)\times s} \quad \underbrace{U_2^\varepsilon}_{(L+1)\times(L+1-s)}] \underbrace{\mathrm{diag}(\sigma_1^\varepsilon, \sigma_2^\varepsilon, \ldots, \sigma_s^\varepsilon, \sigma_{s+1}^\varepsilon, \ldots)}_{(L+1)\times(M-L+1)} [\underbrace{V_1^\varepsilon}_{(M-L+1)\times s} \quad \underbrace{V_2^\varepsilon}_{(M-L+1)\times(M-L+1-s)}]^\star$$

with the singular values $\sigma_1^\varepsilon \geq \sigma_2^\varepsilon \geq \sigma_3^\varepsilon \geq \cdots$. The noise-space correlation function and imaging function become

$$R^\varepsilon(\omega) = \frac{\|\mathcal{P}_2^\varepsilon \phi^L(\omega)\|_2}{\|\phi^L(\omega)\|_2} = \frac{\|U_2^{\varepsilon\star}\phi^L(\omega)\|_2}{\|\phi^L(\omega)\|_2}$$

and

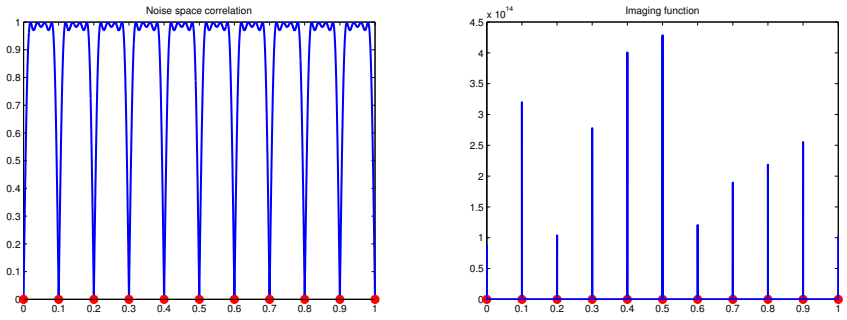
$$J^\varepsilon(\omega) = \frac{1}{R^\varepsilon(\omega)} = \frac{\|\phi^L(\omega)\|_2}{\|\mathcal{P}_2^\varepsilon \phi^L(\omega)\|_2} = \frac{\|\phi^L(\omega)\|_2}{\|U_2^{\varepsilon\star}\phi^L(\omega)\|_2},$$

respectively with $\mathcal{P}_2^\varepsilon = U_2^\varepsilon (U_2^\varepsilon)^\star$.

The MUSIC algorithm is given by

MUSIC for Spectral Estimation
Input: $y^\varepsilon \in \mathbb{C}^{M+1}, s, L$.
1) Form matrix $H^\varepsilon = \mathrm{Hankel}(y^\varepsilon) \in \mathbb{C}^{(L+1)\times(M-L+1)}$.
2) SVD: $H^\varepsilon = [U_1^\varepsilon \; U_2^\varepsilon]\mathrm{diag}(\sigma_1^\varepsilon, \ldots, \sigma_s^\varepsilon, \ldots)[V_1^\varepsilon \; V_2^\varepsilon]^\star$, where $U_1^\varepsilon \in \mathbb{C}^{(L+1)\times s}$.
3) Compute imaging function $J^\varepsilon(\omega) = \|\phi^L(\omega)\|_2 / \|U_2^{\varepsilon\star}\phi^L(\omega)\|_2$.
Output: $\hat{\mathcal{S}} = \{\omega$ corresponding to s largest local maxima of $J^\varepsilon(\omega)\}$.

Figure 1 shows a noise-space correlation function and an imaging function in the noise-free case. True frequencies are exactly located where the noise-space correlation function vanishes and the imaging function peaks.



(a) Noise-space correlation $R(\omega)$ (b) Imaging function $J(\omega)$

Figure 1: Plots of $R(\omega)$ and $J(\omega)$ when $M = 100, L = 50$ and there are 10 equally spaced objects on $\mathbb{T}$ represented by red dots.

The MUSIC algorithm as formulated above requires the number of frequencies s as an input. There are some techniques [14, 33] for evaluating how many objects are present in the event that such information is not available. When $\sigma_s \gg 2\|E\|_2$, s can be easily estimated based on the singular value distribution of H^ε due to Weyl's theorem [49].

Proposition 1 (Weyl's Theorem)**.** $|\sigma_j^\varepsilon - \sigma_j| \le \|E\|_2,\ j = 1, 2, \ldots$

As a result, $\sigma_j^\varepsilon \le \|E\|_2,\ \forall j \ge s+1$ and $\sigma_s^\varepsilon \ge \sigma_s - \|E\|_2$. Hence $\sigma_s^\varepsilon \gg \sigma_{s+1}^\varepsilon$, creating a gap between σ_s^ε and $\{\sigma_j^\varepsilon : j \ge s+1\}$. An example is shown in Figure 3.

Before describing our main results, we pause to define notations to be used in the subsequent sections. For an $m \times n$ matrix A, let $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$ denote the maximum and minimum nonzero singular values of A, respectively. Denote the spectral norm, Frobenius norm and nuclear norm of A by $\|A\|_2$, $\|A\|_F$ and $\|A\|_\star$. Let $x_{\max} = \max_{j=1,\ldots,s} |x_j|$ and $x_{\min} = \min_{j=1,\ldots,s} |x_j|$. The dynamic range of x is defined as $x_{\max}/x_{\min}$. Fixing $\mathcal{S} = \{\omega_1, \ldots, \omega_s\} \subset \mathbb{T}$, we define the matrix $\Phi^{N_1 \to N_2}$ such that

$$\Phi_{kj}^{N_1 \to N_2} = e^{-2\pi i k \omega_j},\ k = N_1, \ldots, N_2,\ j = 1, \ldots, s.$$

For simplicity, we denote $\Phi^M = \Phi^{0 \to M}$.

1.2 Contribution of the present work

The main contribution of the paper is a stability analysis for the MUSIC algorithm with respect to general support set $\mathcal{S}$ and external noise.

In the MUSIC algorithm frequency candidates are identified at the s smallest local minima of the noise-space correlation which measures how much an imaging vector is correlated with the noise space. In noise-free case, the noise-space correlation function $R(\omega)$ vanishes exactly on $\mathcal{S}$. For the noisy case we prove

$$|R^\varepsilon(\omega) - R(\omega)| \le \alpha \|E\|_2, \quad \alpha = \frac{4\sigma_1 + 2\|E\|_2}{(\sigma_s - \|E\|_2)^2} \tag{10}$$

which holds for any support set $\mathcal{S} \subset \mathbb{T}$.

To make the bounds (10) explicit and more meaningful, we prove the discrete Ingham inequalities (Corollary 1) which implies

$$\frac{\sigma_s^2}{L(M-L)} \ge x_{\min}^2 \left(\frac{2}{\pi} - \frac{2}{\pi L^2 q^2} - \frac{4}{L}\right)\left(\frac{2}{\pi} - \frac{2}{\pi (M-L)^2 q^2} - \frac{4}{M-L}\right) \tag{11}$$

$$\frac{\sigma_1^2}{L(M-L)} \le x_{\max}^2 \left(\frac{4\sqrt{2}}{\pi} + \frac{\sqrt{2}}{\pi L^2 q^2} + \frac{3\sqrt{2}}{L}\right)\left(\frac{4\sqrt{2}}{\pi} + \frac{\sqrt{2}}{\pi (M-L)^2 q^2} + \frac{3\sqrt{2}}{M-L}\right) \tag{12}$$

under the gap assumption

$$q = \min_{j \ne l} d(\omega_j, \omega_l) > \max\left(\frac{1}{L}\sqrt{\frac{2}{\pi}}\left(\frac{2}{\pi} - \frac{4}{L}\right)^{-\frac{1}{2}}, \frac{1}{M-L}\sqrt{\frac{2}{\pi}}\left(\frac{2}{\pi} - \frac{4}{M-L}\right)^{-\frac{1}{2}}\right). \tag{13}$$

Furthermore, we prove that for every $\omega_j \in \mathcal{S}$, there exists a local minimizer $\hat{\omega}_j$ of R^ε such that $\hat{\omega}_j \to \omega_j$ as noise decreases to 0.

To relax the restriction on the minimum separation between adjacent frequencies, condition (13) suggests that L should be about $M/2$ and then the resolving power of the present form of MUSIC is as good as $2/M = 2$ RL.

By the results of [1], the spectral norm of the random Hankel matrix E constructed from a zero mean, independently and identically distributed (i.i.d.) sequence of a finite variance is on the order

of $\sqrt{M \log M}$ for $M \gg 1$ while σ_s is on the order of M (with $L \approx M/2$). In this case the factor α in (10) is almost always positive for sufficiently large M regardless of the variance of noise and

$$|R^\varepsilon(\omega) - R(\omega)| = \mathcal{O}(M^{-1/2})$$

up to a logarithmic factor.

Also the super-resolution effect of MUSIC is studied. When the minimum separation between frequencies drops below 1 RL, we show that the noise level that MUSIC can tolerate obeys a power law with respect to the minimum separation with an exponent smaller than an estimate established by Donoho.

Our analysis can be easily extended to other settings where the MUSIC algorithm can be applied, such as the estimation of Directions of Arrivals (DOA) [34] and inverse scattering [8, 13, 31].

1.3 Comparison with other works

Among existing works, [20] is most closely related to the present work. Central to the results of [20] is a stability criterion expressed in terms of the Noise-to-Signal Ratio (NSR) $\mathbb{E}(\|\varepsilon\|_2)/\|y\|_2$, the dynamic range and, when the objects are located exactly on a grid of spacing ≥ 1 RL, the restricted isometry constants from the theory of compressed sensing [4]. The emphasis there is on *sparse* (i.e. undersampling), and typically random, measurement. For the gridless setting considered in the present work, the implications of the analysis in [20] are not explicit due to lack of the restricted isometry property for a well-separated set in the continuum. This barrier is overcome in the present work by the discrete Ingham inequalities and the resulting bounds on singular values.

Other closely related work includes [11] and [7] where Vandermonde decomposition of the Hankel matrix (6) are used to design different algorithms.

In [11] Demanet *et al.* proposed an approach to spectral estimation with a selection step of the support set followed by a pruning step. In the selection step, any ω satisfying $\sin \measuredangle(\phi^L(\omega), \mathrm{Range} H^\varepsilon)$ $\leq \eta$ for some judicious choice of $\eta > 0$ is kept as a frequency candidate based on their estimate

$$\sin \measuredangle(\phi^L(\omega_j), \mathrm{Range} H^\varepsilon) \leq C \frac{s\|E\|_2}{x_{\min}\sigma_{\min}(\Phi^{M-L})\|\phi^L(\omega_j)\|_2}, \quad \forall \omega_j \in \mathcal{S}$$

for some constant $C > 0$. In comparison, our estimates (10)-(12) are more comprehensive as they apply to $\mathbb{T}$ including $\mathcal{S}$ and more explicit due to discrete Ingham inequalities. In addition, the choice of the thresholding parameter η can affect the performance of the algorithm in [11] while MUSIC does not contain any thresholding parameter.

In [7] Chen and Chi exploited the low-rank property of the Hankel matrix H and applied the matrix completion technique to recover a spectrally sparse signal from its partial time-domain samples. The focus of [7] is on the completion and denoising of of data from the partial noisy samples while MUSIC is designed for frequency recovery. A combination of [7] and our work constitutes a new framework for single-snapshot spectral estimation with compressive noisy measurements which is to be discussed in Section 6.

As for frequency recovery, recent progresses center around greedy algorithms and Total Variation (TV) minimization.

The challenge of applying greedy algorithms to (5) while $N \gg M$ lies in the high coherence and ill conditioning of the sensing matrix A. In order to mitigate this effect, we exploited the coherence pattern of A and introduced the techniques of Band exclusion and Local Optimization

(BLO) to enhance standard compressive sensing algorithms. The performance guarantee in [21] assumes $q \geq 3$ RL and ensures reconstruction of $\mathcal{S}$ to the accuracy of 1 RL.

In [6, 5], Candès and Fernandez-Granda proposed TV minimization and showed that, under the assumption of $q \geq 4$ RL, the TV minimizer yields an L_1 reconstruction error linearly proportional to noise with a magnification factor proportional to F^2 where F is the refinement/super-resolution factor. Inspired by this approach, Tang *et al.* [47] developed an atomic norm (equivalent to the TV norm in 1D) minimization for the completion of y from its partial samples and showed exact reconstruction in the noise-free case. Like [20], a main emphasis in [47] is on sparse measurements. Unfortunately, the effect of noise is not considered in [47]. For numerical implementation a SemiDefinite Programming (SDP) on the dual problem is solved in [47, 5] where numerical efficiency and stable retrieval of primal solutions may become a problem.

Historically, Prony was the first to address the problem of spectral estimation [40]. Unfortunately, Prony's method is numerically unstable and numerous modifications were attempted to improve its numerical behavior. Approximate Prony Method (APM) proposed by Beylkin and Monzón in [3] is a major breakthrough for function approximation by exponential sums. Specifically, Beylking and Monzón considered the following problem: given $2N+1$ values of function $f(t)$ on a uniform grid on $[0, 1]$ and a target accuracy $\varepsilon > 0$, they find the minimal number s of complex weights w_j and complex nodes γ_j such that

$$\left| f\left(\frac{k}{2N}\right) - \sum_{j=1}^{s} w_j \gamma_j^k \right| \leq \varepsilon, \ \forall k, 0 \leq k \leq 2N.$$

Many interesting examples were provided in [3]. For instance, the Bessel function $J_0(100\pi t)$ in $[0, 1]$ is approximated by exponential sums of 28 complex nodes with accuracy $\varepsilon = 10^{-10}$ by APM.

In comparison the spectral estimation problem (1) is the identification of $\{\gamma_j\}$ from noisy data, instead of approximation of the signal. For spectral estimation with noisy data, APM's stability may be questionable. The numerical examples of spectral estimation by APM in [39] all have low NSR $= \mathcal{O}(10^{-\delta})$ where $\delta \geq 4$. In contrast our simulations in Section 5 are performed with NSR as large as 0.5. Furthermore, the super-resolution effect of MUSIC is quantitatively documented in Section 4 while it has not been reported in literature whether APM has the capability of localizing closely spaced frequencies.

In terms of discrete Ingham inequalities, Theorem 2 is the first result in which both the gap condition and the upper/lower bounds are explicitly given. In comparison semi-discrete Ingham inequalities in [32, 39, Lemma 3.1] give the correct scaling with respect to the size of the Vandermonde matrix Φ^L without an explicit estimate for the constants (cf. (17) and (20) in Section 2). In other words the previous Ingham inequalities affirm only that the matrix Φ^L has a finite condition number under certain gap condition of $\mathcal{S}$ without an explicit estimate on the magnitude of the condition number.

Detailed numerical comparisons of the MUSIC algorithm with Band-excluded Locally Optimized Orthogonal Matching Pursuit (BLOOMP) of [21], SemiDefinite Programming (SDP) of [6, 5, 47] and Matched filtering using prolates enhanced by the Band-excluded and Locally Optimized technique [18] are presented in Section 5.

Since the SVD step is its primary computational cost, MUSIC has low computational complexity compared to other existing methods. As we will also see, MUSIC is also among the most accurate algorithms. Finally, MUSIC is the only algorithm that can resolve frequencies with com-

plex amplitudes closely spaced below 1 RL. Indeed, the resolution of MUSIC can be arbitrarily small for sufficiently small noise.

The paper is organized as follows. We estimate nonzero singular values of rectangular Vandermonde matrices with nodes on the unit disk in Section 2. Perturbation theory for MUSIC is presented in in Section 3 and super-resolution effect of MUSIC is studied in Section 4. Numerical experiments are provided in Section 5. We finally conclude and discuss extensions of our current work in Section 6.

2 Vandermonde matrices with nodes on the unit circle

Performance of the MUSIC algorithm in the presence of noise is crucially dependent on σ_1 and σ_s, the maximum and minimum nonzero singular values of the noiseless Hankel data matrix. To pave a way for the stability analysis, we discuss singular values of the rectangular Vandermonde matrix Φ^L in this section.

Our estimate is motivated by the classical Ingham inequalities [28, 50, (pp.162-164)] for non-harmonic Fourier series whose exponents satisfy a gap condition. Specifically Ingham inequalities address the stability problem of complex exponential sums in the system $\{e^{2\pi i\omega_j t},\ t\in[-T/2,T/2],\ \omega_j\in\mathbb{R},\ j=1,\dots,s\}$.

Proposition 2. *Let $s\in\mathbb{N}$ and $T>0$ be given. If the ordered frequencies $\{\omega_1,\dots,\omega_s\}$ fulfill the gap condition*

$$\omega_{j+1}-\omega_j\geq q>1/T,\ j=1,\dots,s-1, \tag{14}$$

then the system of complex exponentials $\{e^{-2\pi i\omega_j t},\ t\in[-T/2,T/2],\ j=1,\dots,s\}$ form a Riesz basis of its span in $L^2[-T/2,T/2]$, i.e.,

$$\frac{2}{\pi}\Big(1-\frac{1}{T^2q^2}\Big)\|\mathrm{c}\|_2^2\leq\frac{1}{T}\int_{-T/2}^{T/2}\Big|\sum_{j=1}^{s}\mathrm{c}_j e^{-2\pi i\omega_j t}\Big|^2dt\leq\frac{4\sqrt{2}}{\pi}\Big(1+\frac{1}{4T^2q^2}\Big)\|\mathrm{c}\|_2^2 \tag{15}$$

for all complex vectors $\mathrm{c}=(\mathrm{c}_j)_{j=1}^s\in\mathbb{C}^s$.

Remark 2. *Ingham inequalities can be considered as a generalization of the Parseval's identity for non-harmonic Fourier series. The gap condition is necessary for a positive lower bound in* (15) *but the upper bound always holds.*

We prove a discrete version of Ingham inequalities.

Theorem 2. *Suppose $\mathcal{S}$ satisfies the gap condition*

$$q=\min_{j\neq l}d(\omega_j,\omega_l)>\frac{1}{L}\sqrt{\frac{2}{\pi}}\Big(\frac{2}{\pi}-\frac{4}{L}\Big)^{-\frac{1}{2}}. \tag{16}$$

When L is an even integer,

$$\Big(\frac{2}{\pi}-\frac{2}{\pi L^2q^2}-\frac{4}{L}\Big)\|\mathrm{c}\|_2^2\leq\frac{1}{L}\|\Phi^L\mathrm{c}\|_2^2\leq\Big(\frac{4\sqrt{2}}{\pi}+\frac{\sqrt{2}}{\pi L^2q^2}+\frac{3\sqrt{2}}{L}\Big)\|\mathrm{c}\|_2^2,\ \forall\mathrm{c}\in\mathbb{C}^s. \tag{17}$$

In other words,

$$\frac{1}{L}\sigma_{\max}^2(\Phi^L) \le \frac{4\sqrt{2}}{\pi} + \frac{\sqrt{2}}{\pi L^2 q^2} + \frac{3\sqrt{2}}{L} \tag{18}$$

and

$$\frac{1}{L}\sigma_{\min}^2(\Phi^L) \ge \frac{2}{\pi} - \frac{2}{\pi L^2 q^2} - \frac{4}{L}. \tag{19}$$

When L is an odd integer,

$$\left(\frac{2}{\pi} - \frac{2}{\pi L^2 q^2} - \frac{4}{L}\right)\|\mathrm{c}\|_2^2 \le \frac{1}{L}\|\Phi^L \mathrm{c}\|_2^2 \le \left(1+\frac{1}{L}\right)\left(\frac{4\sqrt{2}}{\pi} + \frac{\sqrt{2}}{\pi(L+1)^2 q^2} + \frac{3\sqrt{2}}{L+1}\right)\|\mathrm{c}\|_2^2,\ \forall \mathrm{c} \in \mathbb{C}^s. \tag{20}$$

Proof of Theorem 2 is provided in Appendix B.

Remark 3. *The difference between the bounds of the discrete and the continuous Ingham inequalities is $\mathcal{O}(1/L)$ which is negligible when L is large. The upper bound in* (17) *holds even when the gap condition* (16) *is violated; however,* (16) *is necessary for the positivity of the lower bound.*

Remark 4. *Some form of discrete Ingham inequalities are developed in [37, 38] for the analysis of the control/observation properties of numerical schemes of the 1-d wave equation. The main result therein is that when time integrals in* (15) *are replaced by discrete sums on a discrete mesh, discrete Ingham inequalities converge to the continuous one as the mesh becomes infinitely fine. Their asymptotic analysis, however, do not provide the non-asymptotic results stated in Theorem 2.*

3 Perturbation of noise-space correlation

In this section we use tools in classical matrix perturbation theory [45, 29] and develop a perturbation estimate on the noise-space correlation function, the key ingredient of the MUSIC algorithm. Our main results are presented in Theorem 3, Corollary 1 and Theorem 4 and proofs are provided in Appendix C.1 and C.2.

Theorem 3. *Suppose $L \ge s$, $M - L + 1 \ge s$ and $\|E\|_2 < \sigma_s$. Then*

$$|R^\varepsilon(\omega) - R(\omega)| \le \|\mathcal{P}_2^\varepsilon - \mathcal{P}_2\|_2 := \sup_{\phi \in \mathbb{C}^{L+1}} \frac{\|\mathcal{P}_2^\varepsilon \phi - \mathcal{P}_2 \phi\|_2}{\|\phi\|_2} \le \frac{4\sigma_1 + 2\|E\|_2}{(\sigma_s - \|E\|_2)^2}\|E\|_2. \tag{21}$$

In particular, for $\omega_j \in \mathcal{S}$, $R(\omega_j) = 0$ and

$$|R^\varepsilon(\omega_j)| \le \frac{2\|E\|_2}{x_{min}\sigma_{\min}((\Phi^{M-L})^T)\|\phi^L(\omega_j)\|_2},\ j = 1, 2, \ldots, s. \tag{22}$$

Remark 5. *While* (21) *is a general perturbation estimate valid on $\mathbb{T}$, including $\mathcal{S}$,* (22) *is a sharper estimate for $\omega_j \in \mathcal{S}$.*

Remark 6. *Suppose noise vector ε contains* i.i.d. *random variables of variance σ^2, $\|E\|_2 \le \|E\|_F = \mathcal{O}(\sigma)$. For fixed M and $\mathcal{S}$,*

$$|R^\varepsilon(\omega) - R(\omega)| = \mathcal{O}(\sigma).$$

Theorem 3 holds for all signal models with *any* support set $\mathcal{S}$. In view of the Vandermonde decomposition (7) we next derive explicit bounds for the perturbation of noise-space correlation by combining Theorem 2 and 3.

Corollary 1. *Let L and $M-L$ be even integers. Suppose $\mathcal{S}$ satisfies the following gap condition*

$$q = \min_{j \neq l} d(\omega_j, \omega_l) > \max\left(\frac{1}{L}\sqrt{\frac{2}{\pi}}\left(\frac{2}{\pi} - \frac{4}{L}\right)^{-\frac{1}{2}}, \frac{1}{M-L}\sqrt{\frac{2}{\pi}}\left(\frac{2}{\pi} - \frac{4}{M-L}\right)^{-\frac{1}{2}} \right). \tag{23}$$

Then

$$|R^{\varepsilon}(\omega) - R(\omega)| \leq \frac{4\alpha_1 + 2\frac{\|E\|_2}{\sqrt{L(M-L)}}}{\left(\alpha_2 - \frac{\|E\|_2}{\sqrt{L(M-L)}}\right)^2} \cdot \frac{\|E\|_2}{\sqrt{L(M-L)}}, \tag{24}$$

where

$$\alpha_1 = x_{max}\sqrt{\left(\frac{4\sqrt{2}}{\pi} + \frac{\sqrt{2}}{\pi L^2 q^2} + \frac{3\sqrt{2}}{L}\right)\left(\frac{4\sqrt{2}}{\pi} + \frac{\sqrt{2}}{\pi(M-L)^2 q^2} + \frac{3\sqrt{2}}{M-L}\right)} \tag{25}$$

and

$$\alpha_2 = x_{min}\sqrt{\left(\frac{2}{\pi} - \frac{2}{\pi L^2 q^2} - \frac{4}{L}\right)\left(\frac{2}{\pi} - \frac{2}{\pi(M-L)^2 q^2} - \frac{4}{M-L}\right)}. \tag{26}$$

Remark 7. *Corollary 1 is stated in the case that both L and $M-L$ are even integers. However, the results hold in other cases with a slightly different α_1 based on* (17) *and* (20).

Remark 8. *As noted in Section 1.2, for i.i.d. noise $\|E\|_2$ grows like $\sqrt{M \log M}$ which is much smaller than $\sqrt{L(M-L)}$ with $L \sim M$ as $M \to \infty$. As a consequence,*

$$|R^{\varepsilon}(\omega) - R(\omega)| = \mathcal{O}\left(\frac{\sqrt{\log M}}{\sqrt{M}}\right)$$

under the gap condition (23).

How close are the MUSIC estimates, namely the s lowest local minimizers of $R^{\varepsilon}(\omega)$, to the true frequencies which are the zeros of $R(\omega)$? While we can not at the moment answer this question directly, the following asymptotic result says that every true frequency has a near-by strict local minimizer of R^{ε} in its vicinity converging to it. Denote $[\phi^L(\omega)]' = d\phi^L(\omega)/d\omega$.

Theorem 4. *Suppose*

$$\mathcal{P}_2[\phi^L(\omega)]'\big|_{\omega=\omega_j} \neq \mathbf{0}, \quad \forall \omega_j \in \mathcal{S}. \tag{27}$$

When $\|E\|_2$ is sufficiently small (details in (63) *and* (64)*), there exists a strict local minimizer $\hat{\omega}_j$ of $R^{\varepsilon}(\omega)$ near ω_j such that*

$$|\hat{\omega}_j - \omega_j| \min_{\xi \in (\omega_j, \hat{\omega}_j)} |Q''(\xi)| \leq 4\alpha\eta(L)\|E\|_2 \tag{28}$$

where $Q(\omega) = R^2(\omega)$, $\eta(L) = 2\pi\sqrt{1^2 + 2^2 + \ldots + L^2}/\sqrt{L+1}$ *and*

$$\alpha = \frac{4\sigma_1 + 2\|E\|_2}{(\sigma_s - \|E\|_2)^2}.$$

Remark 9. *For* i.i.d. *random variables of variance* σ^2, $\|E\|_2 = \mathcal{O}(\sigma)$ *for fixed* M. *Hence*

$$|\hat{\omega}_j - \omega_j| = \mathcal{O}(\sigma)$$

as $\sigma \to 0$.

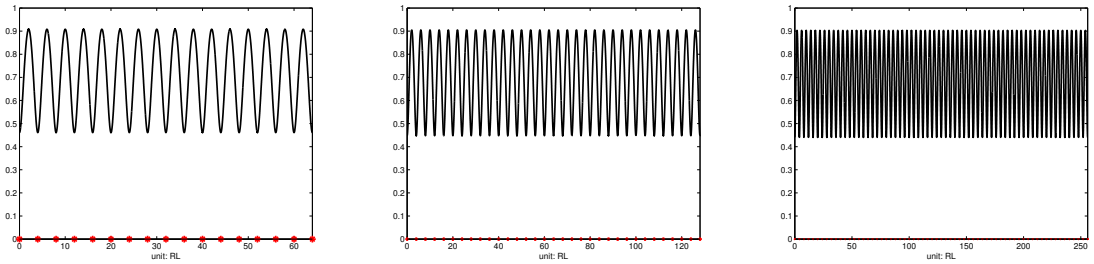
Remark 10. *In view of the identity*

$$Q''(\omega_j) = \frac{2}{L+1}\|\mathcal{P}_2[\phi^L(\omega)]'\|_2^2\Big|_{\omega=\omega_j} \neq 0, \quad \forall j$$

cf. (59), assumption (27) is a generic condition that says the true frequencies are not degenerate minimizers of R^2. *Figure 2 shows* $\|\mathcal{P}_2[\phi^L(\omega)]'\|_2/\|[\phi^L(\omega)]'\|_2$ *for various* M *and suggests that for* $q \geq 4$ *RL,*

$$C_1\|[\phi^L(\omega)]'\|_2^2 \leq \|\mathcal{P}_2[\phi^L(\omega)]'\|_2^2 \leq C_2\|[\phi^L(\omega)]'\|_2^2, \ \forall\ \omega \in [0,1) \tag{29}$$

for some constants $C_1, C_2 > 0$ *independent of* M.



(a) $M = 64$ (b) $M = 128$ (c) $M = 256$

Figure 2: Function $\|\mathcal{P}_2[\phi^L(\omega)]'\|_2/\|[\phi^L(\omega)]'\|_2$ with varied M in the case that frequencies in $\mathcal{S}$ are separated by 4 RL. $L = M/2$ and $\mathcal{S}$ is marked by red dots.

Remark 11. *Under the assumption (23), the right hand side of (28) scales like* $\sigma\sqrt{M \log M}$ *with* $L = M/2$ *for i.i.d. random noise of variance* σ^2. *Under the assumption* (29),

$$|Q''(\omega_j)| = \frac{2}{L+1}\|\mathcal{P}_2[\phi^L(\omega_j)]'\|_2^2 \geq \frac{2C_1\|[\phi^L(\omega_j)]\|_2^2}{L+1} = \frac{2C_1(1^2 + 2^2 + \ldots + L^2)}{L+1} = \mathcal{O}(L^2).$$

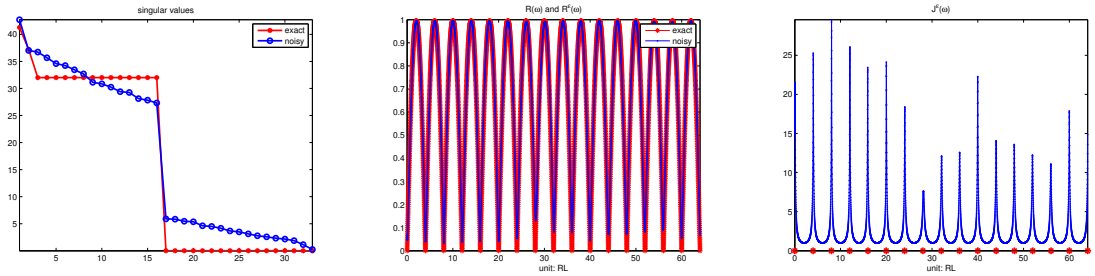
As shown in the proof of Theorem 4 (Step 1),

$$\min_{\xi \in (\omega_j, \hat{\omega}_j)} |Q''(\xi)| > \frac{1}{2}Q''(\omega_j) = \mathcal{O}(M^2), \quad M = 2L \gg 1.$$

As a result

$$|\hat{\omega}_j - \omega_j| = \mathcal{O}\left(\frac{\sqrt{\log M}}{M^{3/2}}\right).$$

Strong stability in the case of well separated objects is demonstrated in Figure 3. Let $M = 64$, $L = 32$ and then 1 RL $= 1/64$. External noise is i.i.d. Gaussian noise, i.e. $\varepsilon \sim N(0, \sigma^2 I)$. Noise-to-Signal Ratio (NSR) $= \mathbb{E}(\|\varepsilon\|_2)/\|y\|_2 = 20\%$. We display singular values of H and H^ε in Figure 3(a), noise-space correlation function $R(\omega)$ and $R^\varepsilon(\omega)$ in Figure 3(b), and imaging function $J^\varepsilon(\omega)$ in Figure 3(c). With a minimum separation of 4 RL imposed on $\mathcal{S}$, $R^\varepsilon(\omega)$ is stably perturbed from $R(\omega)$. More importantly, every peak of $J^\varepsilon(\omega)$ is near a true object in $\mathcal{S}$. Simply extracting s largest local maxima of $J^\varepsilon(\omega)$ yields a good reconstruction.



(a) Singular values of H and H^ε (b) $R(\omega)$ and $R^\varepsilon(\omega)$ (c) $J^\varepsilon(\omega)$

Figure 3: Perturbation of noise-space correlation and imaging function in the case where objects are separated by 4 RL and NSR $= 20\%$. In (c) true objects are located at red dots.

4 Super-resolution effect of MUSIC

Super-resolution refers to the capability of resolving frequencies separated below 1 RL. The super-resolution effect of MUSIC has been numerically demonstrated in various applications [13, 34, 20, 46], but theoretical guarantees are lacking. In this section we aim to analyze the super-resolution effect in light of the preceding results and that of [15].

With $2s$ noiseless data MUSIC guarantees to exactly recover s distinct frequencies. With noisy data, the resulting perturbation of the noise-space correlation function depends crucially on σ_1 and σ_s (Theorem 3). As

$$\sigma_1 \leq \sigma_{\max}(\Phi^L) x_{\max} \sigma_{\max}(\Phi^{M-L}), \tag{30}$$

$$\sigma_s \geq \sigma_{\min}(\Phi^L) x_{\min} \sigma_{\min}(\Phi^{M-L}), \tag{31}$$

the larger $\sigma_{\min}(\Phi^L), x_{\min}, \sigma_{\min}(\Phi^{M-L})$ and the smaller $\sigma_{\max}(\Phi^L), x_{\max}, \sigma_{\max}(\Phi^{M-L})$ are, the less sensitive MUSIC is to noise. It follows that a close-to-unity dynamic range $x_{\max}/x_{\min}$ and good conditioning of Φ^L and Φ^{M-L} are conducive to the stability of MUSIC.

In particular, the denominator $(\sigma_s - \|E\|_2)^2$ on the right hand side of (21) indicates that the amount of noise that can be tolerated by MUSIC is approximately σ_s, which by (31) is at least $x_{\min}\sigma_{\min}(\Phi^L)\sigma_{\min}(\Phi^{M-L})$.

Hence, to understand the super-resolution effect of MUSIC it is essential to estimate the smallest nonzero singular value of Φ^L with closely spaced frequencies. Under certain weakened gap conditions proposed in [2, 15], we provide an explicit upper bound on $\sigma_{\max}(\Phi^L)$ and discuss the possible implication of the bounds on $\sigma_{\min}(\Phi^L)$ in [15] on the super-resolution effect.

4.1 Maximum singular value of Φ^L

Motivated by [2, 15], we consider a sequence $\mathcal{S} = \{\omega_j : j \in \mathbb{Z}\}$ which is *s-periodic* in the sense that $\mathcal{S} \cap [0, 1) = \{\omega_1, \omega_2, \ldots, \omega_s\}$ and $\omega_{j+ks} = k + \omega_j, \forall k \in \mathbb{Z}, j = 1, \ldots, s$. Without loss of generality, we assume $0 \le \omega_1 < \omega_2 < \cdots < \omega_s < 1$.

Define

$$B(q, L) = \begin{cases} \frac{4\sqrt{2}}{\pi} + \frac{\sqrt{2}}{\pi L^2 q^2} + \frac{3\sqrt{2}}{L} & \text{if } L \text{ is even} \\ \left(1 + \frac{1}{L}\right)\left(\frac{4\sqrt{2}}{\pi} + \frac{\sqrt{2}}{\pi (L+1)^2 q^2} + \frac{3\sqrt{2}}{L+1}\right) & \text{if } L \text{ is odd.} \end{cases}$$

Theorem 5. *Suppose $\mathcal{S}$ is an s-periodic sequence and satisfies the following weakened gap condition*

$$|\omega_{j+R} - \omega_j| > R\rho, \ \ j = 1, \ldots, s, \tag{32}$$

for some $R \in \mathbb{Z}^+$ and $\rho \in \mathbb{R}^+$. Then

$$\frac{1}{L} \sum_{k=0}^{L} \left| \sum_{j=1}^{s} \mathrm{c}_j e^{-2\pi i k \omega_j} \right|^2 \le B(R\rho, L) R \|\mathrm{c}\|_2^2, \ \ \forall \mathrm{c} \in \mathbb{C}^s. \tag{33}$$

In other words,

$$\frac{1}{L} \sigma_{\max}^2(\Phi^L) \le B(R\rho, L) R. \tag{34}$$

Proof of Theorem 5 is in Appendix D.

By choosing R and ρ such that $R\rho = 2/L$, in which case there are at most R frequencies in any interval of 4 RL (1 RL = 1/(2L) when L = M/2), we can maintain $B(R\rho, L)$ roughly independent of $L \gg 1$ and obtain the (asymptotic) upper bound $\frac{17\sqrt{2}}{4\pi} R$ for $\sigma_{\max}^2(\Phi^L)/L$. It is noteworthy that the minimum separation between two consecutive frequencies significantly affect the least nonzero singular value but not the largest one.

4.2 Minimum nonzero singular value of Φ^L

Presently we can not prove an *explicit* lower bound for $\sigma_{\min}(\Phi^L)$ when two or more frequencies are spaced below 2 RL (1 RL = 1/(2L) when L = M/2).

Now we recall and reformulate the lower bound established by Donoho [15]. Let $\mathcal{S} = \{\omega_j\}_{j \in \mathbb{Z}}$ be a subset of the lattice $\mathcal{L}(\Delta) = \{k\Delta, k \in \mathbb{Z}\}$ of spacing Δ. Let the Rayleigh index R_* be the *least* positive integer such that

$$|\omega_{j+R_*} - \omega_j| > 2R_*/L, \quad \forall j. \tag{35}$$

In other words, R_* is the size of the largest cluster whose members are separated from each other by less than 4 RL. Define

$$\nu(\Delta, L, R_*) = \min_{\|c\|_2 = 1} \left(\int_{-L/2}^{L/2} \Big| \sum_{j \in \mathbb{Z}} \mathrm{c}_j e^{-2\pi i \omega_j t} \Big|^2 dt \right)^{\frac{1}{2}}. \tag{36}$$

According to [15]

$$\nu(\Delta, L, R_*) \geq \Delta^{2R_*+1}\alpha(L, R_*) \tag{37}$$

where $\alpha(L, R_*)$ is some positive constant depending on L and R_*. No algorithm is proposed for support reconstruction in [15].

The definition (36) is the continuous analog of

$$\sigma_{\min}(\Phi^L) = \min_{\|c\|_2=1} \left(\sum_{k=0}^{L} \Big| \sum_{j=1}^{s} \mathrm{c}_j e^{-2\pi i k \omega_j} \Big|^2 \right)^{\frac{1}{2}}. \tag{38}$$

Hence by making the identification $\Delta = \min_{i \neq j} d(\omega_i, \omega_j) = q$, it is plausible that, even with discrete data and without the lattice substrate, the bound

$$\sigma_{\min}(\Phi^L) \geq q^{2R_*+1}\tilde{\alpha}(L, R_*) \tag{39}$$

holds with some positive constant $\tilde{\alpha}(L, R_*)$ under the condition (35).

Our preceding analysis in Sections 2 and 3 is for the case $R_* = 1$.

As commented above, for a support set with Rayleigh index R_*, the amount of noise that can be tolerated by MUSIC is approximately σ_s which, according to (39), decays at worst like $x_{\min} q^{4R_*+2}$ as $q \to 0$. Our numerical experiments in Section 5.4 show that the noise level that MUSIC can tolerate obeys $q^{e(R_*)}$ for $R_* = 2, 3, 4, 5$ and $e(2) = 3.6691$, $e(3) = 6.0565$, $e(4) = 8.3861$ and $e(5) = 11.2392$, suggesting that $e(R_*) \approx 2.504R_* - 1.4262$.

5 Numerical experiments

A systematic numerical simulation is performed on MUSIC, BLOOMP, SDP and Matched Filtering using prolates in this section, showing that MUSIC combines the advantages of strong stability and low computation complexity for the detection of well-separated frequencies and furthermore only MUSIC yields an exact reconstruction in the noise-free case regardless of the distribution of true frequencies and processes the capability of resolving closely spaced frequencies.

5.1 Algorithms tested.

We compare the performances of various algorithms on the spectral estimation problem (1) with $M = 100$ and i.i.d. Gaussian noise, i.e. $\varepsilon \sim N(0, \sigma^2 I) + iN(0, \sigma^2 I)$. Define the

$$\text{Noise-to-Signal Ratio (NSR)} = \mathbb{E}(\|\varepsilon\|_2)/\|y\|_2 = \sigma\sqrt{2(M+1)}/\|y\|_2. \tag{40}$$

We test and compare the following algorithms.

1. The MUSIC algorithm: As suggested by (24) in Corollary 1 we set $M = 2L$.

2. Band-excluded and Locally Optimized Orthogonal Matching Pursuit (BLOOMP) [21]: BLOOMP works with an arbitrarily fine grid with grid spacing $\ell = \mathrm{RL}/F$ where F is the refinement/super-resolution factor. In the presence of noise it is unnecessary to set an extremely large F. A rule of thumb for the problem of spectral estimation is that $F \sim \mathrm{SNR}$ gives rise to a gridding

error comparable to the external noise [21, Fig. 1]. For instance $F = 20$ is adequate when SNR $\geq 5\%$. When frequencies are separated above 3 RL (i.e., in Fig. 4(b), Fig. 5(b) and Fig. 6(a)(b)), we can set the radii of excluded band and local optimization to be 2 RL and 1 RL, respectively. When frequencies are separated between 2 RL and 3 RL (i.e., in Fig. 6(c)(d), the radii of excluded band and local optimization is set to be 1 RL.

3. SemiDefinite Programming (SDP) [5, 47]: The code is from `http://www.stanford.edu/~cfgranda/superres_sdp_noisy.m` where SDP is solved through CVX, a package for specifying and solving convex programs [24, 25]. Output of SDP is the dual solution of total variation minimization. In the code, frequencies are identified through root findings of a polynomial and amplitudes are solved through least squares. Let $\tilde{\omega} = [\tilde{\omega}_j]_{j=1}^n \in \mathbb{R}^n$ be the frequencies retrieved from the code and $\tilde{x} = [\tilde{x}_j]_{j=1}^n \in \mathbb{C}^n$ be the corresponding amplitudes. Usually n is greater than s. A straightforward way of extracting s reconstructed frequencies is by Hard Thresholding (HT), i.e., picking the frequencies corresponding to the s largest amplitudes in $\tilde{x}$. We also test if the Band Excluded Thresholding (BET) technique introduced in [21] can improve on hard thresholding and enhance the performance of SDP (Fig. 6). BET amounts to trimming $\tilde{\omega} \in \mathbb{R}^n$ to $\hat{\omega} \in \mathbb{R}^s$ as follows.

Band Excluded Thresholding (BET)
Input: $\tilde{\omega}, \tilde{x}, s, r$ (radius of excluded band).
Initialization: $\hat{\omega} = [\ ]$.
Iteration: for $k = 1, \ldots, s$
1) Find j such that $\lvert\tilde{x}_j\rvert = \max_i \lvert\tilde{x}_i\rvert$.
If $\tilde{x}_j = 0$, then go to **Output**.
2) Update the support vector: $\hat{\omega} = [\hat{\omega}\ ;\ \tilde{\omega}_j]$.
3) For $i = 1 : n$
If $\tilde{\omega}_i \in (\tilde{\omega}_j - r, \tilde{\omega}_j + r)$, set $\tilde{x}_i = 0$.
Output: $\hat{\omega}$.

When frequencies are separated by at least 4 RL, we choose $r = 1$ RL.

4. Matched Filtering (MF) using prolates: In [18], Eftekhari and Wakin use matched filtering windowed by the Discrete Prolate Spheroidal (Slepian) Sequence [44] for the same problem while frequencies are extracted by band-excluded and locally optimized thresholding proposed in [21]. In its current form, MF using prolates can not deal with complex-valued amplitudes so it is tested with real-valued amplitudes only.

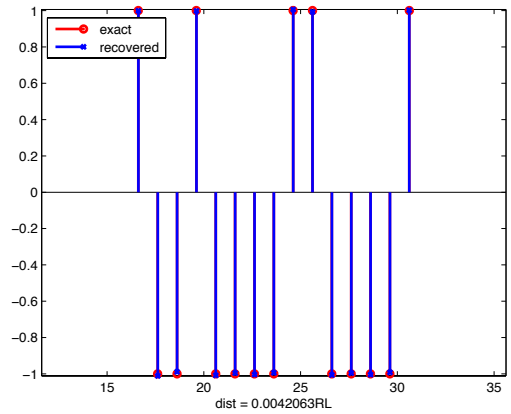
Reconstruction error is measured by Hausdorff distance between the exact ($\mathcal{S}$) and the recovered ($\hat{\mathcal{S}}$) sets of frequencies:

$$d(\hat{\mathcal{S}}, \mathcal{S}) = \max\left\{\max_{\hat{\omega}\in\hat{\mathcal{S}}} \min_{\omega\in\mathcal{S}} d(\hat{\omega}, \omega)\ ,\ \max_{\omega\in\mathcal{S}} \min_{\hat{\omega}\in\hat{\mathcal{S}}} d(\hat{\omega}, \omega)\right\}. \tag{41}$$

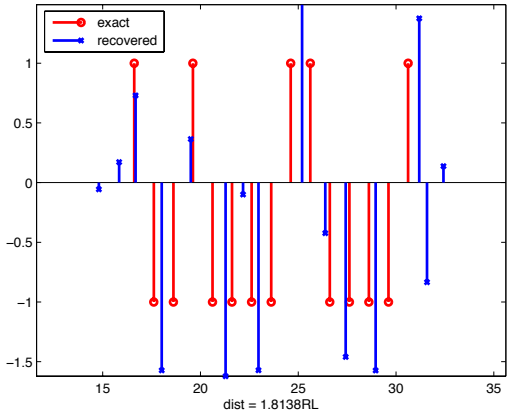
5.2 Noise-free case

In the noise-free case only MUSIC processes a theory of exact reconstruction regardless of the distribution of true frequencies. In theory, BLOOMP requires a separation of 3 RL for approximate

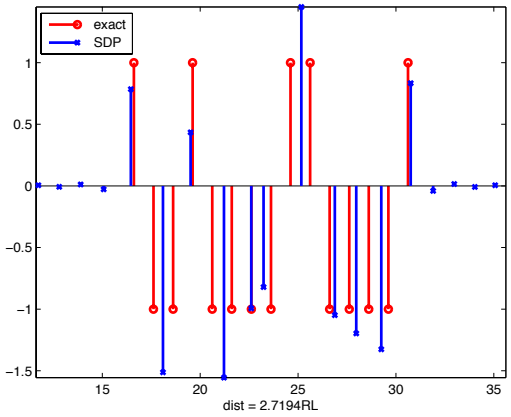
support recovery while SDP requires a separation of 4 RL for exact recovery. In this test we use the four algorithms to recover 15 real-valued amplitudes separated by 1 RL. Figure 4 shows that MUSIC achieves the accuracy of about 0.004 RL while BLOOMP, SDP and MF using prolates essentially fail, which implies that certain separation condition is necessary for BLOOMP, SDP and MF using prolates.



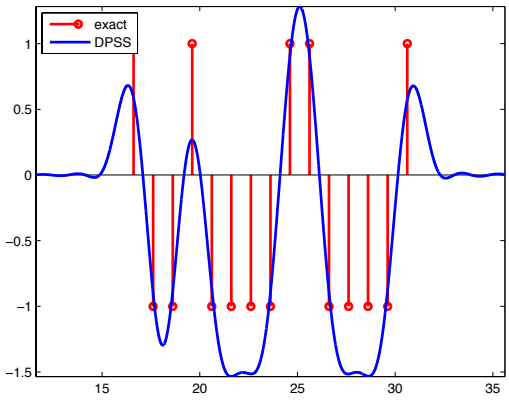
(a) MUSIC. Red: exact; Blue: recovered. $d(\hat{\mathcal{S}}, \mathcal{S}) \approx 0.004$ RL.



(b) BLOOMP. Red: exact; Blue: recovered. $d(\hat{\mathcal{S}}, \mathcal{S}) \approx 1.81$ RL.



(c) SDP. Red: exact; Blue: recovered. Hard thresholding yields $d(\hat{\mathcal{S}}, \mathcal{S}) \approx 2.72$ RL.

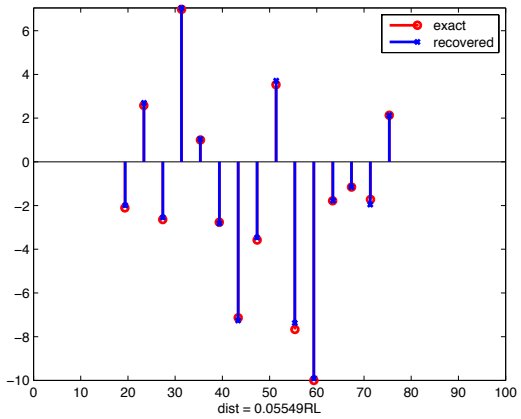


(d) MF using prolates. Red: exact; Blue: inverse Fourier transform of y windowed by the first DPSS sequence.

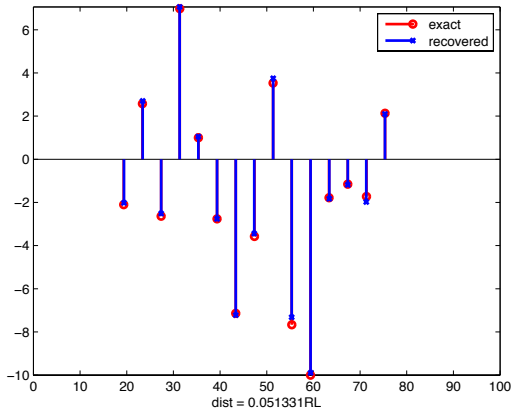
Figure 4: Reconstruction of 15 real-valued frequencies separated by 1 RL. Dynamic range = 1 and NSR = 0%.

5.3 Detection of well-separated frequencies

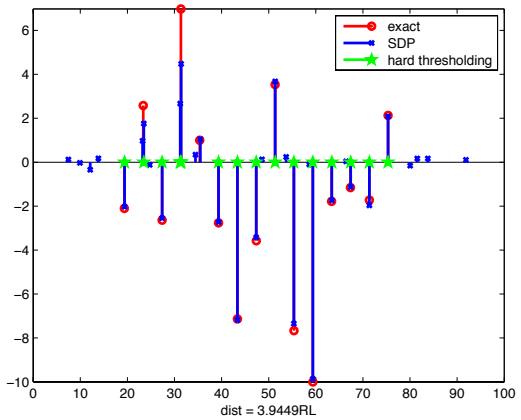
Figure 5 shows reconstructions of 15 real-valued frequencies separated by 4 RL. By extracting 15 largest local maxima of the imaging function $J^{\varepsilon}(\omega)$, MUSIC yields a reconstruction distance of 0.06 RL. As predicted by the theory in [21], every recovered object of BLOOMP is within 1



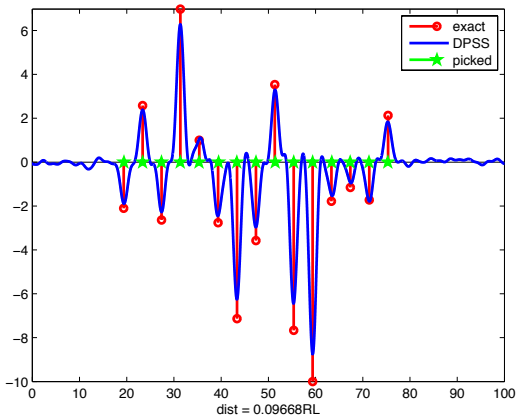
(a) MUSIC. Red: exact; Blue: recovered. $d(\hat{\mathcal{S}}, \mathcal{S}) \approx 0.06$ RL.



(b) BLOOMP. Red: exact; Blue: recovered. $d(\hat{\mathcal{S}}, \mathcal{S}) \approx 0.05$ RL.



(c) SDP. Red: exact; Blue: Primal solution of SDP. Hard thresholding (green) yields $d(\hat{S}, \mathcal{S}) \approx 3.94$ RL. The true amplitude around 33 RL is recovered as two amplitudes and the BET technique can be used to eliminate the smaller one in the step of frequency selection.



(d) MF using prolates. Red: exact; Blue: inverse Fourier transform of y^{ε} windowed by the first DPSS sequence; Green: frequencies selected by the BLO technique. $d(\hat{\mathcal{S}}, \mathcal{S}) \approx 0.10$ RL.

Figure 5: Reconstruction of 15 real-valued amplitudes separated by 4 RL. Dynamic range = 10 and NSR = 10%.

RL distance from a true one. Indeed, in this simulation BLOOMP achieves the best accuracy of 0.05 RL among tested algorithms. The primal solution of SDP is usually not s-sparse and the recovered frequencies tend to cluster around the true ones [5] which degrades the accuracy. The Hausdorff distance between the recovered spikes with the s strongest amplitudes and the true frequencies is 3.94 RL in this simulation. The BET technique can be used to enhance the accuracy of reconstruction and achieve the accuracy of 0.13 RL. Similarly the BLO technique introduced in [21] can be applied to improve the result of Matched filtering windowed by the DPSS sequence (the blue curve in Figure 5(d)) and achieve the accuracy of 0.10 RL.

Figure 6 shows the average errors of 100 trials by SDP with HT, BET-enhanced SDP, BLOOMP and MUSIC for complex-valued objects separated between 4 RL and 5 RL (Fig. 6(a)(b)) or separated between 2 RL and 3 RL (Fig. 6(c)(d)) versus NSR when dynamic range = 1 (Fig. 6(a)(c)) and when dynamic range = 10 (Fig. 6(b)(d)). In this simulation $[0, 1)$ are fully occupied by frequencies satisfying the separation condition and amplitudes x are complex-valued with random phases. Refinement factor F in BLOOMP is adaptive according to the rule: $F = \max(5, \min(1/\mathrm{NSR}, 20))$. Figure 6 shows that BLOOMP is the stablest algorithm while frequencies are separated above 4 RL and MUSIC becomes the stablest one while frequencies are separated between 2 RL and 3 RL. Simply extracting s largest amplitudes from the SDP solution (black curve) is not a good idea and the BET technique (green curve) can mitigate the problem with SDP. The average running time in Figure 6 shows that MUSIC takes about 0.33s for one experiment and is the most efficient one among all methods being tested. SDP needs about 20.5s for one experiment and is computationally most expensive. Running time of BLOOMP is dependent on sparsity s and refinement factor F. The running time of BLOOMP in Fig. 6(c)(d) is more than the time in Fig. 6(a)(b) as $s \in [33, 50]$ in Fig. 6(c)(d) and $s \in [20, 25]$ in Fig. 6(a)(b).

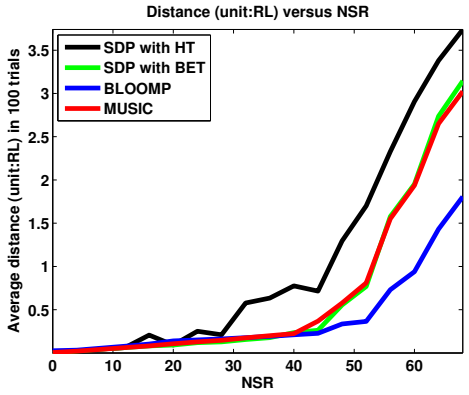
5.4 Super-resolution of MUSIC

Theory in Section 4 implies that MUSIC has super-resolution effect and moreover the noise level that MUSIC can handle follows a power law with respect to the minimum separation of the frequencies. We numerically investigate the $2, 3, 4, 5$-point resolution of MUSIC here as numerical verification.

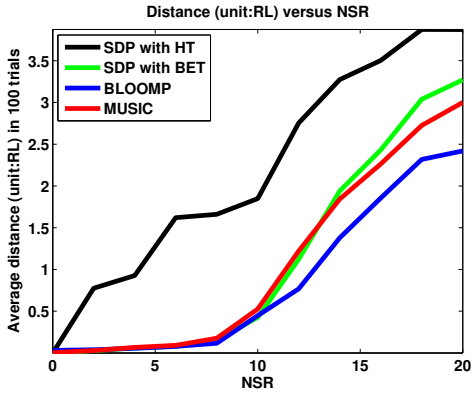
In Figure 7, we consider support set $\mathcal{S}$ containing two, three, four and five equally spaced frequencies. We run MUSIC algorithm on reconstructions of randomly phased complex objects supported on $\mathcal{S}$ with varied separation q and varied NSR for 100 trials and record the average of $d(\mathcal{S}, \hat{\mathcal{S}})/q$. Figure 7 (a)-(d) displays the color plot of the logarithm to the base 2 of average $d(\mathcal{S}, \hat{\mathcal{S}})/q$ with respect to NSR (y-axis) and q (x-axis) in the unit of RL. Frequency localization is considered successful if $d(\mathcal{S}, \hat{\mathcal{S}})/q < 1/2$.

A phase transition occurs in (a)-(d), manifesting MUSIC's capability of resolving two, three, four and five closely spaced complex-valued objects if NSR is below certain level. Theory in Section 4 indicates the noise level that MUSIC can handle scales at worst like q^{4R_*+2} where $R_* = 2$ in Figure 7(a), $R_* = 3$ in Figure 7(b), $R_* = 4$ in Figure 7(c) and $R_* = 5$ in Figure 7(d). The borderline between successful recovery and failure defined by $d(\mathcal{S}, \hat{\mathcal{S}})/q = 1/2$ are marked out in black in Figure 7 (a)-(d). The phase transition curves for $R_* = 2, 3, 4, 5$ are shown in (e) in the ordinary scale and in (f) in log-log scale. It appears that the transition curves can be fitted to a constant times $q^{e(R_*)}$ with $e(2) = 3.6691$, $e(3) = 6.0565$, $e(4) = 8.3861$ and $e(5) = 11.2392$, suggesting that a much smaller exponent $e(R_*) \approx 2.504R_* - 1.4262$ than $4R_* + 2$.

Frequencies separated between 4 RL and 5 RL

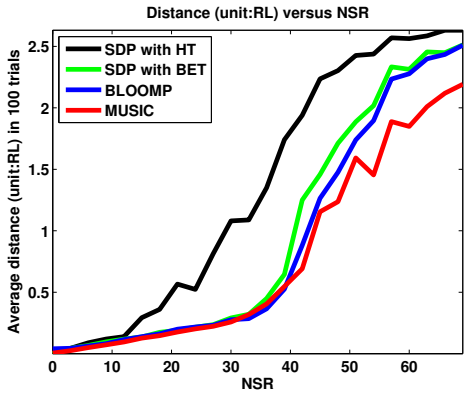


(a) Dynamic range = 1. Average running time for SDP and MUSIC in one experiment is 20.3583s and 0.3627s while the average running time for BLOOMP is 6.3420s ($F = 20$), 3.2788s ($F = 10$) and 1.7610($F = 5$).

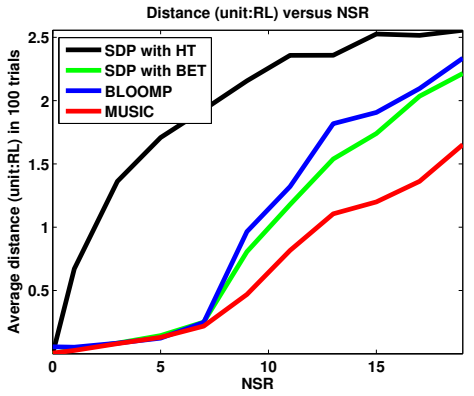


(b) Dynamic range = 10. Average running time for SDP and MUSIC in one experiment is 20.5913s and 0.3661s while the average running time for BLOOMP is 6.2623s ($F = 20$), 3.3030s ($F = 10$) and 1.7542s ($F = 5$).

Frequencies separated between 2 RL and 3 RL

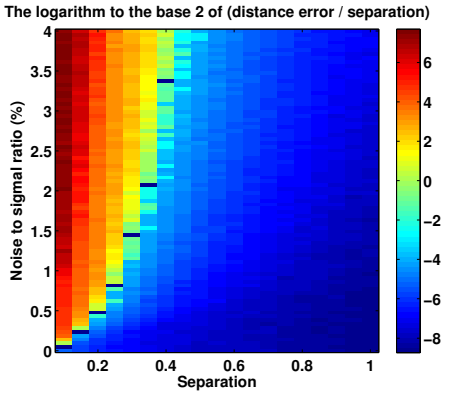


(c) Dynamic range = 1. Average running time for SDP and MUSIC in one experiment is 20.6750s and 0.3334s while the average running time for BLOOMP is 19.8357s ($F = 20$), 11.2349s ($F = 9$) and 8.0850s ($F = 5$).

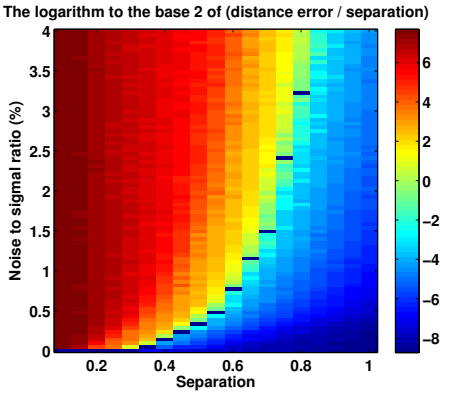


(d) Dynamic range = 10. Average running time for SDP and MUSIC in one experiment is 21.0572s and 0.3321s while the average running time for BLOOMP is 19.9233s ($F = 20$), 11.5190s ($F = 9$) and 8.1054s ($F = 5$).

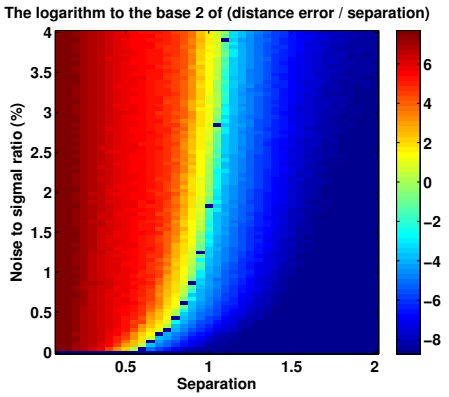
Figure 6: Average error by SDP with HT, BET-enhanced SDP, BLOOMP and MUSIC on complex-valued objects separated between 4 RL and 5 RL (a)(b) or separated between 2 RL and 3 RL (c)(d) versus NSR when dynamic range = 1 (a)(c) and when dynamic range = 10 (b)(d). MF using prolates is not included since it is not designed for complex amplitudes.



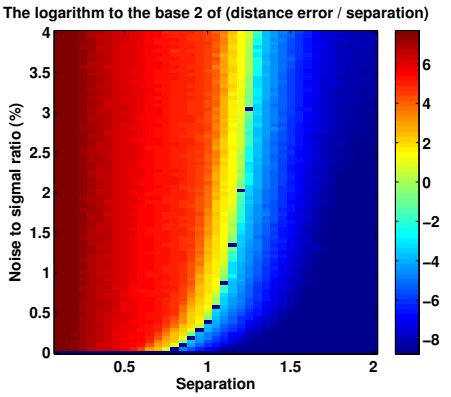
(a) Two-point resolution of MUSIC, $R_* = 2$



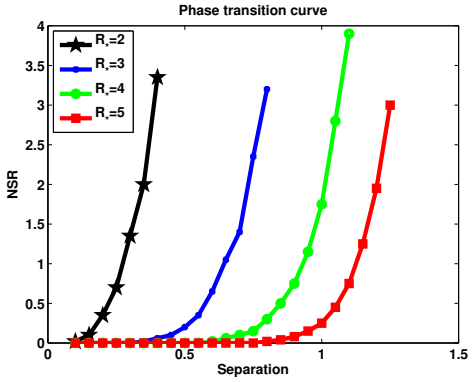
(b) Three-point resolution of MUSIC, $R_* = 3$



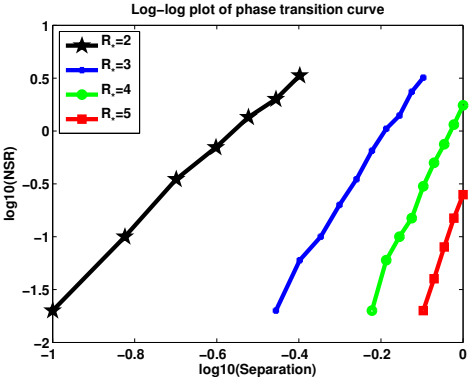
(c) Four-point resolution of MUSIC, $R_* = 4$



(d) Five-point resolution of MUSIC, $R_* = 5$



(e) Phase transition curves



(f) Log-log plot of phase transition curves

Figure 7: Color plots in (a)-(d) shows the logarithm to the base 2 of average $d(\mathcal{S}, \hat{\mathcal{S}})/q$ with respect to NSR (y-axis) and q (x-axis) in the unit of RL. Reconstruction is considered successful if $d(\mathcal{S}, \hat{\mathcal{S}})/q < 1/2$ (from green to black). A clear phase transition is observed. Transition points from which $d(\mathcal{S}, \hat{\mathcal{S}})/q < 1/2$ are marked out by black bars in (a)-(d). Phase-transition curves connecting the black bars for $R_* = 2, 3, 4, 5$ are shown in (e) and in (f) in log-log scale. Least squares fitting for the slope of the log-log plot yields $e(2) = 3.6691$, $e(3) = 6.0565$, $e(4) = 8.3861$ and $e(5) = 11.2392$.

6 Conclusion and extension

We have provided a stability analysis of the MUSIC algorithm for single-snapshot spectral estimation off the grid. We have proved that perturbation of the noise-space correlation by external noise is roughly proportional to the spectral norm of the noise Hankel matrix with a magnification factor given in terms of maximum and minimum nonzero singular values of the Hankel matrix constructed from the noiseless measurements. Under the assumption of frequency separation roughly ≥ 2 RL, the magnification factor is explicitly estimated by means of a new version of discrete Ingham inequalities.

A systematic numerical study has shown that the MUSIC algorithm enjoys strong stability and low computation complexity for the reconstruction of well-separated frequencies. MUSIC is the only algorithm that can recover arbitrarily closely spaced frequencies as long as the noise is sufficiently small. And we have numerically documented the super-resolution effect of MUSIC in terms of the relationship among the minimum separation, the Rayleigh index (the size of largest cluster) and the noise. The results conform to the optimal bound conjectured (and partially proved) by Donoho [15].

Finally we discuss a possible extension of the present work. We became aware of the reference [7] after completing the first draft of this work (`arXiv:1404.1484`). In [7], Chen and Chi used the matrix completion technique to obtain a stable approximation of $\{y(k), k = 0, \ldots, M\}$ from its partial noisy samples. This can be used as the preprocessing denoising step before invoking the single-snapshot MUSIC. Together [7] and the present work constitute a framework for single-snapshot spectral estimation with compressive noisy measurements.

Let y and y^ε, respectively, be the full set of noiseless and noisy data as before. Let the sampling set Λ be a random subset of size m from $\{0, \ldots, M\}$. Let $\mathcal{P}_\Lambda(v)$ be the orthogonal projection of $v \in \mathbb{C}^{M+1}$ onto the space of vectors supported on Λ.

For the noisy compressive data $\mathcal{P}_\Lambda y^\varepsilon$ satisfying $\|\mathcal{P}_\Lambda(y^\varepsilon - y)\|_2 \leq \delta$, [7] proposes the following denoising strategy of Hankel matrix completion

$$\hat{y} = \arg\min_z \|\text{Hankel}(z)\|_\star, \text{ s.t.} \quad \|\mathcal{P}_\Lambda(z - y^\varepsilon)\|_2 \leq \delta$$

where $\|\cdot\|_*$ denotes the nuclear norm.

The total procedure of compressive spectral estimation is given in the following table.

Spectral estimation with compressive measurements
Input: $\mathcal{P}_\Lambda y^\varepsilon \in \mathbb{C}^{M+1}, \delta, s, L$.
1) Matrix completion: $\hat{y} = \arg\min_{z \in \mathbb{C}^{M\times 1}} \|\text{Hankel}(z)\|_\star$, subject to $\|\mathcal{P}_\Lambda(z - y^\varepsilon)\|_2 \leq \delta$
2) Form Hankel matrix $\hat{H} = \text{Hankel}(\hat{y}) \in \mathbb{C}^{(L+1)\times(M-L+1)}$.
3) SVD: $\hat{H} = [\hat{U}_1\ \hat{U}_2]\text{diag}(\hat{\sigma}_1, \ldots, \hat{\sigma}_s, \ldots)[\hat{V}_1\ \hat{V}_2]^\star$, where $\hat{U}_1 \in \mathbb{C}^{(L+1)\times s}$.
4) Compute imaging function $\hat{J}(\omega) = \|\phi^L(\omega)\|_2 / \|\hat{U}_2^\star \phi^L(\omega)\|_2$.
Output: $\hat{\mathcal{S}} = \{\omega$ corresponding to s largest local maxima of $\hat{J}(\omega)\}$.

The following estimate on the difference $\hat{R}(\omega) - R(\omega)$ is obtained by combining [7, Theorem 2] and Theorem 3.

Theorem 6. *Let $R(\omega)$ and $\hat{R}(\omega)$, respectively, be the noise-space correlation functions for the noiseless y and denoised data $\hat{y}$.*

Let Λ of size m be uniformly sampled at random from $\{0,\dots,M\}$. Suppose $\|\mathcal{P}_\Lambda(y^\varepsilon - y)\|_2 \le \delta$. Then there exists a universal constant $C>0$ such that

$$\|\mathrm{Hankel}(\hat{y}) - \mathrm{Hankel}(y)\|_F \le \left(2\sqrt{M+1} + 8(M+1) + \frac{8\sqrt{2}(M+1)^2}{m}\right)\delta \tag{42}$$

and

$$|\hat{R}(\omega) - R(\omega)| \le \frac{4\sigma_1 + \left(2\sqrt{M+1} + 8(M+1) + \frac{8\sqrt{2}(M+1)^2}{m}\right)\delta}{\left[\sigma_s - \left(2\sqrt{M+1} + 8(M+1) + \frac{8\sqrt{2}(M+1)^2}{m}\right)\delta\right]^2}\left(2\sqrt{M+1} + 8(M+1) + \frac{8\sqrt{2}(M+1)^2}{m}\right)\delta. \tag{43}$$

with probability exceeding $1-(M+1)^{-2}$ provided that

$$m > C\mu\gamma s \log^3(M+1) \tag{44}$$

where

$$\mu = \max\left(\frac{L+1}{\sigma^2_{\min}(\Phi^L)}, \frac{M-L+1}{\sigma^2_{\min}(\Phi^{M-L})}\right), \quad \gamma = \max\left(\frac{M+1}{L+1}, \frac{M+1}{M-L+1}\right).$$

Theorem 6 implies that compressive spectral estimation is stable with matrix completion and MUSIC whenever Φ^L, Φ^{M-L} are well-conditioned and the sample size is sufficiently large.

Furthermore with the discrete Ingham inequalities we can give explicit estimate for the right hand side of (43) and (44). In particular, with $L \approx M/2$ and well-separated (>2 RL) frequencies, μ and γ scale like a constant and $m = \mathcal{O}(s\log^3 M)$ suffices for any sufficiently small δ. In this case, the right hand side of (43) is

$$\mathcal{O}\left(\frac{M}{m}\frac{x_{\max}}{x_{\min}}\frac{\delta}{x_{\min}}\right)$$

showing enhanced stability as $M/m \to 1$ where M/m is the compression ratio, $x_{\max}/x_{\min}$ the object's peak-to-trough ratio and $\delta/x_{\min}$ the noise-to-object ratio.

Acknowledgement

Wenjing Liao would like to thank Armin Eftekhari for providing their codes and Gabriel Peyré for helpful comment at 2014 SIAM conference on imaging science.

Appendix A Proof of Theorem 1

In the noise-free case Range(H) and Range(Φ^L) coincide if the matrix $X(\Phi^{M-L})^T$ has full row rank, i.e., Rank $(\Phi^{M-L}) = s$, which is guaranteed on the condition that $M-L+1 \ge s$ and the frequencies in $\mathcal{S}$ are pairwise distinct.

Lemma 1. *If Rank$(\Phi^{M-L}) = s$, then* Range(H) = Range(Φ^L).

Lemma 2. *Rank$(\Phi^L) = s$ if $L+1 \ge s$ and $\omega_k \ne \omega_l,\ \forall k \ne l$.*

Proof. If $L+1 \geq s$, then $s \times s$ square submatrix Ψ of Φ^L is a square Vandermonde matrix whose determinant is given by

$$\det(\Psi) = \prod_{1 \leq i < j \leq s} (e^{-i2\pi\omega_j} - e^{-i2\pi\omega_i}).$$

Clearly, $\det(\Psi) \neq 0$ if and only if $\omega_i \neq \omega_j, i \neq j$. Hence $\text{Rank}(\Psi) = s$ which implies $\text{Rank}(\Phi^L) = s$. □

Similarly, if $L+1 \geq s+1$, the extended matrix $\Phi^L_\omega = [\Phi^L \ \phi^L(\omega)]$ has full column rank for any $\omega \notin \mathcal{S}$. As a consequence, $\omega \in \mathcal{S}$ if and only if $\phi^L(\omega)$ belongs to Range(Φ^L).

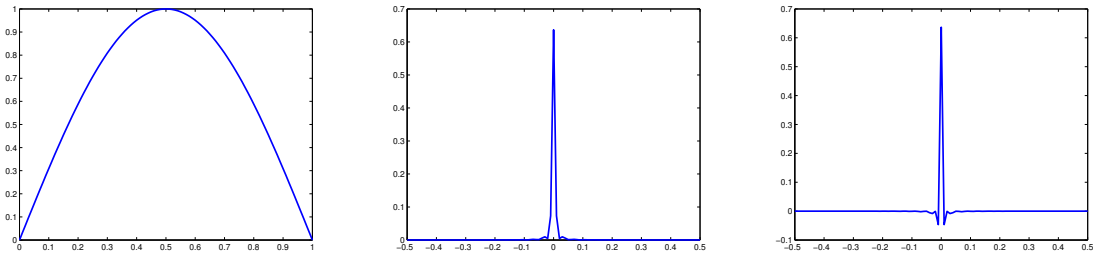
Appendix B Proof of Theorem 2

The proof of Theorem 2 combines techniques used in [28] and [36]. We take

$$g(t) = \cos \pi(t - 0.5)$$

and let

$$G(\omega) = \sum_{k=0}^{L} g(\frac{k}{L}) e^{2\pi i k\omega}. \tag{45}$$



(a) $g(t)$ (b) $|G(\omega)|/L$ (c) Real part of $G(\omega)/L$

Figure 8: Function $g(t)$ and $G(\omega)/L$ when $L = 128$.

Graphs of $g(t)$, $|G(\omega)|/L$ and the real part of $G(\omega)/L$ are shown in Figure 8. Function G has the following properties.

Lemma 3. *1.* $G(\omega + n) = G(\omega)$ *for* $n \in \mathbb{Z}$.

2. $G(-\omega) = e^{-2\pi i L\omega} G(\omega)$ *and* $|G(-\omega)| = |G(\omega)|$.

3. $L(\frac{2}{\pi} - \frac{1}{L}) \leq G(0) \leq L(\frac{2}{\pi} + \frac{1}{L})$.

4. $|G(\omega)| \leq \frac{2}{\pi} \frac{L}{|1-4L^2\omega^2|} + \frac{8}{\pi L}$ *for* $\omega \in [0, 1/2]$.

Proof. 1.

$$G(\omega+n)=\sum_{k=0}^{L}g(\frac{k}{L})e^{2\pi ik(\omega+n)}=\sum_{k=0}^{L}g(\frac{k}{L})e^{2\pi ik\omega}=G(\omega).$$

2.

$$\begin{aligned}G(-\omega)&=\sum_{k=0}^{L}g(\frac{k}{L})e^{-2\pi ik\omega}=\sum_{l=0}^{L}g\left(\frac{L-l}{L}\right)e^{-2\pi i(L-l)\omega}\text{ by letting }l=L-k\\&=\sum_{l=0}^{L}\left[\cos\pi(1-\frac{l}{L}-0.5)\right]e^{-2\pi i(L-l)\omega}=\sum_{l=0}^{L}\left[\cos\pi(\frac{l}{L}-0.5)\right]e^{-2\pi i(L-l)\omega}\\&=\sum_{l=0}^{L}g(\frac{l}{L})e^{-2\pi i(L-l)\omega}=e^{-2\pi iL\omega}\sum_{l=0}^{L}g(\frac{l}{L})e^{2\pi il\omega}=e^{-2\pi iL\omega}G(\omega).\end{aligned}$$

3. On the one hand,

$$\|g\|_1=\int_0^1\cos\pi(t-0.5)dt=\frac{2}{\pi}.$$

On the other hand,

$$G(0)=\sum_0^L g(\frac{k}{L}),$$

and

$$\frac{1}{L}(G(0)-1)\le\|g\|_1\le\frac{1}{L}(G(0)+1).$$

4. According to the Poisson summation formula,

$$\frac{1}{L}G(\omega)=\frac{1}{L}\sum_{k=0}^{L}g(\frac{k}{L})e^{2\pi ik\omega}=\sum_{r=-\infty}^{\infty}\int_0^1 g(z)e^{2\pi iL(\omega-r)z}dz=\frac{2}{\pi}\sum_{r=-\infty}^{\infty}\frac{\cos\pi L(\omega-r)}{1-4L^2(\omega-r)^2}e^{i\pi L(\omega-r)}.\tag{46}$$

Hence

$$\begin{aligned}
&\Big|\frac{1}{L}G(\omega)\Big| \le \frac{2}{\pi}\sum_{r=-\infty}^{\infty}\Big|\frac{\cos\pi L(\omega-r)}{1-4L^2(\omega-r)^2}\Big| \\
&\le \frac{2}{\pi}\frac{1}{|1-4L^2\omega^2|}+\frac{2}{\pi}\sum_{r\neq 0}\frac{1}{4L^2(r-\omega)^2-1} \\
&\le \frac{2}{\pi}\frac{1}{|1-4L^2\omega^2|}+\frac{2}{\pi}\sum_{r\neq 0}\frac{2}{4L^2(r-\omega)^2} \\
&\le \frac{2}{\pi}\frac{1}{|1-4L^2\omega^2|}+\frac{4}{\pi}\Big[\frac{1}{4L^2(\frac{1}{2})^2}+\frac{1}{4L^2(1)^2}+\frac{1}{4L^2(\frac{3}{2})^2}+\frac{1}{4L^2(2)^2}+\dots\Big] \\
&\qquad\qquad \text{as } \omega\in[0,\frac{1}{2}], \\
&\le \frac{2}{\pi}\frac{1}{|1-4L^2\omega^2|}+\frac{4}{\pi}\frac{1}{L^2}\Big[\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\dots\Big] \\
&\le \frac{2}{\pi}\frac{1}{|1-4L^2\omega^2|}+\frac{4}{\pi}\frac{1}{L^2}2=\frac{2}{\pi}\frac{1}{|1-4L^2\omega^2|}+\frac{8}{\pi L^2}.
\end{aligned}$$

In (46) the difference between the discrete and the continuous case lies in

$$\frac{2}{\pi}\sum_{r\neq 0}\frac{\cos\pi L(\omega-r)}{1-4L^2(\omega-r)^2}$$

which is bounded above by $8/(\pi L^2)$, and is therefore negligible when L is sufficiently large. □

We start with the following lemma, which paves the way for the proof of Theorem 2.

Lemma 4. *Suppose objects in $\mathcal{S}$ satisfy the gap condition*

$$d(\omega_j,\omega_l)\ge q>\frac{1}{L}\sqrt{\frac{2}{\pi}}\left(\frac{2}{\pi}-\frac{1}{L}-\frac{8s}{\pi L^2}\right)^{-\frac{1}{2}}. \tag{47}$$

Then

$$\Big(\frac{2}{\pi}-\frac{1}{L}-\frac{2}{\pi L^2q^2}-\frac{8s}{\pi L^2}\Big)\|\mathrm{c}\|_2^2\le\frac{1}{L}\sum_{k=0}^{L}g(\frac{k}{L})\Big|(\Phi^{0\to L}\mathrm{c})_k\Big|^2\le\Big(\frac{2}{\pi}+\frac{1}{L}+\frac{2}{\pi L^2q^2}+\frac{8s}{\pi L^2}\Big)\|\mathrm{c}\|_2^2 \tag{48}$$

for all $\mathrm{c}\in\mathbb{C}^s$.

Proof.

$$\sum_{k=0}^{L} g(\frac{k}{L})\Big|(\Phi^{0\to L}\mathrm{c})_k\Big|^2 = \sum_{k=0}^{L} g(\frac{k}{L})\overline{\sum_{j=1}^{s}\mathrm{c}_j e^{-2\pi ik\omega_j}}\sum_{l=1}^{s}\mathrm{c}_l e^{-2\pi ik\omega_l}$$
$$= \sum_{j=1}^{s}\sum_{l=1}^{s}\overline{\mathrm{c}_j}\mathrm{c}_l\sum_{k=0}^{L} g(\frac{k}{L})e^{2\pi ik(\omega_j-\omega_l)} = \sum_{j=1}^{s}\sum_{l=1}^{s} G(\omega_j-\omega_l)\overline{\mathrm{c}_j}\mathrm{c}_l$$
$$= G(0)\|\mathrm{c}\|_2^2 + \sum_{j=1}^{s}\sum_{l\neq j} G(\omega_j-\omega_l)\overline{\mathrm{c}_j}\mathrm{c}_l.$$

It follows from the triangle inequality that

$$G(0)\|\mathrm{c}\|_2^2 - \sum_{j=1}^{s}\sum_{l\neq j}|G(\omega_j-\omega_l)\overline{\mathrm{c}_j}\mathrm{c}_l| \le \sum_{k=0}^{L} g(\frac{k}{L})\Big|(\Phi^{0\to L}\mathrm{c})_k\Big|^2 \le G(0)\|\mathrm{c}\|_2^2 + \sum_{j=1}^{s}\sum_{l\neq j}|G(\omega_j-\omega_l)\overline{\mathrm{c}_j}\mathrm{c}_l|,$$

where $\sum_{j=1}^{s}\sum_{l\neq j}|G(\omega_j-\omega_l)\overline{\mathrm{c}_j}\mathrm{c}_l|$ can be estimated through Property 4 in Lemma 3.

$$\sum_{j=1}^{s}\sum_{l\neq j}|G(\omega_j-\omega_l)\overline{\mathrm{c}_j}\mathrm{c}_l| \le \sum_{j=1}^{s}\sum_{l\neq j}|G(\omega_j-\omega_l)|\frac{|\mathrm{c}_j|^2+|\mathrm{c}_l|^2}{2} = \sum_{j=1}^{s}|\mathrm{c}_j|^2\sum_{l\neq j}|G(\omega_j-\omega_l)|$$
$$= \sum_{j=1}^{s}|\mathrm{c}_j|^2\sum_{l\neq j}|G(d(\omega_j,\omega_l))| \le \sum_{j=1}^{s}|\mathrm{c}_j|^2\sum_{l\neq j}\Big[\frac{2}{\pi}\frac{L}{|1-4L^2d^2(\omega_j,\omega_l)|}+\frac{8}{\pi L}\Big]$$
$$\le \sum_{j=1}^{s}|\mathrm{c}_j|^2\frac{4}{\pi}\sum_{n=1}^{\lfloor\frac{s}{2}\rfloor}\Big[\frac{L}{4L^2n^2q^2-1}+\frac{4}{L}\Big] \text{ as frequencies in } \mathcal{S} \text{ are pairwise separated by } q>\frac{1}{L}$$
$$\le \sum_{j=1}^{s}|\mathrm{c}_j|^2\frac{4}{\pi}\Big[\frac{2s}{L}+\sum_{n=1}^{\infty}\frac{L}{4L^2n^2q^2-1}\Big] \le \sum_{j=1}^{s}|\mathrm{c}_j|^2\frac{4}{\pi}\Big[\frac{2s}{L}+\frac{1}{L^2q^2}\sum_{n=1}^{\infty}\frac{L}{4n^2-1}\Big] \tag{49}$$
$$\le \sum_{j=1}^{s}|\mathrm{c}_j|^2\frac{4}{\pi}\Big[\frac{2s}{L}+\frac{1}{Lq^2}\frac{1}{2}\sum_{n=1}^{\infty}\Big(\frac{1}{2n-1}-\frac{1}{2n+1}\Big)\Big] = \sum_{j=1}^{s}|\mathrm{c}_j|^2\frac{4}{\pi}\Big[\frac{2s}{L}+\frac{1}{Lq^2}\frac{1}{2}\Big]$$
$$= \|\mathrm{c}\|_2^2\frac{2}{\pi}\Big(\frac{1}{Lq^2}+\frac{4s}{L}\Big),$$

where $\lfloor\frac{s}{2}\rfloor$ denotes the nearest integer smaller than or equal to $s/2$. Kernel g in (48) is crucial for the convergence of the series in (49).

Therefore

$$G(0)\|\mathrm{c}\|_2^2 - \frac{2}{\pi}\Big(\frac{1}{Lq^2}+\frac{4s}{L}\Big)\|\mathrm{c}\|_2^2 \le \sum_{k=0}^{L} g(\frac{k}{L})\Big|(\Phi^{0\to L}\mathrm{c})_k\Big|^2 \le G(0)\|\mathrm{c}\|_2^2 + \frac{2}{\pi}\Big(\frac{1}{Lq^2}+\frac{4s}{L}\Big)\|\mathrm{c}\|_2^2.$$

The equation above along with Property 3 in Lemma 3 yields

$$L\Big(\frac{2}{\pi}-\frac{1}{L}-\frac{2}{\pi L^2q^2}-\frac{8s}{\pi L^2}\Big)\|\mathrm{c}\|_2^2 \le \sum_{k=0}^{L} g(\frac{k}{L})\Big|(\Phi^{0\to L}\mathrm{c})_k\Big|^2 \le L\Big(\frac{2}{\pi}+\frac{1}{L}+\frac{2}{\pi L^2q^2}+\frac{8s}{\pi L^2}\Big)\|\mathrm{c}\|_2^2.$$

The gap condition (47) is derived from the positivity condition of the lower bound, i.e.,

$$\frac{2}{\pi} - \frac{1}{L} - \frac{2}{\pi L^2 q^2} - \frac{8s}{\pi L^2} > 0.$$

□

Proof of Theorem 2 is given below.

Proof. Given that $\mathcal{S} = \{\omega_1, \ldots, \omega_s\} \subset [0, 1)$ and frequencies are separated above $1/L$, there are no more than L frequencies in $\mathcal{S}$, i.e., $s < L$.

The lower bound in Theorem 4 follows from Lemma 4 as

$$\begin{aligned}\|\Phi^{0\to L}\mathrm{c}\|_2^2 &= \sum_{k=0}^{L}\Big|(\Phi^{0\to L}\mathrm{c})_k\Big|^2 = \sum_{k=0}^{L}\Big|(\Phi^{0\to L}\mathrm{c})_k\Big|^2 \geq \sum_{k=0}^{L} g(\frac{k}{L})\Big|(\Phi^{0\to L}\mathrm{c})_k\Big|^2 \\ &\geq L\Big(\frac{2}{\pi} - \frac{1}{L} - \frac{2}{\pi L^2 q^2} - \frac{8s}{\pi L^2}\Big) > L\Big(\frac{2}{\pi} - \frac{2}{\pi L^2 q^2} - \frac{4}{L}\Big).\end{aligned}$$

The gap condition (16) in Theorem 4 is derived from the positivity condition of the lower bound, i.e.,

$$\Big(\frac{2}{\pi} - \frac{2}{\pi L^2 q^2} - \frac{4}{L}\Big) > 0.$$

We prove the upper bound in Theorem 4 in two cases: L is even or L is odd.

Case 1: L is even. First we substitute L with $2L$ in (48) and obtain

$$\sum_{k=0}^{2L} g(\frac{k}{2L})\Big|(\Phi^{0\to 2L}\mathrm{c})_k\Big|^2 \leq 2L\Big(\frac{2}{\pi} + \frac{1}{2L} + \frac{2}{4\pi L^2 q^2} + \frac{8s}{4\pi L^2}\Big)\|\mathrm{c}\|_2^2. \tag{50}$$

Let $D^{\frac{L}{2}} = \mathrm{diag}(e^{-2\pi i\omega_1\frac{L}{2}}, e^{-2\pi i\omega_2\frac{L}{2}}, \ldots, e^{-2\pi i\omega_s\frac{L}{2}})$ and $D^{-\frac{L}{2}} = (D^{\frac{L}{2}})^{-1}$. On the one hand,

$$\begin{aligned}&\sum_{k=0}^{2L} g(\frac{k}{2L})\Big|(\Phi^{0\to 2L}D^{-\frac{L}{2}}\mathrm{c})_k\Big|^2 \geq \sum_{k=L/2}^{3L/2} g(\frac{k}{2L})\Big|(\Phi^{0\to 2L}D^{-\frac{L}{2}}\mathrm{c})_k\Big|^2 \geq \sum_{k=L/2}^{3L/2} g(\frac{1}{4})\Big|(\Phi^{0\to 2L}D^{-\frac{L}{2}}\mathrm{c})_k\Big|^2 \\ &= \frac{1}{\sqrt{2}}\sum_{k=L/2}^{3L/2}\Big|(\Phi^{0\to 2L}D^{-\frac{L}{2}}\mathrm{c})_k\Big|^2 = \frac{1}{\sqrt{2}}\sum_{k=0}^{L}\Big|(\Phi^{\frac{L}{2}\to\frac{3L}{2}}D^{-\frac{L}{2}}\mathrm{c})_k\Big|^2 = \frac{1}{\sqrt{2}}\sum_{k=0}^{L}\Big|(\Phi^{0\to L}D^{\frac{L}{2}}D^{-\frac{L}{2}}\mathrm{c})_k\Big|^2 \\ &= \frac{1}{\sqrt{2}}\sum_{k=0}^{L}\Big|(\Phi^{0\to L}\mathrm{c})_k\Big|^2.\end{aligned}$$

On the other hand, (50) implies

$$\begin{aligned}\sum_{k=0}^{2L} g(\frac{k}{2L})\Big|(\Phi^{0\to 2L}D^{-\frac{L}{2}}\mathrm{c})_k\Big|^2 &\leq 2L\Big(\frac{2}{\pi} + \frac{1}{2L} + \frac{2}{4\pi L^2 q^2} + \frac{8s}{4\pi L^2}\Big)\|D^{-\frac{L}{2}}\mathrm{c}\|_2^2 \\ &= L\Big(\frac{4}{\pi} + \frac{1}{L} + \frac{1}{\pi L^2 q^2} + \frac{4s}{\pi L^2}\Big)\|\mathrm{c}\|_2^2.\end{aligned}$$

As a result

$$\|\Phi^L \mathrm{c}\|_2^2 = \sum_{k=0}^{L} \left|(\Phi^{0\to L}\mathrm{c})_k\right|^2 \le L\Big(\frac{4\sqrt{2}}{\pi}+\frac{\sqrt{2}}{L}+\frac{\sqrt{2}}{\pi L^2 q^2}+\frac{4\sqrt{2}s}{\pi L^2}\Big)\|\mathrm{c}\|_2^2 < L\Big(\frac{4\sqrt{2}}{\pi}+\frac{\sqrt{2}}{\pi L^2 q^2}+\frac{3\sqrt{2}}{L}\Big)\|\mathrm{c}\|_2^2.$$

When L is an even integer,

$$\Big(\frac{2}{\pi}-\frac{2}{\pi L^2 q^2}-\frac{4}{L}\Big)\|\mathrm{c}\|_2^2 \le \frac{1}{L}\|\Phi^L \mathrm{c}\|_2^2 \le \Big(\frac{4\sqrt{2}}{\pi}+\frac{\sqrt{2}}{\pi L^2 q^2}+\frac{3\sqrt{2}}{L}\Big)\|\mathrm{c}\|_2^2.$$

Case 2: L is odd.

$$\begin{aligned}\|\Phi^L \mathrm{c}\|_2^2 &= \sum_{k=0}^{L} \left|(\Phi^{0\to L}\mathrm{c})_k\right|^2 \le \sum_{k=0}^{L+1} \left|(\Phi^{0\to L+1}\mathrm{c})_k\right|^2 \\ &< (L+1)\Big(\frac{4\sqrt{2}}{\pi}+\frac{\sqrt{2}}{\pi(L+1)^2 q^2}+\frac{3\sqrt{2}}{L+1}\Big)\|\mathrm{c}\|_2^2. \end{aligned} \tag{51}$$

Eq. (51) above follows from Case 1 as $L+1$ is an even integer. In summary, when L is an odd integer,

$$\left(\frac{2}{\pi}-\frac{2}{\pi L^2 q^2}-\frac{4}{L}\right)\|\mathrm{c}\|_2^2 \le \frac{1}{L}\|\Phi^L \mathrm{c}\|_2^2 \le \left(1+\frac{1}{L}\right)\left(\frac{4\sqrt{2}}{\pi}+\frac{\sqrt{2}}{\pi(L+1)^2 q^2}+\frac{3\sqrt{2}}{L+1}\right)\|\mathrm{c}\|_2^2.$$

□

Appendix C Proof of Theorems in Section 3

C.1 Proof of Theorem 3

Proof. Let $\mathcal{H} = HH^\star$, $\mathcal{H}^\varepsilon = H^\varepsilon H^{\varepsilon\star}$ and $\mathcal{E} = HE^\star + EH^\star + EE^\star$. Then

$$\mathcal{H}^\varepsilon = \mathcal{H} + \mathcal{E}, \tag{52}$$

and

$$\mathcal{H} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1\Sigma_1^\star & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^\star \\ U_2^\star \end{bmatrix}, \tag{53}$$

$$\mathcal{H}^\varepsilon = \begin{bmatrix} U_1^\varepsilon & U_2^\varepsilon \end{bmatrix} \begin{bmatrix} \Sigma_1^\varepsilon\Sigma_1^{\varepsilon\star} & 0 \\ 0 & \Sigma_2^\varepsilon\Sigma_2^{\varepsilon\star} \end{bmatrix} \begin{bmatrix} U^{\varepsilon\star}_1 \\ U^{\varepsilon\star}_2 \end{bmatrix}, \tag{54}$$

where $\Sigma_1 = \mathrm{diag}(\sigma_1,\ldots,\sigma_s), \Sigma_1^\varepsilon = \mathrm{diag}(\sigma_1^\varepsilon,\ldots,\sigma_s^\varepsilon)$, and $\Sigma_2^\varepsilon = \mathrm{diag}(\sigma_{s+1}^\varepsilon, \sigma_{s+2}^\varepsilon,\ldots)$.

Combining (52) and (54) yields

$$\begin{bmatrix} U_1^\star \\ U_2^\star \end{bmatrix} (\mathcal{H}+\mathcal{E}) \begin{bmatrix} U_1^\varepsilon & U_2^\varepsilon \end{bmatrix} = \begin{bmatrix} U_1^\star U_1^\varepsilon \Sigma_1^\varepsilon\Sigma_1^{\varepsilon\star} & U_1^\star U_2^\varepsilon\Sigma_2^\varepsilon\Sigma_2^{\varepsilon\star} \\ U_2^\star U_1^\varepsilon\Sigma_1^\varepsilon\Sigma_1^{\varepsilon\star} & U_2^\star U_2^\varepsilon\Sigma_2^\varepsilon\Sigma_2^{\varepsilon\star} \end{bmatrix}. \tag{55}$$

On the one hand, the (2,1) entries on both sides of (55) are equal such that

$$U_2^\star \mathcal{H} U_1^\varepsilon + U_2^\star \mathcal{E} U_1^\varepsilon = U_2^\star U_1^\varepsilon \Sigma_1^\varepsilon \Sigma_1^{\varepsilon\star}.$$

Based on (53), we obtain $U_2^\star \mathcal{H} = \mathbf{0}$ and therefore

$$U_2^\star \mathcal{E} U_1^\varepsilon (\Sigma_1^\varepsilon \Sigma_1^{\varepsilon\star})^{-1} = U_2^\star U_1^\varepsilon$$

which implies

$$\|U_2^\star U_1^\varepsilon\|_2 \leq \frac{\|\mathcal{E}\|_2}{(\sigma_s^\varepsilon)^2} \tag{56}$$

On the other hand, the (1,2) entries on both sides of (55) are equal such that

$$U_1^\star \mathcal{H} U_2^\varepsilon + U_1^\star \mathcal{E} U_2^\varepsilon = U_1^\star U_2^\varepsilon \Sigma_2^\varepsilon \Sigma_2^{\varepsilon\star}.$$

Based on (53), we obtain $U_1^\star \mathcal{H} = \Sigma_1 \Sigma_1^\star U_1^\star$ and therefore

$$\Sigma_1 \Sigma_1^\star U_1^\star U_2^\varepsilon + U_1^\star \mathcal{E} U_2^\varepsilon = U_1^\star U_2^\varepsilon \Sigma_2^\varepsilon \Sigma_2^{\varepsilon\star}.$$

For any $\phi \in \mathbb{C}^{L+1-s}$,

$$\begin{aligned}\sigma_s^2 \|U_1^\star U_2^\varepsilon \phi\|_2 &\leq \|\Sigma_1 \Sigma_1^\star U_1^\star U_2^\varepsilon \phi\|_2 \leq \|U_1^\star \mathcal{E} U_2^\varepsilon\|_2 \|\phi\|_2 + \|U_1^\star U_2^\varepsilon \Sigma_2^\varepsilon \Sigma_2^{\varepsilon\star} \phi\|_2 \\ &\leq \|\mathcal{E}\|_2 \|\phi\|_2 + \|U_1^\star U_2^\varepsilon\|_2 (\sigma_{s+1}^\varepsilon)^2 \|\phi\|_2,\end{aligned}$$

so

$$\frac{\|U_1^\star U_2^\varepsilon \phi\|_2}{\|\phi\|_2} \leq \frac{\|\mathcal{E}\|_2 + (\sigma_{s+1}^\varepsilon)^2 \|U_1^\star U_2^\varepsilon\|_2}{\sigma_s^2}.$$

By taking the supremum over $\phi \in \mathbb{C}^{L+1-s}$ on the left, we obtain

$$\|U_1^\star U_2^\varepsilon\|_2 \leq \frac{\|\mathcal{E}\|_2 + (\sigma_{s+1}^\varepsilon)^2 \|U_1^\star U_2^\varepsilon\|_2}{\sigma_s^2},$$

and then

$$\|U_1^\star U_2^\varepsilon\|_2 \leq \frac{\|\mathcal{E}\|_2}{\sigma_s^2 - (\sigma_{s+1}^\varepsilon)^2}. \tag{57}$$

Let $\mathcal{P}_1$ and $\mathcal{P}_1^\varepsilon$ be orthogonal projects onto the subspace spanned by columns of U_1 and U_1^ε respectively. For any $\phi \in \mathbb{C}^{L+1}$

$$\begin{aligned}&\frac{\|\mathcal{P}_2^\varepsilon \phi - \mathcal{P}_2 \phi\|_2}{\|\phi\|_2} = \frac{\|\mathcal{P}_1 \mathcal{P}_2^\varepsilon \phi + \mathcal{P}_2 \mathcal{P}_2^\varepsilon \phi - \mathcal{P}_2 \phi\|_2}{\|\phi\|_2} = \frac{\|\mathcal{P}_1 \mathcal{P}_2^\varepsilon \phi - \mathcal{P}_2 \mathcal{P}_1^\varepsilon \phi\|_2}{\|\phi\|_2} \\ &= \frac{\|U_1 U_1^\star U_2^\varepsilon U_2^{\varepsilon\star} \phi - U_2 U_2^\star U_1^\varepsilon U_1^{\varepsilon\star} \phi\|_2}{\|\phi\|_2} \leq \|U_1^\star U_2^\varepsilon\|_2 + \|U_2^\star U_1^\varepsilon\|_2.\end{aligned} \tag{58}$$

Eq. (58) together with (56) and (57) imply

$$|R^\varepsilon(\omega) - R(\omega)| \leq \|\mathcal{P}_2^\varepsilon - \mathcal{P}_2\|_2 = \sup_{\phi \in \mathbb{C}^{L+1}} \frac{\|\mathcal{P}_2^\varepsilon \phi - \mathcal{P}_2 \phi\|_2}{\|\phi\|_2} \leq \left[\frac{1}{(\sigma_s^\varepsilon)^2} + \frac{1}{\sigma_s^2 - (\sigma_{s+1}^\varepsilon)^2}\right] \|\mathcal{E}\|_2.$$

Meanwhile

$$\|\mathcal{E}\|_2 \le 2\|H\|_2\|E\|_2 + \|E\|_2^2 \le (2\sigma_1 + \|E\|_2)\|E\|_2,$$

and therefore

$$\|\mathcal{P}_2^\varepsilon - \mathcal{P}_2\|_2 \le (2\sigma_1 + \|E\|_2)\left[\frac{1}{(\sigma_s^\varepsilon)^2} + \frac{1}{\sigma_s^2 - (\sigma_{s+1}^\varepsilon)^2}\right]\|E\|_2.$$

According to Proposition 1, $\sigma_s^\varepsilon \ge \sigma_s - \|E\|_2$ and $\sigma_{s+1}^\varepsilon \le \|E\|_2$. Then

$$\frac{1}{(\sigma_s^\varepsilon)^2} + \frac{1}{\sigma_s^2 - (\sigma_{s+1}^\varepsilon)^2} \le \frac{2}{(\sigma_s - \|E\|_2)^2},$$

which implies that

$$\|\mathcal{P}_2^\varepsilon - \mathcal{P}_2\|_2 \le \frac{2(2\sigma_1 + \|E\|_2)}{(\sigma_s - \|E\|_2)^2}\|E\|_2.$$

In particular, while ω is restricted on $\mathcal{S}$, a sharper upper bound in (22) is derived as follows:

Since $M - L + 1 \ge s$ and true frequencies are pairwise distinct, $X(\Phi^{M-L})^T$ has full row rank. Denote $Y^\varepsilon = U_1^\varepsilon \Sigma_1^\varepsilon V_1^{\varepsilon\star} + U_2^\varepsilon \Sigma_2^\varepsilon V_2^{\varepsilon\star}$ where $\Sigma_1^\varepsilon = \mathrm{diag}(\sigma_1^\varepsilon, \ldots, \sigma_s^\varepsilon)$ and $\Sigma_2^\varepsilon = \mathrm{diag}(\sigma_{s+1}^\varepsilon, \sigma_{s+2}^\varepsilon, \ldots)$. Multiplying $U_2^{\varepsilon\star}$ on the left and the pseudo-inverse of $X(\Phi^{M-L})^T$ on the right of

$$U_1^\varepsilon \Sigma_1^\varepsilon V_1^{\varepsilon\star} + U_2^\varepsilon \Sigma_2^\varepsilon V_2^{\varepsilon\star} = \Phi^L X(\Phi^{M-L})^T + E$$

yields

$$\Sigma_2^\varepsilon V_2^{\varepsilon\star}[X(\Phi^{M-L})^T]^\dagger = U_2^{\varepsilon\star}\Phi^L + U_2^{\varepsilon\star}E[X(\Phi^{M-L})^T]^\dagger.$$

Then

$$U_2^{\varepsilon\star}\phi^L(\omega_j) = U_2^{\varepsilon\star}\Phi^L e_j = U_2^{\varepsilon\star}E[X(\Phi^{M-L})^T]^\dagger e_j - \Sigma_2^\varepsilon V_2^{\varepsilon\star}[X(\Phi^{M-L})^T]^\dagger e_j,$$

and

$$\|\mathcal{P}_2^\varepsilon \phi^L(\omega_j)\|_2 = \|U_2^{\varepsilon\star}\phi^L(\omega_j)\|_2 \le \frac{\|E\|_2 + \sigma_{s+1}^\varepsilon}{\sigma_{\min}(X(\Phi^{M-L})^T)} \le \frac{2\|E\|_2}{\sigma_{\min}(X(\Phi^{M-L})^T)} \le \frac{2\|E\|_2}{x_{\min}\sigma_{\min}((\Phi^{M-L})^T)}.$$

Therefore

$$R^\varepsilon(\omega_j) = \frac{\|\mathcal{P}_2^\varepsilon \phi^L(\omega_j)\|_2}{\|\phi^L(\omega_j)\|_2} \le \frac{2\|E\|_2}{x_{\min}\sigma_{\min}((\Phi^{M-L})^T)\|\phi^L(\omega_j)\|_2}.$$

□

C.2 Proof of Theorem 4

Proof. Let

$$Q(\omega) = R^2(\omega) = \frac{\phi^L(\omega)^\star U_2 U_2^\star U_2 U_2^\star \phi^L(\omega)}{\|\phi^L(\omega)\|_2^2} = \frac{\phi^L(\omega)^\star U_2 U_2^\star U_2 U_2^\star \phi^L(\omega)}{L+1}$$

and

$$Q^\varepsilon(\omega) = [R^\varepsilon(\omega)]^2 = \frac{\phi^L(\omega)^\star U_2^\varepsilon U^{\varepsilon\star}{}_2 U_2^\varepsilon U^{\varepsilon\star}{}_2 \phi^L(\omega)}{L+1}.$$

Both $Q(\omega)$ and $Q^\varepsilon(\omega)$ are smooth functions and

$$\begin{aligned} Q'(\omega) &= \frac{\phi^L(\omega)^\star U_2U_2^\star U_2U_2^\star[\phi^L(\omega)]' + [\phi^L(\omega)]'^\star U_2U_2^\star U_2U_2^\star\phi^L(\omega)}{L+1} \\ &= \frac{\langle \mathcal{P}_2\phi^L(\omega), \mathcal{P}_2[\phi^L(\omega)]'\rangle + \langle \mathcal{P}_2[\phi^L(\omega)]', \mathcal{P}_2\phi^L(\omega)\rangle}{L+1}, \end{aligned}$$

$$\begin{aligned} Q''(\omega) &= \frac{\phi^L(\omega)^\star U_2U_2^\star U_2U_2^\star[\phi^L(\omega)]'' + 2[\phi^L(\omega)]'^\star U_2U_2^\star U_2U_2^\star[\phi^L(\omega)]' + [\phi^L(\omega)]''^\star U_2U_2^\star U_2U_2^\star\phi^L(\omega)}{L+1} \\ &= \frac{\langle \mathcal{P}_2\phi^L(\omega), \mathcal{P}_2[\phi^L(\omega)]''\rangle + 2\|\mathcal{P}_2[\phi^L(\omega)]'\|_2^2 + \langle \mathcal{P}_2[\phi^L(\omega)]'', \mathcal{P}_2\phi^L(\omega)\rangle}{L+1}. \end{aligned} \tag{59}$$

Let $D(\omega) = Q^\varepsilon(\omega) - Q(\omega)$ and then

$$[Q^\varepsilon(\omega)]' = Q'(\omega) + D'(\omega),$$

$$[Q^\varepsilon(\omega)]'' = Q''(\omega) + D''(\omega).$$

First, we derive an upper bound of $|D'(\omega)|$ and $|D''(\omega)|$ in terms of α, L and $\|E\|_2$.

$$\begin{aligned} (L+1)|D'(\omega)| &= |\langle \mathcal{P}_2^\varepsilon\phi^L(\omega), \mathcal{P}_2^\varepsilon[\phi^L(\omega)]'\rangle - \langle \mathcal{P}_2\phi^L(\omega), \mathcal{P}_2[\phi^L(\omega)]'\rangle \\ &\quad + \langle \mathcal{P}_2^\varepsilon[\phi^L(\omega)]', \mathcal{P}_2^\varepsilon\phi^L(\omega)\rangle - \langle \mathcal{P}_2[\phi^L(\omega)]', \mathcal{P}_2\phi^L(\omega)\rangle| \\ &= |\langle \mathcal{P}_2^\varepsilon\phi^L(\omega), \mathcal{P}_2^\varepsilon[\phi^L(\omega)]'\rangle - \langle \mathcal{P}_2^\varepsilon\phi^L(\omega), \mathcal{P}_2[\phi^L(\omega)]'\rangle \\ &\quad + \langle \mathcal{P}_2^\varepsilon\phi^L(\omega), \mathcal{P}_2[\phi^L(\omega)]'\rangle - \langle \mathcal{P}_2\phi^L(\omega), \mathcal{P}_2[\phi^L(\omega)]'\rangle \\ &\quad + \langle \mathcal{P}_2^\varepsilon[\phi^L(\omega)]', \mathcal{P}_2^\varepsilon\phi^L(\omega)\rangle - \langle \mathcal{P}_2^\varepsilon[\phi^L(\omega)]', \mathcal{P}_2\phi^L(\omega)\rangle \\ &\quad + \langle \mathcal{P}_2^\varepsilon[\phi^L(\omega)]', \mathcal{P}_2\phi^L(\omega)\rangle - \langle \mathcal{P}_2[\phi^L(\omega)]', \mathcal{P}_2\phi^L(\omega)\rangle| \\ &\le \|\mathcal{P}_2^\varepsilon\phi^L(\omega)\|_2\|\mathcal{P}_2^\varepsilon - \mathcal{P}_2\|_2\|[\phi^L(\omega)]'\|_2 + \|\mathcal{P}_2[\phi^L(\omega)]'\|_2\|\mathcal{P}_2^\varepsilon - \mathcal{P}_2\|_2\|\phi^L(\omega)\|_2 \\ &\quad + \|\mathcal{P}_2^\varepsilon[\phi^L(\omega)]'\|_2\|\mathcal{P}_2^\varepsilon - \mathcal{P}_2\|_2\|\phi^L(\omega)\|_2 + \|\mathcal{P}_2\phi^L(\omega)\|_2\|\mathcal{P}_2^\varepsilon - \mathcal{P}_2\|_2\|[\phi^L(\omega)]'\|_2 \\ &\le 4\|\mathcal{P}_2^\varepsilon - \mathcal{P}_2\|_2\|\phi^L(\omega)\|_2\|[\phi^L(\omega)]'\|_2. \end{aligned}$$

Since $[\phi^L(\omega)]' = -2\pi i[0\ e^{-2\pi i\omega}\ 2e^{-2\pi i2\omega}\ 3e^{-2\pi i3\omega}\ \dots\ Le^{-2\pi iL\omega}]^T$, we have $\|[\phi^L(\omega)]'\|_2 = \eta(L)\sqrt{L+1}$ where $\eta(L) = 2\pi\sqrt{1^2+2^2+\dots+L^2}/\sqrt{L+1}$ and then

$$|D'(\omega)| \le 4\alpha\eta(L)\|E\|_2, \tag{60}$$

by Theorem 3.

Applying the same technique to $D''(\omega)$ yields

$$|D''(\omega)| \le 4\alpha\left[\eta^2(L) + \zeta(L)\right]\|E\|_2, \tag{61}$$

where $\zeta(L) = (2\pi)^2\sqrt{1^4+2^4+\dots+L^4}/\sqrt{L+1}$.

Next we prove that for each $\omega_j \in \mathcal{S}$, there exists a strict local minimizer $\hat{\omega}_j$ of R^ε near ω_j satisfying (28).

Since ω_j is a strict local minimizer of $Q(\omega)$ and

$$\mathcal{P}_2[\phi^L(\omega)]'\big|_{\omega=\omega_j} \ne \mathbf{0}, \quad \forall\omega_j \in \mathcal{S} \tag{62}$$

we have

$$Q'(\omega_j) = 0 \text{ and } Q''(\omega_j) > 0.$$

We break the following argument into several steps:

Step 1 Let $Q''(\omega_j) = 2m_j$. $m_j > 0$ due to assumption (62). Thanks to the smoothness of Q'', there exists $\delta_j > 0$ such that

$$Q''(\omega) > m_j, \ \forall \omega \in (\omega_j - \delta_j, \omega_j + \delta_j).$$

For sufficiently small noise satisfying

$$4\alpha[\eta^2(L) + \zeta(L)]\|E\|_2 < m_j/2, \tag{63}$$

we have

$$[Q^\varepsilon(\omega)]'' > m_j/2, \ \forall \omega \in (\omega_j - \delta_j, \omega_j + \delta_j).$$

Step 2 Consider $Q'(\omega_j - \delta_j)$ and $Q'(\omega_j + \delta_j)$. There exist $\kappa_1, \kappa_2 \in (0, \delta_j)$ such that

$$\begin{aligned} Q'(\omega_j - \delta_j) &= Q'(\omega_j) + Q''(\omega_j - \kappa_1)(-\delta_j) \\ Q'(\omega_j + \delta_j) &= Q'(\omega_j) + Q''(\omega_j + \kappa_1)\delta_j. \end{aligned}$$

From Step 1, $Q''(\omega_j - \kappa_1) > m_j$ and $Q''(\omega_j + \kappa_1) > m_j$, so $Q'(\omega_j - \delta_j) < -m_j\delta_j$ and $Q'(\omega_j + \delta_j) > m_j\delta_j$. According to (60), when noise is sufficiently small such that

$$4\alpha\eta(L)\|E\|_2 < m_j\delta_j/2, \tag{64}$$

we have

$$[Q^\varepsilon(\omega_j - \delta_2)]' < -m_j\delta_j/2 < 0 \text{ and } [Q^\varepsilon(\omega_j + \delta_3)]' > m_j\delta_j/2 > 0.$$

$[Q^\varepsilon(\omega)]'$ is a smooth function so there exists $\hat{\omega}_j \in (\omega_j - \delta_j, \omega_j + \delta_j)$ such that $[Q^\varepsilon(\hat{\omega}_j)]' = 0$ by intermediate value theorem.

Step 3 From Step 2 and Step 3 we have obtained an open interval containing ω_j: $(\omega_j - \delta_j, \omega_j + \delta_j)$ such that

$$Q''(\omega) > m_j \text{ and } [Q^\varepsilon(\omega)]'' > m_j/2, \ \forall \omega \in (\omega_j - \delta_j, \omega_j + \delta_j).$$

Also there exists $\hat{\omega}_j \in (\omega_j - \delta_j, \omega_j + \delta_j)$ such that

$$[Q^\varepsilon(\hat{\omega}_j)]' = 0 \text{ and } [Q^\varepsilon(\hat{\omega}_j)]'' > 0,$$

so $\hat{\omega}_j$ is a strict local minimizer of $Q^\varepsilon(\omega)$ and $R^\varepsilon(\omega)$.

Step 4 Furthermore

$$\begin{aligned} 0 = [Q^\varepsilon(\hat{\omega}_j)]' &= Q'(\hat{\omega}_j) + D'(\hat{\omega}_j) \\ &= Q'(\omega_j) + Q''(\xi_j)(\hat{\omega}_j - \omega_j) + D'(\hat{\omega}_j), \text{ for some } \xi_j \in (\omega_j, \hat{\omega}_j), \end{aligned}$$

through Taylor expansion of $Q'(\omega)$ at $\omega = \omega_j$. Since $Q'(\omega_j) = 0$,

$$|\hat{\omega}_j - \omega_j| \min_{\xi \in (\omega_j, \hat{\omega}_j)} |Q''(\xi)| \le |\hat{\omega}_j - \omega_j||Q''(\xi_j)| \le |D'(\hat{\omega}_j)| \le 4\alpha\eta(L)\|E\|_2 \tag{65}$$

Proof of Theorem 4 ends here. Next we provide the argument for the validation of Remark 9 and Remark 11.

Remark 9 Suppose noise vector ε contains i.i.d. random variables of variance σ^2. For fixed M, $\|E\|_2 \le \|E\|_F^2 = \mathcal{O}(\sigma)$. Since $\min_{\xi\in(\omega_j,\hat{\omega}_j)} |Q''(\xi)| > m_j$, we have $|\hat{\omega}_j - \omega_j| \to 0$ as $\sigma \to 0$ and

$$|\hat{\omega}_j - \omega_j| = \mathcal{O}(\sigma).$$

Remark 11 The asymptotic rate of $\hat{\omega}_j \to \omega_j$ as $M = 2L \to \infty$ in the case of $q \ge 4$ RL is discussed in Remark 11. Here we show that condition (63) and (64) hold as $M = 2L \to \infty$ under assumption (29).

The left hand side (l.h.s.) of (63) scales like $M\sqrt{M\log M}$. On the right hand side (r.h.s.),

$$m_j = \frac{2\|\mathcal{P}_2[\phi^L(\omega_j)]'\|_2^2}{L+1} \ge \frac{2C_1}{L+1}\|[\phi^L(\omega_j)]'\|_2^2 \sim M^2, \text{ as } M = 2L \to \infty.$$

Therefore the r.h.s. of (63) grows faster than the l.h.s. and (63) holds as $M = 2L \to \infty$.

Next we show that (64) is also valid as $M \to \infty$. First

$$\begin{aligned} Q'''(\omega) =& \frac{\langle \mathcal{P}_2\phi^L(\omega), \mathcal{P}_2[\phi^L(\omega)]'''\rangle + \langle \mathcal{P}_2[\phi^L(\omega)]''', \mathcal{P}_2\phi^L(\omega)\rangle}{L+1} \\ &+ \frac{3\langle \mathcal{P}_2[\phi^L(\omega)]', \mathcal{P}_2[\phi^L(\omega)]''\rangle + 3\langle \mathcal{P}_2[\phi^L(\omega)]'', \mathcal{P}_2[\phi^L(\omega)]'\rangle}{L+1} \end{aligned}$$

and then

$$\begin{aligned} |Q'''(\omega)| &\le \frac{2\|\phi^L(\omega)\|_2\|[\phi^L(\omega)]'''\|_2 + 6\|[\phi^L(\omega)]'\|_2\|[\phi^L(\omega)]''\|_2}{L+1} \\ &= \frac{2\sqrt{L+1}\sqrt{1^6+2^6+\ldots+L^6} + 6\sqrt{1^2+2^2+\ldots+L^2}\sqrt{1^4+2^4+\ldots+L^4}}{L+1} \\ &= \mathcal{O}(L^3) \text{ as } L \to \infty \\ &= \mathcal{O}(M^3) \text{ as } M = 2L \to \infty. \end{aligned}$$

In Step 1, for all $\omega \in (\omega_j - \delta_j, \omega_j + \delta_j)$, $|Q''(\omega) - Q''(\omega_j)|$ can be as large as $m_j = \mathcal{O}(M^2)$. Meanwhile $Q''(\omega) = Q''(\omega_j) + (\omega - \omega_j)Q'''(\kappa)$ for some $\kappa \in (\omega_j, \omega)$ and then $|Q''(\omega) - Q''(\omega_j)| = |\omega - \omega_j||Q'''(\kappa)|$. Since $|Q''(\omega) - Q''(\omega_j)|$ can be as large as $\mathcal{O}(M^2)$ and $Q'''(\kappa) \le \mathcal{O}(M^3)$, $|\omega - \omega_j|$ can be as large as $\mathcal{O}(1/M)$. In other words, $\delta_j = \mathcal{O}(1/M)$ in Step 1.

In (64), the l.h.s. scales like $\sqrt{M\log M}$ and the r.h.s. $= m_j\delta_j/2 = \mathcal{O}(M^2/M) = \mathcal{O}(M)$ as $M = 2L \to \infty$. As a result, (64) holds while $M = 2L \to \infty$ under assumption (29).

□

Appendix D Proof of Theorem 5

Proof. We partition $\mathcal{S}$ into the subsets $\mathcal{S}_m = \{\omega_{m+kR} : k \in \mathbb{Z}\}, m = 1, \ldots, R$, each of which satisfies the gap condition: $d(\omega_j, \omega_l) > R\rho$ for all $\omega_j, \omega_l \in \mathcal{S}_m$ with $j \ne l$. According to Theorem 2,

$$\frac{1}{L}\sum_{k=0}^{L}\left|\sum_{\omega_j\in\mathcal{S}_m\cap[0,1)} \mathrm{c}_j e^{-2\pi i k\omega_j}\right|^2 \le B(R\rho, L)\sum_{\omega_j\in\mathcal{S}_m\cap[0,1)}|\mathrm{c}_j|^2, \ m = 1, \ldots, R.$$

As $|z_1 + \ldots + z_R|^2 \leq R(|z_1|^2 + \ldots + |z_R|^2)$,

$$\frac{1}{L}\sum_{k=0}^{L}\left|\sum_{\omega_j \in \mathcal{S} \cap [0,1)} \mathrm{c}_j e^{-2\pi i k \omega_j}\right|^2 = \frac{1}{L}\sum_{k=0}^{L}\left|\sum_{m=1}^{R}\sum_{\omega_j \in \mathcal{S}_m \cap [0,1)} \mathrm{c}_j e^{-2\pi i k \omega_j}\right|^2$$

$$\leq \frac{R}{L}\sum_{m=1}^{R}\sum_{k=0}^{L}\left|\sum_{\omega_j \in \mathcal{S}_m \cap [0,1)} \mathrm{c}_j e^{-2\pi i k \omega_j}\right|^2 \leq R\sum_{m=1}^{R}\sum_{\omega_j \in \mathcal{S}_m \cap [0,1)} B(R\rho, L)|\mathrm{c}_j|^2$$

$$\leq B(R\rho, L)R\sum_{m=1}^{R}\sum_{\omega_j \in \mathcal{S}_m \cap [0,1)} |\mathrm{c}_j|^2 = B(R\rho, L)R\|\mathrm{c}\|_2^2.$$

□

References

[1] R. Adamczak, "A few remarks on the operator norm of random Toeplitz matrices", *Journal of Theoretical Probability* **23**, pp.85-108, 2010.

[2] C. Baiocchi, V. Komornik and P. Loreti, "Ingham-Beurling type theorems with weakened gap conditions," *Acta Mathematica Hungarica* **97(1)**, pp.55-95, 2002.

[3] G. Beylkin and L. Monzón, "On approximation of functions by exponential sums," *Applied and Computational Harmonic Analysis* **19(1)**, pp.17-48, 2005.

[4] E. J. Candès, "The restricted isometry property and its implications for compressed sensing", *Comptes Rendus Mathematique* **346(9)**, pp.589-592, 2008.

[5] E. J. Candès and C. Fernandez-Granda,"Super-resolution from noisy data", *Journal of Fourier Analysis and Applications* **19(6)**, pp.1229-1254, 2013.

[6] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution", *Communications on Pure and Applied Mathematics* **67(6)**, pp.906-956, June 2014.

[7] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," arXiv:1304.8126, 2013.

[8] M. Cheney, "The linear sampling method and the MUSIC algorithm", *Inverse Problems* **17(4)**, pp.591, 2001.

[9] M. Cheney, "A mathematical tutorial on synthetic aperture radar", *SIAM review* **43(2)**, pp.301-312, 2001.

[10] Y. Chi, A. Pezeshki, L. Scharf, A. Pezeshki, and R. Calderbank, "Sensitivity to basis mismatch in compressed sensing", *IEEE Transactions on Signal Processing* **59(5)**, pp.2182-2195, 2011.

[11] L. Demanet, D. Needell and N. Nguyen, "Super-resolution via superset selection and pruning", *Proceedings of the 10th International Conference on Sampling Theory and Applications*, 2013.

[12] A. J. Den Dekker and A. van den Bos, "Resolution: a survey", *Journal of Optical Society of America A* **14(3)**, pp. 547-557, 1997.

[13] A. J. Devaney, "Super-resolution processing of multi-static data using time reversal and MUSIC", *Journal of the Acoustical Society of America*, 2000.

[14] A. Di, "Multiple source location - a matrix decomposition approach", *IEEE Transactions on Acoustics, Speech and Signal Processing* **33(5)**, pp. 1086-1091, 1985.

[15] D.L. Donoho, "Superresolution via sparsity constraints", *SIAM Journal on Mathematical Analysis* **23**, pp.1309-1331, 1992.

[16] D. L. Donoho, "Compressed sensing", *IEEE Transactions on Information Theory* **52(4)**, pp.1289-1306, 2006.

[17] M.F. Duarte and R.G. Baraniuk, "Spectral compressive sensing", *Applied and Computational Harmonic Analysis* **35(1)**, pp. 111-129, 2013.

[18] A. Eftekhari and M. B. Wakin, "Greed is super: a new iterative method for super-resolution", in preparation.

[19] A. Fannjiang, T. Strohmer and P. Yan, "Compressed remote sensing of sparse objects", *SIAM Journal on Imaging Sciences* **3(3)**, pp.596-618, 2010.

[20] A. Fannjiang, "The MUSIC algorithm for sparse objects: a compressed sensing analysis", *Inverse Problems* **27**, 035013, 2011.

[21] A. Fannjiang, and W. Liao, "Coherence pattern-guided compressive sensing with unresolved grids", *SIAM Journal on Imaging Sciences* **5(1)**, pp.179-202, 2012.

[22] A. Fannjiang and W. Liao, "Mismatch and resolution in compressive imaging", *Proceedings of SPIE on Wavelets and Sparsity XIV* **8138**, 2012.

[23] A. Fannjiang and W. Liao, "Super-resolution by compressive sensing algorithms", *Asilomar conference on signals, systems and computers, IEEE Computer Society*, 2012.

[24] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming", version 2.0 beta. `http://cvxr.com/cvx`, September 2013.

[25] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs", Recent Advances in Learning and Control (a tribute to M. Vidyasagar), V. Blondel, S. Boyd, and H. Kimura, editors, pp.95-110, Lecture Notes in Control and Information Sciences, Springer, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

[26] M. Herman and T. Strohmer, "General Deviants: An Analysis of Perturbations in Compressed Sensing", *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Compressive Sensing* **4(2)**, pp.342-349, 2010.

[27] Y. Hua and T. 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)**, pp.814-824, 1990.

[28] A. E. Ingham, "Some trigonometrical inequalities with applications to the theory of series", *Mathematische Zeitschrift* **41(1)**, pp.367-379, 1936.

[29] I. Ipsen, "The eigenproblem and invariant subspaces: perturbation theory", GW Stewart. Birkhuser Boston, pp.71-93, 2010.

[30] S. M. Kay and S. L. Marple Jr, "Spectrum analysis - a modern perspective", *Proceedings of the IEEE* **69(11)**, pp.1380-1419, 1981.

[31] A. Kirsch, "The MUSIC algorithm and the factorization method in inverse scattering theory for inhomogeneous media", *Inverse problems* **18**, pp. 1025-1040, 2002.

[32] V. Komornik and P. Loreti, "Semi-discrete Ingham-type inequalities," *Applied Mathematics and Optimization* **55(2)**, pp.203-218, 2007.

[33] H. Krim and J. H. Cozzens, "A data-based enumeration technique for fully correlated signals", *IEEE Transactions on Signal Processing* **42(7)**, pp.1662-1668, 1994.

[34] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach", *IEEE Signal Processing Magazine* **13(4)**, pp.67-94, 1996.

[35] S. Y. Kung, K. S. Arun, and D. V. Bhaskar Rao, "State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem," *Journal of the Optical Society of America*, **73(12)**, pp.1799-1811, 1983.

[36] S. Kunis and D. Potts, "Stability results for scattered data interpolation by trigonometric polynomials", *SIAM Journal on Scientific Computing* **29(4)**, pp. 1403-1419, 2007.

[37] M. Negreanu and E. Zuazua, "Discrete Ingham inequalities and applications", *Comptes Rendus Mathematique* **338(4)**, pp.281-286, 2004.

[38] M. Negreanu and E. Zuazua, "Discrete Ingham inequalities and applications", *SIAM Journal on Numerical Analysis* **44(1)**, pp.412-448, 2006.

[39] D. Potts and M. Tasche, "Parameter estimation for multivariate exponential sums," *Electronic Transactions on Numerical Analysis* **40**, pp.204-224, 2013.

[40] G. R. B. de Prony, "Essai Experimentale et Analytique", *J. de L'Ecole Polytechnique* **2**, pp. 24-76, 1795.

[41] B. D. Rao and K. S. Arun, "Model based processing of signals: A state space approach", *Proceedings of the IEEE* **80(2)**, pp.283-309, 1992.

[42] R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation", Ph.D. thesis, Stanford Univ., Stanford, CA, Nov. 1981.

[43] R. O. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Transactions on Antennas and Propagation* **34(3)**, pp.276-280, 1986.

[44] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty - V: The discrete case", *Bell System Technical Journal* **57(5)**, pp.1371-1430, 1978.

[45] G. W. Stewart and J. G. Sun, *Matrix perturbation theory*, 1990.

[46] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, New Jersey, 2005.

[47] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid", *IEEE Transactions on Information Theory* **59(11)**, pp. 7465-7490, 2013.

[48] D. J. Thomson, "Spectrum estimation and harmonic analysis", *Proceedings of the IEEE* **70(9)**, pp.1055-1096, 1982.

[49] H. Weyl, "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)", *Mathematische Annalen* **71(4)**, pp.441-479, 1912.

[50] R. M. Young, *An Introduction to Non-Harmonic Fourier Series*, Academic Press, New York, 1980.