

# Super-resolution of generalized spikes and spectra of confluent Vandermonde matrices\*

Dmitry Batenkov<sup>1</sup> and Nuha Diab<sup>1</sup>

<sup>1</sup>Department of Applied Mathematics, Tel-Aviv University, Tel-Aviv, Israel.  
*dbatenkov@tauex.tau.ac.il, nuhadiab@tauex.tau.ac.il*

March 25, 2022

## Abstract

We study the problem of super-resolution of a linear combination of Dirac distributions and their derivatives on a one-dimensional circle from noisy Fourier measurements. Following numerous recent works on the subject, we consider the geometric setting of “partial clustering”, when some Diracs can be separated much below the Rayleigh limit. Under this assumption, we prove sharp asymptotic bounds for the smallest singular value of a corresponding rectangular confluent Vandermonde matrix with nodes on the unit circle. As a consequence, we derive matching lower and upper min-max error bounds for the above super-resolution problem, under the additional assumption of nodes belonging to a fixed grid.

## 1 Introduction

### 1.1 Background

The problem of computational super-resolution (SR) is to recover the fine details of an unknown object from inaccurate measurements of inherently low resolution [1]. In recent years, there is much interest in the problem of reconstructing a signal modelled by a linear combination of Dirac  $\delta$ -distributions (e.g. [2]–[11] and references therein):

$$\mu(x) = \sum_{j=1}^s a_j \delta_{\xi_j} \quad a_j \in \mathbb{C}, \delta_{\xi_j} = \delta(x - \xi_j), \xi_j \in [-\pi, \pi] \quad (1)$$

from noisy and bandlimited Fourier measurements:

$$y_k := \hat{\mu}(k) + \eta_k, \quad \hat{\mu}(k) = \langle \mu, e^{-ikx} \rangle, \quad k = 0, 1, \dots, M, \quad |\eta_k| \leq \varepsilon. \quad (2)$$

For the model (1) we have  $\hat{\mu}(k) = \sum_{j=1}^s a_j e^{ik\xi_j}$ , and therefore the measurement vector  $y = \{y_k\}_{k=0}^M$  can be expressed as

$$y = Va + \eta \in \mathbb{C}^{M+1} \quad (3)$$

where  $V$  is the  $(M + 1) \times s$  Vandermonde matrix with the nodes on the unit circle:

$$V := [e^{ik\xi_j}]_{k=0, \dots, M}^{j=1, \dots, s}.$$

In order to describe the stability of this inverse problem, suppose that the nodes  $\xi_j$  belong to a grid of step size  $\Delta$  and define the super-resolution factor (SRF) as  $\frac{1}{(M\Delta)}$ . Suppose that at most  $\ell \leq s$  nodes form a

---

\*This research was supported by Israel Science Foundation grant 1792/20 and by Lower Saxony - Israel collaboration grant from the Volkswagen Foundation.

”cluster” of size  $O(\Delta)$  (to be rigorously defined below). In the ”super-resolution regime”  $SRF \gg 1$  [4], [5] showed that  $\sigma_{\min}(V)$  scales like  $SRF^{\ell-1}$  and consequently the worst-case reconstruction error rate of the coefficients of  $\mu$  as in (1) from noisy measurements (2) is of the order  $SRF^{2\ell-1}\epsilon$ . Despite the great amount of research devoted to the subject, there is currently no known tractable algorithm which provably achieves these min-max bounds for all signals of interest [3].

## 1.2 Our contributions

In this work we extend the methods and results of [4], [5] to the model

$$\mu = \sum_{j=1}^s a_j \delta_{\xi_j} + b_j \delta'_{\xi_j}, \quad (4)$$

where  $\delta'$  is the distributional derivative of the Dirac delta. The Vandermonde matrix  $V$  in (3) is replaced by the so-called *confluent Vandermonde matrix*  $U$ , which is defined (up to normalization) as:

$$U := [e^{ik\xi_j} k e^{i(k-1)\xi_j}]_{\substack{j=1,\dots,s \\ k=0,\dots,M}}.$$

Under the partial clustering assumptions, in Theorem 3.1 and 3.2 we prove a sharp lower and upper bounds for the smallest singular value of  $U$  in the super-resolution regime, and show that it scales like  $SRF^{2\ell-1}$ . These bounds are proved by extending the decimation approach from [4] for the lower bound on  $\sigma_{\min}(U)$ , and by extending the finite difference approximation approach from [5] for the upper bound, further generalizing it to any node vector  $\boldsymbol{\xi}$  satisfying the clustering assumptions. In addition, our proof technique for bounding the remainder part in the upper bound of the smallest singular value can be applied to gain a slight improvement in Proposition 2.10 in [5] by relaxing the conditions on  $M, \Delta$ .

As a consequence, in Theorem 3.3 we also obtain sharp min-max bounds of order  $SRF^{4\ell-1}\epsilon$  for the problem of sparse superresolution of signals (4) on a grid by extending the corresponding technique from [4].

Also, we show numerically that the well-known ESPRIT method for exponential fitting (appropriately extended to handle higher multiplicities) is optimal, meaning that it attains the min-max error bounds we established in Theorem 3.3 for the recovered parameters of the signal (4).

In relation to prior work on the subject, in [12] the authors give a stability estimate for the more general model (5) with arbitrary fixed  $n$ , however assuming that the number of measurements  $N+1$  equals the number of unknowns. Evaluating their estimate for our model and notation, their bound is of order  $\Delta^{4\ell-1}\epsilon^{\frac{1}{2}}$ , while ours in the same case is  $\Delta^{4\ell-1}\epsilon$ . In contrast, [13] established the bound in the super-resolution setting of a single cluster (and off-grid nodes) to be of order  $SRF^{4\ell}\epsilon$ , while we derive the min-max rate  $SRF^{4\ell-1}\epsilon$ .

## 1.3 Discussion

Naturally, our results and techniques pave the way to analyzing the general model

$$\mu = \sum_{j=1}^s \sum_{l=0}^n a_{j,l} \delta_{\xi_j}^{(l)} \quad (5)$$

in the clustered super-resolution regime. The applications of this model include modern sampling theory beyond the Nyquist rate, algebraic signal recovery, and multi-exponential analysis, to name a few (see [12]–[16] and references therein). At the same time, we believe that several recent developments on the basic model (1) can be utilized to the more general setting, as follows.

- Recently, [2] succeeded to establish sharp bounds for the entire spectrum of  $V$  without requiring the entire node set to be contained in a small interval of length  $\frac{\pi}{s^2}$ . We believe that similar techniques could be applied to (4) in order to eliminate the above restriction.

- Optimal scaling of the constants in the above bounds for the spectrum of  $V$  using harmonic analysis techniques has been investigated in [17], and it would be interesting to extend these methods to (5).
- While we obtain min-max rates for nodes on a grid, we expect to get similar rates for the "off-grid" model as in [3], where the node locations can be any real number. Furthermore, it should be possible to establish component-wise bounds for the coefficients of different orders and for the nodes themselves, as done in [12], [13], [16] for the more restrictive geometric settings of the problem.

Going back to the model (1), confluent Vandermonde matrices appear naturally in the perturbation analysis of the nonlinear least squares problems [3], [13] for exponential fitting, and we expect our methods to be applicable in this context as well.

The paper is organized as follows. In section 2 we establish some notation. In section 3 we formulate the main results, which are proved in section 4. Finally, in section 5 we present numerical evidence confirming our bounds.

## 2 Preliminaries

### 2.1 Notation

**Definition 2.1.** For  $N \in \mathbb{N}$  and a vector  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_s)$  of pairwise distinct real nodes  $\xi_j \in (-\pi, \pi]$ , we define the rectangular  $(2N + 1) \times 2s$  **confluent Vandermonde matrix**  $U_N(\boldsymbol{\xi})$  as

$$U_N(\boldsymbol{\xi}) := \frac{1}{\sqrt{2N}} [z_j^k \quad kz_j^{k-1}]_{k=0, \dots, 2N}^{j=1, \dots, s}$$

s.t.  $z_j = \exp(i\xi_j)$ .

The main subject of the paper is the scaling of the smallest singular value of  $U_N$  when some of the nodes of  $\boldsymbol{\xi}$  nearly collide (become very close to each other).

**Definition 2.2** (wraparound distance). For  $t \in \mathbb{R}$ , we denote

$$\|t\|_{\mathbb{T}} := |\text{Arg} \exp(it)| = |t \bmod (-\pi, \pi]|,$$

where  $\text{Arg}(z)$  is the principal value of the argument of  $z \in \mathbb{C} \setminus \{0\}$ , taking values in  $(-\pi, \pi]$ .

**Definition 2.3** (minimal separation). Given a vector of  $s$  distinct nodes  $\mathbf{x} := (x_1, \dots, x_s)$  with  $x_j \in (-\pi, \pi]$ , we define the minimal separation (in wraparound sense) as

$$\Delta := \Delta(\mathbf{x}) = \min_{i \neq j} \|x_i - x_j\|_{\mathbb{T}}.$$

**Definition 2.4.** The node vector  $\mathbf{x} = (x_1, \dots, x_s) \subset (-\frac{\pi}{2}, \frac{\pi}{2}]$  is said to form a  $(\Delta, \rho, s, \ell, \tau)$ -clustered configuration for some  $\Delta > 0$ ,  $2 \leq \ell \leq s$ ,  $\ell - 1 \leq \tau \leq \frac{\pi}{\Delta}$ , and  $\rho \geq 0$  if for each  $x_j$  there exist at most  $\ell$  distinct nodes

$$\mathbf{x}^{(j)} = \{x_{j,k}\}_{k=1, \dots, r_j} \subset \mathbf{x}, \quad 1 \leq r_j \leq \ell, \quad x_{j,1} \equiv x_j,$$

such that the following conditions are satisfied:

1. For any  $y \in \mathbf{x}^{(j)} \setminus \{x_j\}$ , we have

$$\Delta \leq \|y - x_j\|_{\mathbb{T}} \leq \tau \Delta.$$

2. For any  $y \in \mathbf{x} \setminus \mathbf{x}^{(j)}$ , we have

$$\|y - x_j\|_{\mathbb{T}} \geq \rho.$$

**Definition 2.5.** For  $\Delta > 0$  let  $M = \lfloor \frac{\pi}{2\Delta} \rfloor$  and denote by  $\mathcal{T}_\Delta$  the discrete grid

$$\mathcal{T}_\Delta := \{k\Delta, \quad k = -M, \dots, M\} \subset \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right].$$

Further define  $G := G(\Delta) = |\mathcal{T}_\Delta| = 2M + 1$ .

**Definition 2.6.** For  $\Delta, \rho, s, \ell, \tau$  as in Definition 2.4, let  $\mathcal{R} := \mathcal{R}(\Delta, \rho, s, \ell, \tau)$  be the set of point distributions of the form  $\mu = \sum_{j=1}^s a_j \delta_{t_j} + b_j \delta'_{t_j}$  where  $t_j \in \mathcal{T}_\Delta$  and  $a_j, b_j \in \mathbb{C}$  for all  $j = 1, \dots, s$ , while  $\mathbf{t} = (t_1, \dots, t_s)$  forms a  $(\Delta, \rho, s, \ell, \tau)$ -clustered configuration.

**Definition 2.7.** For fixed  $N \in \mathbb{N}$ ,  $\epsilon > 0$ ,  $N\Delta < 1$  and  $\mu \in \mathcal{R}(\Delta, \rho, s, \ell, \tau)$ , let

$$B_\epsilon^N(\mu) := \left\{ y \in \mathbb{C}^{2N+1} : \left( \sum_{k=0}^{2N} |y_k - \hat{\mu}(k)|^2 \right)^{\frac{1}{2}} < \epsilon \right\},$$

where  $\hat{\mu}(k)$  are the Fourier coefficients as defined in (2).

**Definition 2.8.** Let  $\mathcal{A} := \mathcal{A}(\mathcal{R}, N, \epsilon)$  be the set of functions  $\varphi$  that maps each  $y \in \cup_{\mu \in \mathcal{R}} B_\epsilon^N(\mu)$  to a discrete distribution  $\varphi_y \in \mathcal{R}(\Delta, \rho, s, \ell, \tau)$ .

**Definition 2.9.** For  $\mu = \sum_{j=1}^s a_j \delta_{t_j} + b_j \delta'_{t_j}$ , the norm  $\|\mu\|_2$  is the discrete  $\ell_2$  norm of the coefficients vector:

$$\|\mu\|_2 := \left( \sum_{j=1}^s |a_j|^2 + |b_j|^2 \right)^{\frac{1}{2}}.$$

**Definition 2.10 (Min-Max Error).** The  $\ell^2$  min-max error for the on-the-grid model is

$$\mathcal{E}(\mathcal{R}, N, \epsilon) = \inf_{\varphi \in \mathcal{A}} \sup_{\mu \in \mathcal{R}} \sup_{y \in B_\epsilon^N(\mu)} \|\varphi_y - \mu\|_2,$$

where  $\varphi_y := \varphi(y) \in \mathcal{R}$ .

## 3 Main Results

### 3.1 Optimal bounds for the smallest singular value

As in previous works on the subject, the main quantity of interest is the smallest singular value of  $U_N$ .

**Theorem 3.1.** For each  $s \in \mathbb{N}$  there exists a constant  $C_1 = C_1(s)$  such that for any  $4\tau\Delta \leq \min(\rho, \frac{1}{s^2})$ , any  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_s) \subset \frac{1}{s^2} \left( -\frac{\pi}{2}, \frac{\pi}{2} \right]$  forming a  $(\Delta, \rho, s, \ell, \tau)$ -clustered configuration, and any  $N$  satisfying

$$\max \left( \frac{4\pi s}{\rho}, 4s^3 \right) \leq N \leq \frac{\pi s}{\tau\Delta}$$

we have

$$\sigma_{\min}(U_N(\boldsymbol{\xi})) \geq C_1 \cdot (N\Delta)^{2\ell-1}.$$

**Theorem 3.2.** For each  $s \in \mathbb{N}$  there exists a constant  $C_2 = C_2(\ell, \tau)$  such that for any  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_s)$  forming a  $(\Delta, \rho, s, \ell, \tau)$ -clustered configuration, and any  $N$  satisfying  $N \leq \frac{1}{2\Delta}$  we have

$$\sigma_{\min}(U_N(\boldsymbol{\xi})) \leq C_2 \cdot (N\Delta)^{2\ell-1}.$$

The proofs of the above results are given in Sections 4.2 and 4.3, respectively. For the lower bound, we extend the decimation technique from [4] to the confluent setting. For the upper bound we generalize the approach from [5] to hold for any clustered configuration  $\boldsymbol{\xi}$ . Furthermore, our proof technique can be used to slightly improve the condition (2.9) in Proposition 2.10 in [5] by requiring only that  $N\Delta \leq \text{const}$  instead of  $N^{3/2}\Delta \leq \text{const}$ .

### 3.2 Stable superresolution of generalized spikes of order 1

In our setting, we assume that the spike locations are restricted to a discrete grid of step size  $\Delta$ . In effect, our results show that  $\mathcal{E} \asymp SRF^{4\ell-1}\epsilon$  as  $SRF \rightarrow \infty$ .

**Theorem 3.3.** Fix  $s \geq 1, 2 \leq \ell \leq s, \epsilon > 0$ . Put  $SRF := \frac{1}{N\Delta}$ . Then the following hold:

1. For any  $\rho \geq 0, \ell - 1 \leq \tau$ , and  $M \geq \pi$ , there exists  $\beta \geq M$  such that for all sufficiently small  $\Delta \leq \Delta_0$ , it holds that

$$\mathcal{E}(\hat{\mathcal{R}}(\Delta, \rho, s, \ell, \tau), N, \epsilon) \leq C_{s,\ell} SRF^{4\ell-1}\epsilon, \quad SRF = \beta$$

for some constant  $C_{s,\ell}$  depending only on  $s$  and  $\ell$ , where  $\hat{\mathcal{R}} := \left\{ \mu : \mu \in \mathcal{R}, \text{supp}(\mu) \subset \frac{1}{s^2} \left( -\frac{\pi}{2}, \frac{\pi}{2} \right) \right\}$ .

2. For any  $\rho \geq 0, \ell - 1 \leq \tau$  and  $N\Delta \leq \frac{1}{2}$ , it holds that

$$\mathcal{E}(\mathcal{R}(\Delta, \rho, s, \ell, \tau), N, \epsilon) \geq C_{\ell,\tau} SRF^{4\ell-1}\epsilon$$

for some constant  $C_{\ell,\tau}$  depending only on  $\ell$  and  $\tau$ .

The proof is presented in Section 4.4, largely repeating the arguments from [4], [5], together with the above established bounds on  $\sigma_{\min}(U_N)$ .

## 4 Proofs

### 4.1 Square confluent Vandermonde matrices

**Definition 4.1.** For  $s \in \mathbb{N}$  and vector  $\mathbf{z} = (z_1, \dots, z_s)$  of pairwise distinct complex nodes  $|z_j| = 1$ , we define the square  $2s \times 2s$  confluent Vandermonde matrix

$$\mathbf{U}_{2s}(\mathbf{z}) := [z_j^k \quad kz_j^{k-1}]_{\substack{j=1,\dots,s \\ k=0,\dots,2s-1}}$$

**Theorem 4.1** ([18]). Let  $\mathbf{x} = (x_1, \dots, x_n)$  be a vector of pairwise distinct complex numbers and let

$$b_\lambda = \max \left( 1 + |x_\lambda|, 1 + 2(1 + |x_\lambda|) \sum_{\nu=1, \nu \neq \lambda}^n \frac{1}{|x_\nu - x_\lambda|} \right).$$

Then

$$\|\mathbf{U}_{2n}^{-1}(\mathbf{x})\|_\infty \leq \max_{1 \leq \lambda \leq n} b_\lambda \left( \prod_{\nu=1, \nu \neq \lambda}^n \frac{1 + |x_\nu|}{|x_\nu - x_\lambda|} \right)^2$$

**Proposition 4.1.** Let  $\mathbf{z} = (z_1, \dots, z_s)$  be a vector of pairwise distinct complex nodes with  $|z_j| = 1, j = 1, \dots, s$ . Denote by  $\delta_{j,k}$  the angular distance between  $z_j$  and  $z_k$ :

$$\delta_{j,k} := \delta_{j,k}(\mathbf{z}) = \left| \text{Arg} \left( \frac{z_j}{z_k} \right) \right| = \left| \text{Arg}(z_j) - \text{Arg}(z_k) \bmod(-\pi, \pi) \right|.$$

Then

$$\sigma_{\min}(\mathbf{U}_{2s}(\mathbf{z})) \geq \frac{\pi^{2(1-s)}}{\sqrt{2s}} \min_{1 \leq j \leq s} \gamma_j \prod_{k \neq j} \delta_{j,k}^2,$$

where

$$\gamma_j = \min \left( \frac{1}{2}, \frac{2}{5\pi} \left( \sum_{k \neq j} \delta_{j,k}^{-1} \right)^{-1} \right)$$

*Proof.* By Theorem 4.1 we have

$$\|\mathbf{U}_{2s}^{-1}(\mathbf{z})\|_{\infty} \leq 2^{2(s-1)} \max_{1 \leq j \leq s} b_j \prod_{k \neq j} |z_k - z_j|^{-2}, \quad (6)$$

where

$$b_j = \max \left( 2, 5 \sum_{k \neq j}^s |z_j - z_k|^{-1} \right).$$

For any  $|\theta| \leq \frac{\pi}{2}$ , we have

$$\frac{\pi}{2} |\theta| \leq \sin |\theta| \leq |\theta|$$

and since for any  $z_j \neq z_k$

$$|z_j - z_k| = \left| 1 - \frac{z_j}{z_k} \right| = 2 \sin \left| \frac{1}{2} \text{Arg} \left( \frac{z_j}{z_k} \right) \right| = 2 \sin \left| \frac{\delta_{j,k}}{2} \right|,$$

we therefore obtain

$$\frac{\pi}{2} \delta_{j,k} \leq |z_j - z_k| \leq \delta_{j,k}.$$

Plugging into (6), we have

$$\sigma_{\max}(\mathbf{U}_{2s}^{-1}(\mathbf{z})) \leq \sqrt{2s} \|\mathbf{U}_{2s}^{-1}(\mathbf{z})\|_{\infty} \leq \pi^{2(s-1)} \sqrt{2s} \max_{1 \leq j \leq s} b_j \prod_{k \neq j}^s \delta_{j,k}^{-2}$$

and

$$b_j = \max \left( 2, \frac{5\pi}{2} \sum_{k \neq j}^s \delta_{j,k}^{-1} \right).$$

This finishes the proof with  $\gamma_j := b_j^{-1}$ . □

## 4.2 Proof of Theorem 3.1.

### 4.2.1 Overview of the proof

First we use the **Decimation** technique that has first been introduced in [4]. It states that there exists a certain blow-up factor  $\lambda$  such that the mapped nodes  $\{e^{i\lambda x_j}\}$  attain "good" separation properties. Second, for any such  $\lambda$  of order  $O(N)$ , we can partition the rectangular confluent Vandermonde matrix into squared well-conditioned confluent matrices and use this partition to bound  $\sigma_{\min}$  from below.

In order to use the corresponding results from [4], we introduce an auxiliary bandwidth parameter  $\Omega$ .

**Definition 4.2.** For  $N, s \in \mathbb{N}$ , a vector  $\mathbf{x} = (x_1, \dots, x_s)$  of pairwise distinct real nodes  $x_j \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right]$ , and a bandwidth parameter  $2N \geq \Omega > 0$ , let  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_s)$  where  $\xi_j = \frac{x_j \Omega}{N}$ . Then we define

$$U_N(\mathbf{x}, \Omega) := U_N(\boldsymbol{\xi}) = U_N\left(\frac{\Omega}{N} \mathbf{x}\right) = \frac{1}{\sqrt{2N}} \left[ \exp\left(ik \frac{x_j \Omega}{N}\right) \quad k \exp\left(i(k-1) \frac{x_j \Omega}{N}\right) \right]_{k=0, \dots, 2N}^{j=1, \dots, s} \in \mathbb{C}^{(2N+1) \times (2s)}.$$

### 4.2.2 The existence of an admissible decimation

We can now use a key result from [4].

**Lemma 4.1** (Lemma 4.1 in [4]). *Let  $\mathbf{x}$  form a  $(\Delta, \rho, s, \ell, \tau)$  clustered configuration, and suppose that  $\frac{4\pi s}{\rho} \leq \Omega \leq \frac{\pi s}{\tau \Delta}$ . Then, for any  $0 \leq \xi \leq 1$ , there exists a set  $I \subset \left[\frac{\Omega}{2s}, \frac{\Omega}{s}\right]$  of total measure  $\frac{\Omega}{2s} \xi$  such that for every  $\lambda \in I$ , the following holds for every  $x_j \in \mathbf{x}$ :*

1.

$$\|\lambda y - \lambda x_j\|_{\mathbb{T}} \geq \lambda \Delta \geq \frac{\Delta \Omega}{2s} \quad \forall y \in \mathbf{x}^{(j)} \setminus \{x_j\}$$

2.

$$\|\lambda y - \lambda x_j\|_{\mathbb{T}} \geq \frac{1-\xi}{s^2} \pi \quad \forall y \in \mathbf{x} \setminus \mathbf{x}^{(j)}$$

Furthermore, the set  $I^c := \left[ \frac{\Omega}{2s}, \frac{\Omega}{s} \right] \setminus I$  is a union of at most  $\frac{s^2}{2} \left\lceil \frac{\Omega}{4s} \right\rceil$  intervals.

Fix  $\xi = \frac{1}{2}$  and consider the set  $I$  given by the above Lemma. Let us also fix a finite and positive integer  $N$  and consider the set of  $2N + 1$  equispaced points in  $[0, 2\Omega]$ :

$$P_N := \left\{ k \frac{\Omega}{N} \right\}_{k=0, \dots, 2N}.$$

**Proposition 4.2.** *If  $N > 2s^3 \left\lceil \frac{\Omega}{4s} \right\rceil$ , then  $P_N \cap I \neq \emptyset$ .*

*Proof.* Exactly as the proof of Proposition 4.2 in [4]. □

We are now in a position to extend the main result from [4] to the confluent setting.

**Theorem 4.2.** *There exists a constant  $C = C(s)$  such that for any  $\mathbf{x}$  forming a  $(\Delta, \rho, s, \ell, \tau)$ -clustered configuration, and any  $\Omega$  satisfying*

$$\frac{4\pi s}{\rho} \leq \Omega \leq \frac{\pi s}{\tau \Delta},$$

*we have*

$$\sigma_{\min}(U_N(\mathbf{x}, \Omega)) \geq C \cdot (\Delta \Omega)^{2\ell-1} \quad \text{whenever } N > 2s^3 \left\lceil \frac{\Omega}{4s} \right\rceil.$$

*Proof.* Similarly to the proof of theorem 3.2 in [4], for any subset  $R \subset \{0, \dots, 2N\}$  let  $U_{N,R}$  be the submatrix of  $U_N$  containing only the rows in  $R$ . In particular, if  $\{0, \dots, 2N\} = R_1 \dot{\cup} \dots \dot{\cup} R_p$  then

$$\sigma_{\min}^2(U_N) \geq \sum_{n=1}^p \sigma_{\min}^2(U_{N,R_n}).$$

By Lemma 4.1 and Proposition 4.2, there exists  $m \in \mathbb{N}$ ,  $0 \leq m \leq 2N$  such that

$$u_j := x_j \frac{\Omega}{N} m = \lambda x_j$$

with

$$\begin{aligned} \frac{\tau}{2s}(\Delta \Omega) &\geq \|u_j - u_k\|_{\mathbb{T}} \geq \frac{1}{2s}(\Delta \Omega) && \forall x_k \in \mathbf{x}^{(j)} \setminus \{x_j\}; \\ \pi &\geq \|u_j - u_k\|_{\mathbb{T}} \geq \frac{\pi}{2s^2} && \forall x_k \in \mathbf{x} \setminus \mathbf{x}^{(j)}. \end{aligned}$$

Since  $\lambda \leq \frac{\Omega}{s}$  we conclude that  $2ms \leq 2N$ .

We will divide  $U_N$  to  $m$  squared matrices of size  $2s \times 2s$  in the following form:

$$\begin{aligned} R_0 &= \{0, m, \dots, (2s-1)m\}, \\ R_1 &= \{1, m+1, \dots, (2s-1)m+1\}, \\ &\vdots \\ R_{m-1} &= \{m-1, 2m-1, \dots, 2sm-1\}. \end{aligned}$$

For  $k = 0, 1, \dots, m-1$  each  $U_{N, R_k}$  is a square confluent Vandermonde matrix, and it can be checked by direct computation that

$$U_{N, R_k}(\boldsymbol{\nu}) = \frac{1}{\sqrt{2N}} \mathbf{U}_{2s}(\boldsymbol{\nu}) D(\mathbf{z}, m) T(\mathbf{z}, k)$$

where  $\boldsymbol{\nu} = \{\exp(iu_j)\}_{j=1}^s$  and  $\mathbf{z} = \{\exp(ix_j \frac{\Omega}{N})\}_{j=1}^s$ , with

$$D(\mathbf{z}, m) = \text{diag}(1, \dots, 1, mz_1^{m-1}, \dots, mz_s^{m-1}),$$

$$T(\mathbf{z}, r) := \begin{bmatrix} z_1^r & \dots & 0 & rz_1^{r-1} & \dots & 0 \\ \vdots & \ddots & \vdots & & \ddots & \vdots \\ 0 & \dots & z_s^r & 0 & \dots & rz_s^{r-1} \\ 0 & \dots & 0 & z_1^r & \dots & 0 \\ \vdots & \ddots & \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & z_s^r \end{bmatrix}.$$

Recall the well-known formula for a block matrix inverse.

**Lemma 4.2** (e.g. [19]). *Consider the block upper triangular matrix*

$$\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}.$$

*It is invertible if and only if both  $A$  and  $D$  are invertible, and its inverse is given by*

$$\begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}$$

**Lemma 4.3.** *For  $r \in \mathbb{Z}$ ,  $s, m \in \mathbb{N}$ ,  $m \neq 0$  and vector  $\mathbf{z} = (z_1, \dots, z_s)$  of pairwise distinct complex nodes with  $|z_j| = 1$  we have*

$$\|P^{-1}(\mathbf{z}, r, m)\|_{\infty} = \left| \frac{r}{m} \right| + 1,$$

where

$$P(\mathbf{z}, r, m) = D(\mathbf{z}, m) T(\mathbf{z}, r).$$

*Proof.* By direct computation,

$$P(\mathbf{z}, r, m) := \begin{bmatrix} z_1^r & \dots & 0 & rz_1^{r-1} & \dots & 0 \\ \vdots & \ddots & \vdots & & \ddots & \vdots \\ 0 & \dots & z_s^r & 0 & \dots & rz_s^{r-1} \\ 0 & \dots & 0 & mz_1^{r+m-1} & \dots & 0 \\ \vdots & \ddots & \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & mz_n^{r+m-1} \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

where

$$A := \text{diag}(z_1^r, \dots, z_s^r), \quad B := \text{diag}(rz_1^{r-1}, \dots, rz_s^{r-1}),$$

$$C := \text{diag}(mz_1^{r+m-1}, \dots, mz_s^{r+m-1}).$$

By Lemma 4.2 we get

$$P^{-1}(\mathbf{z}, r, m) = \begin{bmatrix} A^{-1} & -A^{-1}BC^{-1} \\ 0 & C^{-1} \end{bmatrix},$$



where  $-A^{-1}BC^{-1} = \text{diag}(-\frac{r}{m}z_1^{-(r+m)}, \dots, -\frac{r}{m}z_s^{-(r+m)})$ . Thus

$$\|P^{-1}(\mathbf{z}, r, m)\|_\infty = |z_k^{-r}| + \left| -\frac{r}{m}z_k^{-(r+m)} \right| = 1 + \left| \frac{r}{m} \right|.$$

□

Now, let us take a look at  $\gamma_j$  from Proposition 4.1:

$$\gamma_j = \min \left( \frac{1}{2}, \frac{2}{5\pi} \left( \sum_{k \neq j}^s \delta_{j,k}^{-1} \right)^{-1} \right),$$

where

$$\delta_{j,k} := \delta_{j,k}(\boldsymbol{\nu}).$$

We will show two properties:

1.

$$\begin{aligned} \sum_{k \neq j} \delta_{j,k}^{-1} &\leq \frac{2s\ell}{\Delta\Omega} + \frac{2(s-\ell)s^2}{\pi} \leq \frac{2s\ell\pi + 2(s-\ell)s^2\Delta\Omega}{\Delta\Omega\pi} \\ \frac{2}{5\pi} \left( \sum_{k \neq j} \delta_{j,k}^{-1} \right)^{-1} &\geq \frac{\Delta\Omega\pi}{5\pi(s\ell\pi + (s-\ell)s^2\Delta\Omega)} \geq \frac{1}{5} \frac{\Delta\Omega}{\pi s^2 + s^3\Delta\Omega} = \frac{\Delta\Omega}{5s^2} \frac{1}{(\pi + s\Delta\Omega)} \end{aligned}$$

Given that  $\Delta\Omega < \frac{\pi s}{\tau}$ , we have

$$\begin{aligned} \Omega\Delta s + \pi &\leq \frac{\pi s^2}{\tau} + \pi \leq \frac{\pi s^2}{\ell-1} + \pi \leq \pi(1+s^2) \\ \Rightarrow \frac{2}{5\pi} \left( \sum_{k \neq j} \delta_{j,k}^{-1} \right)^{-1} &\geq \kappa(s)\Delta\Omega, \end{aligned} \tag{P1}$$

where

$$\kappa(s) = \frac{1}{5\pi s^2(1+s^2)}.$$

2. Using  $2 \leq \ell \leq s$  we get

$$\begin{aligned} \frac{2}{5\pi} \left( \sum_{k \neq j} \delta_{j,k}^{-1} \right)^{-1} &\leq \left( \left( \frac{2s}{\tau\Delta\Omega} \right) \ell + \frac{1}{\pi}(s-\ell) \right)^{-1} \frac{2}{5\pi} \\ &= \frac{2}{5} \frac{\tau\Delta\Omega}{2s\ell\pi + (s-\ell)\tau\Delta\Omega} \leq \frac{2}{5} \frac{\tau\Delta \frac{\pi s}{\tau\Delta}}{2s\ell\pi + (s-\ell)\tau\Delta\Omega} \\ &\leq \frac{1}{5} \frac{\pi s}{s\ell\pi} \leq \frac{1}{5} \frac{1}{\ell} \leq \frac{1}{10} \leq \frac{1}{2} \\ &\Rightarrow \frac{2}{5\pi} \left( \sum_{k \neq j} \delta_{j,k}^{-1} \right)^{-1} \leq \frac{1}{2}. \\ \Rightarrow \gamma_j &= \min \left( \frac{1}{2}, \frac{2}{5\pi} \left( \sum_{k \neq j} \delta_{j,k}^{-1} \right)^{-1} \right) = \frac{2}{5\pi} \left( \sum_{k \neq j} \delta_{j,k}^{-1} \right)^{-1}. \end{aligned} \tag{P2}$$

Using Proposition 4.1 and Lemma 4.3 we are going to bound from below the smallest singular value of the *square* confluent Vandermonde matrix:

$$\begin{aligned}
\sigma_{\min}(U_{N,R,r}) &= \sigma_{\min}\left(\frac{1}{\sqrt{2N}}\mathbf{U}_{2s}(\boldsymbol{\nu})D(\mathbf{z},m)T(\mathbf{z},r)\right) \geq \sigma_{\min}(\mathbf{U}_{2s}(\boldsymbol{\nu}))(\sqrt{2N}\|P^{-1}(\mathbf{z},r,m)\|_{\infty})^{-1} \\
&\geq \frac{1}{\sqrt{2s \cdot 2N}}\pi^{2(1-s)}\left(1 + \left|\frac{r}{m}\right|\right)^{-1} \min_{1 \leq j \leq s} \gamma_j \prod_{k \neq j}^s \delta_{j,k}^2(\boldsymbol{\nu}) \\
&= \frac{\pi^{2(1-s)}}{\sqrt{2s}} \frac{\kappa(s)}{\sqrt{2N}} \left(1 + \left|\frac{r}{m}\right|\right)^{-1} (\Delta\Omega)^{2\ell-1} \\
&= \frac{\tilde{\kappa}(s)}{\sqrt{2N}} \left(1 + \left|\frac{r}{m}\right|\right)^{-1} (\Delta\Omega)^{2\ell-1}
\end{aligned}$$

where  $\tilde{\kappa}(s) = \frac{\pi^{2(1-s)}}{\sqrt{2s}}\kappa(s)$ . Ahead of the last step we used properties (P1) and (P2).

Finally, we can bound from below the smallest singular value of the *rectangular* confluent Vandermonde matrix:

$$\begin{aligned}
\sigma_{\min}^2(U_N) &\geq \sum_{r=0}^{m-1} \left(1 + \left|\frac{r}{m}\right|\right)^{-2} \frac{\tilde{\kappa}^2(s)}{2N} (\Delta\Omega)^{2(2\ell-1)} \\
&\geq \frac{\tilde{\kappa}^2(s)}{2N} (\Delta\Omega)^{2(2\ell-1)} \sum_{r=0}^{m-1} (2)^{-2} \\
&\geq \frac{m\tilde{\kappa}^2(s)}{8N} (\Delta\Omega)^{4\ell-2} \\
&\geq \frac{\tilde{\kappa}^2(s)}{16s} (\Delta\Omega)^{4\ell-2}.
\end{aligned}$$

We used the fact that  $m = \frac{\lambda N}{\Omega} \geq \frac{\Omega N}{2s\Omega} = \frac{N}{2s}$ .

To summarize, the final result for Theorem 4.2 is

$$\sigma_{\min}(U_N(\mathbf{x},\Omega)) \geq C_1(s)(\Delta\Omega)^{2\ell-1}, \quad C_1(s) := \frac{\tilde{\kappa}(s)}{\sqrt{16s}}. \quad \square$$

*Proof of Theorem 3.1.* Similar to [4, Corollary 3.6]), for  $\Omega := \frac{N}{s^2}$  and any  $\boldsymbol{\xi} \subset \frac{1}{s^2}(-\frac{\pi}{2}, \frac{\pi}{2}]$  forming a  $(\Delta, \rho, s, \ell, \tau)$ -clustered configuration with the conditions of Theorem 3.1, we have  $\mathbf{x} = \frac{N}{\Omega}\boldsymbol{\xi} \subset (-\frac{\pi}{2}, \frac{\pi}{2}]$  which forms a  $(\tilde{\Delta}, \tilde{\rho}, s, \ell, \tau)$ -clustered configuration with  $\tilde{\Delta} := s^2\Delta$  and  $\tilde{\rho} := s^2\rho$ . Clearly,  $4\tau\tilde{\Delta} \leq s^2\rho = \tilde{\rho}$  and also

$$\Omega s^2 = N \geq 4s^3 \Rightarrow \frac{\Omega}{4s} \geq 1 \Rightarrow \frac{2\Omega}{4s} \geq \left\lceil \frac{\Omega}{4s} \right\rceil \Rightarrow N = \Omega s^2 \geq 2s^3 \left\lceil \frac{\Omega}{4s} \right\rceil,$$

thus the conditions of Theorem 4.2 are satisfied for  $\mathbf{x}, \Omega, \tilde{\rho}, \tilde{\Delta}, \tau$ . Therefore

$$\sigma_{\min}(U_N(\boldsymbol{\xi})) = \sigma_{\min}(U_N(\mathbf{x},\Omega)) \geq C \cdot (\tilde{\Delta}\Omega)^{2\ell-1} = C \cdot \left(\frac{N}{\Omega}\Delta\Omega\right)^{2\ell-1} = C \cdot (\Delta N)^{2\ell-1},$$

finishing the proof of Theorem 3.1. □

### 4.3 Proof of Theorem 3.2.

**Definition 4.3.** For  $M, s \in \mathbb{N}$  and a vector  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_s)$  of pairwise distinct real nodes  $\omega_j \in \mathbb{T}$ , let  $\Phi_M$  denote the  $(M+1) \times 2s$  confluent Vandermonde matrix

$$\Phi_M(\boldsymbol{\omega}) = \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ z_1 & \dots & z_s & 1 & \dots & 1 \\ z_1^2 & \dots & z_s^2 & 2z_1 & \dots & 2z_s \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ z_1^M & \dots & z_s^M & Mz_1^{M-1} & \dots & Mz_s^{M-1} \end{pmatrix},$$

and let  $V_M$  denote the  $(M+1) \times 2s$  pascal Vandermonde matrix

$$V_M(\boldsymbol{\omega}) = \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ z_1 & \dots & z_s & 2\pi iz_1 & \dots & 2\pi iz_s \\ z_1^2 & \dots & z_s^2 & 4\pi iz_1^2 & \dots & 4\pi iz_s^2 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ z_1^M & \dots & z_s^M & M2\pi iz_1^M & \dots & M2\pi iz_s^M \end{pmatrix},$$

where  $z_j = \exp(-2\pi i\omega_j)$  and  $\mathbb{T}$  is the periodic interval  $[0, 1)$ .

By direct computation we get

$$V_M = \Phi_M H$$

with  $H = \text{diag}(1, \dots, 1, 2\pi iz_1, \dots, 2\pi iz_s)$ .

Inspired by the proof of Proposition 2.10 in [5], we will consider  $\boldsymbol{\omega} = \frac{\boldsymbol{\xi}}{-2\pi} + \frac{1}{2}$  where  $\boldsymbol{\xi}$  is a  $(\Delta, \rho, s, \ell, \tau)$ -clustered configuration and a suitable vector  $u$  in order to obtain an upper bound for

$$\sigma_{\min}(\Phi_M(\boldsymbol{\omega})) = \min_{u \in \mathbb{C}^{2s}, u \neq 0} \frac{\|\Phi_M(\boldsymbol{\omega})u\|_2}{\|u\|_2}.$$

Put  $\alpha := M\Delta$ , assume that  $M < \frac{1}{\Delta}$  and let  $\boldsymbol{\omega} = \{\omega_j\}_{j=1}^s$  be defined w.l.o.g by  $\omega_j = \tau_j \frac{\alpha}{M}$  where  $\tau_1 = 0$ ,  $\tau_j < \tau_{j+1}$ ,  $\tau_\ell \leq \tau$  for  $1 \leq j \leq \ell$ , while  $\{\omega_j\}_{\ell+1}^s$  are arbitrary.

**Definition 4.4.** We consider the vector  $u \in \mathbb{C}^{2s}$  defined by

$$\begin{aligned} u_j &:= \left(\frac{\alpha}{M}\right)^{2\ell-1} A_j & 1 \leq j \leq \ell & \quad (\star) \\ u_{s+j} &:= \left(\frac{\alpha}{M}\right)^{2\ell-1} 2\pi iz_j B_j & 1 \leq j \leq \ell & \\ u_j = u_{s+j} &= 0 & \text{otherwise,} & \end{aligned}$$

where  $z_j := \exp(-2\pi i\omega_j)$  and  $A_j, B_j$  are as given by equation  $(\star)$  from appendix A.1.

Let  $\tilde{u}_j := u_j$ ,  $\tilde{u}_{s+j} := \frac{z_j^{-1}}{2\pi i} u_{s+j}$  for  $1 \leq j \leq \ell$  and  $\tilde{u}_j = \tilde{u}_{s+j} = 0$  otherwise. To estimate  $\|\Phi u\|_2$ , we identify  $u$  with the discrete distribution

$$\mu := \sum_{j=1}^{\ell} \tilde{u}_j \delta_{\tau_j \frac{\alpha}{M}} + \tilde{u}_{s+j} \delta'_{\tau_j \frac{\alpha}{M}}. \quad (7)$$

We also define a modified Dirichlet kernel  $D_M \in C^\infty(\mathbb{T})$  by

$$D_M(\omega) := \sum_{m=0}^M \exp(2\pi im\omega). \quad (8)$$

**Lemma 4.4.** For  $\mu$  and  $D_M$  as defined in (7), (8), the following is true:

$$\sum_{m=0}^M |\hat{\mu}(m)|^2 = \|\mu * D_M\|_{L^2(\mathbb{T})}^2.$$

The proof of the above lemma is in appendix A.4. Thus, observe the following:

$$\|\Phi_M u\|_2^2 = \|V_M H^{-1} u\|_2^2 = \|V_M \tilde{u}\|_2^2 = \sum_{m=0}^M |(V_M \tilde{u})_m|^2 = \sum_{m=0}^M |\hat{\mu}(m)|^2 = \|\mu * D_M\|_{L^2(\mathbb{T})}^2.$$

As shown in appendix A.1, we see that for all  $\omega \in \mathbb{T}$

$$\begin{aligned} (\mu * D_M)(\omega) &= \sum_{j=1}^{\ell} \left( \tilde{u}_j D_M \left( \omega - \frac{\tau_j \alpha}{M} \right) + \tilde{u}_{s+j} D'_M \left( \omega - \frac{\tau_j \alpha}{M} \right) \right) \\ &= \left( \frac{\alpha}{M} \right)^{2\ell-1} D_M^{(2\ell-1)}(\omega) + \left( \frac{\alpha}{M} \right)^{2\ell-1} \{R_A(\omega) + R_B(\omega)\}, \end{aligned} \quad (9)$$

where  $R_A(\omega)$  and  $R_B(\omega)$  are written explicitly in section A.1 in the appendix.

By the Bernstein inequality for trigonometric polynomials [20], we have

$$\|D_M^{(2\ell-1)}\|_{L^2(\mathbb{T})} \leq (2\pi M)^{2\ell-1} \|D_M\|_{L^2(\mathbb{T})} = \sqrt{M+1} (2\pi M)^{2\ell-1}. \quad (10)$$

**Lemma 4.5.** For  $R_A(\omega)$  and  $R_B(\omega)$  as defined in section A.1 in the appendix, we have

$$\begin{aligned} \|R_A(\omega)\|_{L^2(\mathbb{T})} &\leq \left( \frac{\alpha}{M} \right)^{2\ell} \sqrt{M+1} (2\pi M)^{2\ell} \frac{\tau^{2\ell}}{(2\ell-1)!} \sum_{i=1}^{\ell} |A_i|, \\ \|R_B(\omega)\|_{L^2(\mathbb{T})} &\leq \left( \frac{\alpha}{M} \right)^{2\ell-1} \sqrt{M+1} (2\pi M)^{2\ell} \frac{\tau^{2\ell-1}}{(2\ell-1)!} \sum_{i=1}^{\ell} |B_i|. \end{aligned}$$

**Lemma 4.6.** For  $1 \leq i \leq \ell$  and  $A_j, B_j$  as defined in appendix A.1, we can bound the following expressions as follows:

$$\begin{aligned} \sum_{i=1}^{\ell} |A_i| &\leq C_A(\ell, \tau) \left( \frac{M}{\alpha} \right)^{2\ell-1}, \\ \sum_{i=1}^{\ell} |B_i| &\leq C_B(\ell, \tau) \left( \frac{M}{\alpha} \right)^{2\ell-2}. \end{aligned}$$

The proofs of Lemmas 4.5 and 4.6 are shown in appendix A.2 and A.3 respectively.

Combining (10) and Lemmas 4.6 and 4.5 we get:

$$\begin{aligned} \|\mu * D_M\|_{L^2(\mathbb{T})} &\leq \left( \frac{\alpha}{M} \right)^{2\ell-1} \left( \|D_M^{(2\ell-1)}\|_{L^2(\mathbb{T})} + \|R_A(\omega)\|_{L^2(\mathbb{T})} + \|R_B(\omega)\|_{L^2(\mathbb{T})} \right) \\ &\leq \sqrt{M+1} (2\pi \alpha)^{2\ell-1} (1 + \tilde{C}(\ell, \tau) 2\pi \alpha). \end{aligned} \quad (11)$$

The proof of the following lemma is in appendix A.5.

**Lemma 4.7.** Let  $u \in \mathbb{C}^{2s}$  be defined by equation  $(\star)$  from Definition 4.4, then

$$\|u\|_2 \geq \tilde{C}_3(\ell, \tau) := \frac{(2\ell-1)!}{4\ell^3 \tau^{2\ell-1}}. \quad (12)$$

Combining (11) and (12) we get:

$$\frac{\|\Phi_M u\|_2}{\|u\|_2} \leq \frac{\sqrt{M+1}(2\pi\alpha)^{2\ell-1}(1+\tilde{C}(\ell,\tau)2\pi\alpha)}{\tilde{C}_3(\ell,\tau)} \leq \hat{C}(\ell,\tau)\sqrt{M+1}(2\pi\alpha)^{2\ell-1}.$$

**Proposition 4.3.** For  $N, s, d \in \mathbb{N}$  and vector  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_s)$  of pairwise distinct real nodes  $\xi_j \in (-\pi, \pi]$ , let  $\Phi_{2N}$  be as in definition 4.3. Then, the following decomposition holds:

$$\tilde{\Phi}_{2N}(\boldsymbol{\eta}) := \frac{1}{\sqrt{2N}} \Phi_{2N} \left( \frac{\boldsymbol{\xi}}{-2\pi} + \frac{1}{2} \right) = \frac{1}{\sqrt{2N}} E_1 \Phi_{2N} \left( \frac{\boldsymbol{\xi}}{-2\pi} \right) E_2 = E_1 U_N(\boldsymbol{\xi}) E_2,$$

where

$$E_1 = \text{diag}(1, e^{-2\pi i \frac{1}{2}}, \dots, e^{-2\pi i M \frac{1}{2}})_{2s \times 2s} \quad \text{and} \quad E_2 = \text{diag}(1, \dots, 1, e^{2\pi i \frac{1}{2}}, \dots, e^{2\pi i \frac{1}{2}})_{2s \times 2s}.$$

Therefore,  $\tilde{\Phi}_{2N}(\boldsymbol{\eta})$  and  $U_N(\boldsymbol{\xi})$  are unitary equivalent and thus have the same singular values.

Finally, by Proposition 4.3 and setting  $M = 2N$  we get:

$$\frac{\boldsymbol{\xi}}{-2\pi} + \frac{1}{2} = \boldsymbol{\omega} \Rightarrow \boldsymbol{\xi} = 2\pi\boldsymbol{\omega} - \pi,$$

and

$$\sigma_{\min}(U_N(\boldsymbol{\xi})) = \sigma_{\min} \left( \tilde{\Phi}_{2N} \left( \frac{\boldsymbol{\xi}}{-2\pi} + \frac{1}{2} \right) \right) = \sigma_{\min}(\tilde{\Phi}_{2N}(\boldsymbol{\omega})) \leq C_2(\ell, \tau)(2\pi\alpha)^{2\ell-1},$$

completing the proof of Theorem 3.2.

## 4.4 Proof of Theorem 3.3

### 4.4.1 Notation

**Definition 4.5** (Pascal-Vandermonde matrix). For  $\mathbf{t} = (t_1, \dots, t_s) \in \mathbb{R}^{1 \times s}$  and  $z_j = e^{it_j}$  let

$$H := H(\mathbf{t}) = \text{diag}(1, \dots, 1, -iz_1, \dots, -iz_s)_{2s \times 2s}, \quad P_N(\mathbf{t}) = U_N(\mathbf{t})H(\mathbf{t}).$$

Every discrete distribution  $\mu = \sum_{j=1}^s a_j \delta_{t_j} + b_j \delta'_{t_j} \in \mathcal{R}$  can be identified with a sparse vector  $x_\mu \in \mathbb{C}_{\mathcal{R}}^{2G} \subset \mathbb{C}^{2G}$ , where  $G = G(\Delta)$  from definition 2.5,  $\|x\|_0 = \#\{i | x_i \neq 0\}$ ,

$$(x_\mu)_i := \begin{cases} a_j & t_j = (i-1)\Delta \in \mathbf{t} \wedge 1 \leq i \leq G \\ -ib_j & t_j = (i-1-G)\Delta \in \mathbf{t} \wedge G+1 \leq i \leq 2G \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

for  $j = 1, \dots, 2s$  and

$$\mathbb{C}_{\mathcal{R}(\Delta, \rho, s, \ell, \tau)}^{2G} := \left\{ x_\nu : \nu \in \mathcal{R}(\Delta, \rho, s, \ell, \tau), x_\nu \in \mathbb{C}^{2G}, \|x_\nu\|_0 \leq 2s \text{ and } x_\nu \text{ as defined in (13)} \right\}.$$

A direct computation shows that for every  $\omega \in \mathbb{R}$

$$\hat{\mu}(\omega) = \sum_{j=1}^s a_j e^{i\omega t_j} - i\omega b_j e^{i\omega t_j}.$$

Thus we can write

$$\begin{pmatrix} \hat{\mu}(0) \\ \hat{\mu}(1) \\ \vdots \\ \hat{\mu}(G-1) \end{pmatrix}_{(G \times 1)} = \frac{1}{\sqrt{G}} (\mathcal{F}_G \quad \mathcal{F}'_G) x_\mu,$$

where  $\mathcal{F}_G = [e^{ikj\Delta}]_{k=0, \dots, G-1}^{j=0, \dots, G-1}$  is the  $G \times G$  Discrete Fourier matrix and  $\mathcal{F}'_G = \text{diag}(0, \dots, G-1)\mathcal{F}_G$ .

**Corollary 4.1.** *Assume that  $2N + 1 \leq G$ , let  $\mathcal{F}_{2N+1}$  be the first  $2N + 1$  rows of the  $G \times G$  DFT matrix and let  $\tilde{\mathcal{F}}_{2N} = \frac{1}{\sqrt{2N}} (\mathcal{F}_{2N+1} \quad \mathcal{F}'_{2N+1})$ . Then,*

$$\tilde{\mathcal{F}}_{2N} x_\mu = P_N v_\mu = U_N w_\mu ,$$

where

$$v_\mu = \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ b_1 \\ \vdots \\ b_s \end{pmatrix}, \quad w_\mu = H v_\mu = \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ -iz_1 b_1 \\ \vdots \\ -iz_s b_s \end{pmatrix} .$$

#### 4.4.2 Proof of the upper bound

As in [5], we choose any  $\varphi$  such that  $\varphi_y \in \{\nu : \|\tilde{\mathcal{F}}_{2N} x_\nu - y\|_2 \leq \epsilon\}$ . Note that  $x_\mu$  satisfies the same constraint  $\|\tilde{\mathcal{F}}_{2N} x_\mu - y\|_2 \leq \epsilon$ , which means that such  $\varphi_y$  exists. Then we have:

$$\mathcal{E}(\mathcal{R}, N, \epsilon) \leq \sup_{\mu \in \mathcal{R}} \sup_{y \in B_\epsilon^N(\mu)} \|\varphi_y - \mu\|_2 .$$

By Lemma 4.7 in [4] there exists  $\Delta_0$  such that for all  $\Delta \leq \Delta_0$  and any  $x_\mu \in \mathbb{C}_{\mathcal{R}(\Delta, \rho, s, \ell, \tau)}^{2G}$ , we have

$$\begin{aligned} \mathbf{t} &:= \text{supp}(\varphi_y - \mu) \in \mathcal{R}(\Delta, \rho', s', \ell', \tau') \\ &\Rightarrow x_{\varphi_y} - x_\mu \in \mathbb{C}_{\mathcal{R}(\Delta, \rho', s', \ell', \tau')}^{2G} \end{aligned}$$

where  $s' \leq 2s$ ,  $\ell' \leq 2\ell$ ,  $\tau' \geq 1$  and  $\rho' = 8sM\tau'\Delta$ . In particular,  $\rho' > 4\tau'\Delta$ , and therefore, by applying Theorem 3.1, we obtain that for all  $N$  satisfying

$$\max \left( \frac{\pi s'}{2sM\tau'\Delta} = \frac{4\pi s'}{\rho}, 4s'^3 \right) \leq N \leq \frac{\pi s'}{\tau'\Delta}$$

and for  $\frac{1}{N\Delta} := \beta = M\tau' \frac{2s}{\pi s'} \geq M$ , we have:

$$\begin{aligned} \frac{2\epsilon}{\|\varphi_y - \mu\|_2} &\geq \frac{\left| \|\tilde{\mathcal{F}}_{2N} x_{\varphi_y} - y\|_2 - \|\tilde{\mathcal{F}}_{2N} x_\mu - y\|_2 \right|}{\|\varphi_y - \mu\|_2} \geq \frac{\|\tilde{\mathcal{F}}_{2N}(x_{\varphi_y} - x_\mu)\|_2}{\|\varphi_y - \mu\|_2} \\ &= \frac{\|P_N(\mathbf{t})(v_{\varphi_y} - v_\mu)\|_2}{\|v_{\varphi_y} - v_\mu\|_2} = \frac{\|U_N(\mathbf{t})(H(\mathbf{t})v_{\varphi_y} - H(\mathbf{t})v_\mu)\|_2}{\|H^{-1}(\mathbf{t})(H(\mathbf{t})v_{\varphi_y} - H(\mathbf{t})v_\mu)\|_2} \\ &\geq \frac{\|U_N(\mathbf{t})(w_{\varphi_y} - w_\mu)\|_2}{\|H^{-1}(\mathbf{t})\|_2 \|w_{\varphi_y} - w_\mu\|_2} \geq \sigma_{\min}(U_N(\mathbf{t})) \\ &\geq C_{s', \ell'}(N\Delta)^{2\ell' - 1} \geq C_{2s, 2\ell}(N\Delta)^{4\ell - 1}. \quad \square \end{aligned}$$

#### 4.4.3 Proof of the lower bound

Pick any  $(\Delta, \rho, 2s, 2\ell, \tau)$ -clustered configuration  $\mathbf{t} = (t_1, \dots, t_{2s})$ . Let  $w \in \mathbb{C}^{4s}$  be a unit norm singular vector of  $U_N(\mathbf{t})$ , put  $v = H^{-1}(\mathbf{t})w$  and define the corresponding  $\mu = \sum_{j=1}^{2s} (v)_{2j-1} \delta_{t_j} + (v)_{2j} \delta'_{t_j} \in \mathcal{R}(\Delta, \rho, 2s, 2\ell, \tau)$  (so in fact according to our previous notation  $w = w_\mu$  and  $v = v_\mu$ ). By this construction, we obtain

$$\sigma := \sigma_{\min}(U_N(\text{supp}(\mu))) = \|U_N w_\mu\|_2 = \|P_N v_\mu\|_2 = \|\tilde{\mathcal{F}}_{2N} x_\mu\|_2 .$$

Write  $\mathbf{t}$  as a disjoint union of two  $(\Delta, \rho, s, \ell, \tau)$ -clustered configurations  $\mathbf{t}_1, \mathbf{t}_2$ , implying that  $\mu = \mu_1 - \mu_2$  where  $\text{supp } \mu_i = \mathbf{t}_i$  and  $\mu_i \in \mathcal{R}(\Delta, \rho, s, \ell, \tau)$  for  $i = 1, 2$ . Let  $x_i = \frac{\epsilon}{\sigma} x_{\mu_i} \in \mathbb{C}_{\mathcal{R}(\Delta, \rho, s, \ell, \tau)}^{2G}$ , for  $i = 1, 2$ , so that  $\frac{\epsilon}{\sigma} x_{\mu} = x_1 - x_2$ .

Now suppose we are given the data:

$$y = \tilde{\mathcal{F}}_{2N} x_1 = \tilde{\mathcal{F}}_{2N} x_2 + \tilde{\mathcal{F}}_{2N}(x_1 - x_2).$$

Let  $e := \tilde{\mathcal{F}}_{2N}(x_1 - x_2) \in \mathbb{C}^{2N+1}$ . The previous equations imply:

$$\|e\|_2 = \|\tilde{\mathcal{F}}_{2N}(x_1 - x_2)\|_2 = \frac{\epsilon}{\sigma} \|\tilde{\mathcal{F}}_{2N} x_{\mu}\|_2 = \epsilon.$$

For an arbitrary  $\varphi$  we have

$$\begin{aligned} \frac{\epsilon}{\sigma} &= \frac{\epsilon}{\sigma} \|w_{\mu}\|_2 = \frac{\epsilon}{\sigma} \|x_{\mu}\|_2 \\ &= \|x_1 - x_2\|_2 \\ &\leq \|x_1 - x_{\varphi_y}\|_2 + \|x_2 - x_{\varphi_y}\|_2 \\ &\leq 2 \max_{k=1,2} \|x_k - x_{\varphi_y}\|_2 \end{aligned}$$

and so by definition of  $\mathcal{E}$  and Theorem 3.2 we conclude that

$$\Rightarrow \mathcal{E}(\mathcal{R}, N, \epsilon) \geq \inf_{\varphi \in \mathcal{A}} \max_{k=1,2} \|x_{\varphi_y} - x_k\|_2 \geq \frac{\epsilon}{2\sigma} \geq \frac{\epsilon}{2C_{\tau,2\ell}(N\Delta)^{4\ell-1}}. \quad \square$$

## 5 Numerical experiments

In order to validate the bounds of Theorems 3.1 and 3.2, we computed  $\sigma_{\min}(U_N)$  for varying values of  $\Delta, N, \ell, s$  and the actual clustering configurations. As before, we put  $SRF := \frac{1}{N\Delta}$ . We checked two clustering scenarios:

1. Figure 1a - A single equispaced cluster of size  $\ell$  in  $[-\frac{\pi}{2}, -\frac{\pi}{2} + \ell\Delta]$  with the rest of the nodes equally spaced and maximally separated in  $(-\frac{\pi}{2} + \ell\Delta, \frac{\pi}{2}]$ .
2. Figure 1b - A multi-cluster configuration with the first equispaced cluster of size  $\ell_1$  in  $[-\frac{\pi}{2}, -\frac{\pi}{2} + \ell_1\Delta]$  and the second equispaced cluster of size  $\ell_2$  in  $[\frac{\pi}{2} - \ell_2\Delta, \frac{\pi}{2}]$  with the rest of the nodes equally spaced and maximally separated in  $(-\frac{\pi}{2} + \ell_1\Delta, \frac{\pi}{2} - \ell_2\Delta)$ .

We also show in figure 2 that the vector  $u$  defined in  $(\star)$  is indeed an approximate minimal singular vector, by plotting the Rayleigh quotient  $\frac{\|\Phi_M(\omega)u\|_2}{\|u\|_2}$  versus the minimal singular value  $\sigma_{\min}(\Phi_M(\omega))$ , where  $\Phi_M$  is the confluent Vandermonde matrix as in Definition 4.3 and  $\omega$  is a single-cluster equispaced configuration.

Finally, in order to validate the bounds of Theorem 3.3, we computed the  $\ell^2$  min-max error  $\mathcal{E}$  as in Definition 2.10 and also the  $\ell^2$  errors of estimating the nodes  $\mathcal{E}_{\xi}$ , and the coefficients  $\mathcal{E}_a, \mathcal{E}_b$  of the worst-case discrete distribution  $\mu$  defined by (7) assuming  $s = \ell$ . We used the ESPRIT (Estimation of Signal Parameters via Rotation Invariance Techniques) [21] method for recovering the nodes  $\{t_j\}_{j=1}^s$  (see more about this method in appendix B). ESPRIT is considered to be one of the best performing subspace methods for estimating parameters of model (1) with white Gaussian noise. Originally developed in the context of frequency estimation [22], it has been generalized to the full model (5) in [14]. Recently it has been shown that if the noise level  $\epsilon$  in the measurements (2) is sufficiently small, the error committed by ESPRIT for estimating the nodes of the simple model (1) is nearly min-max [23]. Consequently, we conjecture the same near-optimal behaviour in the model (4). In order to recover the coefficients  $a_j, b_j$  we solve a linear system of equations by the Least Squares method:

$$\min \|U_N(\tilde{\xi})v_{\mu} - y\|_2,$$

where  $\tilde{\xi}$  are the recovered nodes. Note that we prove the theoretical bound to the on-grid model however the ESPRIT algorithm recovers the nodes without taking the grid assumption into account. We have checked two cases:

1. Figure 3a - A single equispaced cluster of size  $s = \ell = 2$  with error  $\epsilon = 10^{-12}$ .
2. Figure 3b - A single equispaced cluster of size  $s = \ell = 3$  with error  $\epsilon = 10^{-12}$ .

Our results suggest that the ESPRIT method might indeed be optimal, meaning that it attains the min-max error bounds we established in Theorem 3.3 for the recovered parameters of signal (4). Note that all figures are in logarithmic scale.

The code for the above experiments is available at <https://github.com/Gnflu/SR-of-conVan-sys.git>

## A Computations for Theorem 3.2

### A.1 Finite difference coefficients

We seek approximation of the form:

$$D_M^{(2\ell-1)}(\omega) \approx \sum_{i=1}^{\ell} A_i D_M(x_i) + B_i D'_M(x_i) = S_{A,B}(\omega)$$

where

$$x_i = \omega - \tau_i \Delta, \quad \Delta = \frac{\alpha}{M},$$

and

$$S_{A,B}(\omega) = \sum_{i=1}^{\ell} A_i D_M(x_i) + B_i D'_M(x_i) = \sum_{i=1}^{\ell} A_i D_M(\omega + x_i - \omega) + B_i D'_M(\omega + x_i - \omega).$$

Let  $h_i := x_i - \omega = -\tau_i \Delta = \tau_i h$ ,  $h = -\Delta$ . Then by Taylor expansion of  $D_M(\omega + h_i)$  and using the integral form of the remainder we have:

$$\begin{aligned} S_{A,B}(\omega) &= \sum_{i=1}^{\ell} A_i \left( \sum_{k=0}^{2\ell-1} \frac{D_M^{(k)}(\omega)}{k!} h_i^k + \int_{\omega}^{\omega+h_i} \frac{D_M^{(2\ell)}(t)}{(2\ell-1)!} (\omega + h_i - t)^{2\ell-1} dt \right) \\ &+ B_i \left( \sum_{k=0}^{2\ell-2} \frac{D_M^{(k+1)}(\omega)}{k!} h_i^k + \int_{\omega}^{\omega+h_i} \frac{D_M^{(2\ell)}(t)}{(2\ell-2)!} (\omega + h_i - t)^{2\ell-2} dt \right) \end{aligned}$$

By the change of variable  $t = \omega + h_i r$  we have  $dt = h_i dr$  and therefore

$$\begin{aligned} S_{A,B}(\omega) &= \sum_{k=0}^{2\ell-1} \frac{D_M^{(k)}(\omega)}{k!} \left( \sum_{i=1}^{\ell} A_i h_i^k \right) + \sum_{k=0}^{2\ell-2} \frac{D_M^{(k+1)}(\omega)}{k!} \left( \sum_{i=1}^{\ell} B_i h_i^k \right) \\ &+ \frac{1}{(2\ell-1)!} \sum_{i=1}^{\ell} A_i \int_0^1 D_M^{(2\ell)}(\omega + h_i r) (h_i - h_i r)^{2\ell-1} h_i dr \\ &+ \frac{1}{(2\ell-2)!} \sum_{i=1}^{\ell} B_i \int_0^1 D_M^{(2\ell)}(\omega + h_i r) (h_i - h_i r)^{2\ell-2} h_i dr \\ &= P(\omega) + R_A(\omega) + R_B(\omega), \end{aligned}$$



where

$$\begin{aligned}
P(\omega) &= \sum_{k=0}^{2\ell-1} \frac{D_M^{(k)}(\omega)}{k!} \left( \sum_{i=1}^{\ell} A_i h_i^k \right) + \sum_{k=0}^{2\ell-2} \frac{D_M^{(k+1)}(\omega)}{k!} \left( \sum_{i=1}^{\ell} B_i h_i^k \right), \\
R_A(\omega) &= \frac{1}{(2\ell-1)!} \sum_{i=1}^{\ell} A_i \int_0^1 D_M^{(2\ell)}(\omega + h_i r) (h_i - h_i r)^{2\ell-1} h_i dr, \\
R_B(\omega) &= \frac{1}{(2\ell-2)!} \sum_{i=1}^{\ell} B_i \int_0^1 D_M^{(2\ell)}(\omega + h_i r) (h_i - h_i r)^{2\ell-2} h_i dr.
\end{aligned}$$

We seek  $A_1, \dots, A_\ell$  and  $B_1, \dots, B_\ell$  so that  $P(\omega) \equiv D_M^{(2\ell-1)}(\omega)$ , thus the following equations should be fulfilled:

1.

$$\sum_{i=1}^{\ell} A_i = 0$$

2.

$$\sum_{i=1}^{\ell} \left( A_i + \frac{k B_i}{h_i} \right) h_i^k = 0 \quad k = 1, \dots, 2\ell - 2$$

3.

$$\sum_{i=1}^{\ell} \left( A_i + \frac{(2\ell-1) B_i}{h_i} \right) h_i^{2\ell-1} = (2\ell-1)!$$

This is equivalent to solving the following linear system of equations:

$$U_{2\ell} \begin{pmatrix} A_1 \\ \vdots \\ A_\ell \\ B_1 \\ \vdots \\ B_\ell \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ (2\ell-1)! \end{pmatrix},$$

where

$$U_{2\ell} = \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ h_1 & \dots & h_\ell & 1 & \dots & 1 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ h_1^{2\ell-1} & \dots & h_\ell^{2\ell-1} & (2\ell-1)h_1^{2\ell-2} & \dots & (2\ell-1)h_\ell^{2\ell-2} \end{pmatrix}.$$

Thus  $A_j, B_j$  are given by:

$$\begin{pmatrix} A_1 \\ \vdots \\ A_\ell \\ B_1 \\ \vdots \\ B_\ell \end{pmatrix} = U_{2\ell}^{-1} \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ (2\ell-1)! \end{pmatrix}. \quad (\star\star)$$

In particular, if  $U_{2\ell}^{-1} = \begin{pmatrix} V \\ W \end{pmatrix}$  then  $A_j = (2\ell-1)!v_{j,2\ell}$  and  $B_j = (2\ell-1)!w_{j,2\ell}$ , where  $V, W \in \mathbb{C}^{\ell \times 2\ell}$  and  $v_{i,j}, w_{i,j}$  denote the  $(i,j)$ th entry of  $V, W$  respectively.

## A.2 Proof of Lemma 4.5

Let

$$h_* := \arg \max_{h_i} \left| \int_0^1 D_M^{(2\ell)}(\omega + h_i r) (h_i - h_i r)^{2\ell-1} h_i dr \right|$$

Using the Cauchy-Schwartz inequality we have

$$\begin{aligned} \|R_A(\omega)\|_{L^2(\mathbb{T})}^2 &= \frac{1}{(2\ell-1)!^2} \left\| \sum_{i=1}^{\ell} A_i \int_0^1 D_M^{(2\ell)}(\omega + h_i r) (h_i - h_i r)^{2\ell-1} h_i dr \right\|_{L^2(\mathbb{T})}^2 \\ &= \frac{1}{(2\ell-1)!^2} \int_0^1 \left| \sum_{i=1}^{\ell} A_i \int_0^1 D_M^{(2\ell)}(\omega + h_i r) (h_i - h_i r)^{2\ell-1} h_i dr \right|^2 d\omega \\ &\leq \frac{1}{(2\ell-1)!^2} \left( \sum_{i=1}^{\ell} |A_i| \right)^2 \int_0^1 \left| \int_0^1 D_M^{(2\ell)}(\omega + h_* r) (h_* - h_* r)^{2\ell-1} h_* dr \right|^2 d\omega \\ &\leq \frac{1}{(2\ell-1)!^2} \left( \sum_{i=1}^{\ell} |A_i| \right)^2 \int_0^1 \left( \int_0^1 |D_M^{(2\ell)}(\omega + h_* r)|^2 dr \int_0^1 |h_* - h_* r|^{4\ell-2} h_*^2 dr \right) d\omega \\ &\leq \frac{1}{(2\ell-1)!^2} \left( \sum_{i=1}^{\ell} |A_i| \right)^2 \int_0^1 \left( \int_0^1 |D_M^{(2\ell)}(\omega + h_* r)|^2 dr |\tau \Delta|^{4\ell} \right) d\omega \\ &\leq \frac{1}{(2\ell-1)!^2} \left( \sum_{i=1}^{\ell} |A_i| \right)^2 |\tau \Delta|^{4\ell} \int_0^1 \left( \int_0^1 |D_M^{(2\ell)}(\omega + h_* r)|^2 d\omega \right) dr \\ &\leq \frac{1}{(2\ell-1)!^2} \left( \sum_{i=1}^{\ell} |A_i| \right)^2 |\tau \Delta|^{4\ell} \|D_M^{(2\ell)}\|_{L^2(\mathbb{T})}^2 \\ &\leq \frac{\tau^{4\ell}}{(2\ell-1)!^2} \left( \sum_{i=1}^{\ell} |A_i| \right)^2 \Delta^{4\ell} (M+1) (2\pi M)^{4\ell} \end{aligned}$$

Similarly we get that:

$$\|R_B(\omega)\|_{L^2(\mathbb{T})}^2 \leq \frac{\tau^{4\ell-2}}{(2\ell-1)!^2} \left( \sum_{i=1}^{\ell} |B_i| \right)^2 \Delta^{4\ell-2} (M+1) (2\pi M)^{4\ell}$$

## A.3 Proof of Lemma 4.6

Recall the definitions of  $v_{i,j}$  and  $w_{i,j}$  from Appendix A.1. From expressions (3.10) and (3.12) evaluated in Gautchi's paper [24], and using that  $M \leq \frac{1}{\Delta}$  we have:

1. On one hand

$$\begin{aligned}
\sum_{\mu=1}^{2\ell} |v_{n,\mu}| &\leq \left( \left| 1 + 2h_n \sum_{\nu \neq n} \frac{1}{(h_n - h_\nu)} \right| + 2 \left| \sum_{\nu \neq n} \frac{1}{(h_n - h_\nu)} \right| \right) \prod_{\nu \neq n} \left( \frac{1 + |h_\nu|}{|h_n - h_\nu|} \right)^2 \\
&\leq \left( \left| 1 + 2\tau_n h \sum_{\nu \neq n} \frac{1}{(\tau_n - \tau_\nu)h} \right| + 2 \left| \sum_{\nu \neq n} \frac{1}{(\tau_n - \tau_\nu)h} \right| \right) \prod_{\nu \neq n} \left( \frac{1 + |\tau_\nu h|}{|(\tau_n - \tau_\nu)h|} \right)^2 \\
&\leq (|1 + 2\tau(\ell - 1)| + \left| \frac{2\ell}{h} \right|) \left( \frac{1 + \tau|h|}{|h|} \right)^{2\ell-2} \\
&= \left( 1 + 2\tau\ell - 2\tau + 2\ell \frac{M}{\alpha} \right) \left( \frac{M}{\alpha} + \tau \right)^{2\ell-2} \\
&\leq 2\ell \left( \frac{M}{\alpha} + \tau \right)^{2\ell-1} \\
&\leq C_v(\ell, \tau) \left( \frac{M}{\alpha} \right)^{2\ell-1}
\end{aligned}$$

2. On the other hand,

$$\begin{aligned}
\sum_{\mu=1}^{2\ell} |w_{n,\mu}| &\leq (1 + |h_n|) \prod_{\nu \neq n} \left( \frac{1 + |h_\nu|}{|h_n - h_\nu|} \right)^2 \leq (1 + \tau|h|) \left( \frac{1 + \tau|h|}{|h|} \right)^{2\ell-2} \\
&= \left( 1 + \tau \frac{\alpha}{M} \right) \left( \frac{M}{\alpha} + \tau \right)^{2\ell-2} \leq 2 \left( \frac{M}{\alpha} + \tau \right)^{2\ell-2} \\
&\leq C_w(\ell, \tau) \left( \frac{M}{\alpha} \right)^{2\ell-2}
\end{aligned}$$

Now we can evaluate the following expressions:

1.

$$\begin{aligned}
\sum_{i=0}^{\ell} |A_i| &= (2\ell - 1)! \sum_{i=1}^{\ell} |v_{i,2\ell}| \leq (2\ell - 1)! \sum_{i=1}^{\ell} \sum_{\mu=1}^{2\ell} |v_{i,\mu}| \\
&\leq \ell(2\ell - 1)! C_v(\ell, \tau) \left( \frac{M}{\alpha} \right)^{2\ell-1} \\
&= C_A(\ell, \tau) \left( \frac{M}{\alpha} \right)^{2\ell-1}
\end{aligned}$$

2.

$$\begin{aligned}
\sum_{i=0}^{\ell} |B_i| &= (2\ell - 1)! \sum_{i=1}^{\ell} |w_{i,2\ell}| \leq (2\ell - 1)! \sum_{i=1}^{\ell} \sum_{\mu=1}^{2\ell} |w_{i,\mu}| \\
&\leq \ell(2\ell - 1)! C_w(\ell, \tau) \left( \frac{M}{\alpha} \right)^{2\ell-2} \\
&= C_B(\ell, \tau) \left( \frac{M}{\alpha} \right)^{2\ell-2}
\end{aligned}$$

#### A.4 Proof of Lemma 4.4

For any tempered distribution  $\mu$  supported in  $\mathbb{T}$ , we will show that the following is true:

$$\sum_{m=0}^M |\hat{\mu}(m)|^2 = \|\mu * D_M\|_{L^2(\mathbb{T})}^2$$



$$\begin{aligned}
x &= \begin{pmatrix} I & 0 \\ 0 & hI \end{pmatrix} L^{-1} D^{-1} b = \begin{pmatrix} I & 0 \\ 0 & hI \end{pmatrix} L^{-1} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ (2\ell-1)!h^{1-2\ell} \end{pmatrix} = \begin{pmatrix} (L^{-1})_{1,2\ell}(2\ell-1)!h^{1-2\ell} \\ \vdots \\ (L^{-1})_{\ell,2\ell}(2\ell-1)!h^{1-2\ell} \\ (L^{-1})_{\ell+1,2\ell}(2\ell-1)!h^{2-2\ell} \\ \vdots \\ (L^{-1})_{2\ell,2\ell}(2\ell-1)!h^{2-2\ell} \end{pmatrix} \\
\|x\|_2^2 &= \sum_{i=1}^{\ell} (L^{-1})_{i,2\ell}^2 (2\ell-1)!^2 h^{2(1-2\ell)} + \sum_{i=1}^{\ell} (L^{-1})_{\ell+i,2\ell}^2 (2\ell-1)!^2 h^{2(2-2\ell)} \geq \sum_{i=1}^{2\ell} (L^{-1})_{i,2\ell}^2 (2\ell-1)!^2 \Delta^{2-4\ell} \\
\sum_{i=1}^{2\ell} (L^{-1})_{i,2\ell}^2 &= \|L^{-1}(0, \dots, 0, 1)^T\|_2^2 \geq \min_{\|w\|_2=1} \|L^{-1}w\|_2^2 = \sigma_{\min}^2(L^{-1}) = \frac{1}{\sigma_{\max}^2(L)} = \frac{1}{\|L\|_2^2} \geq \frac{1}{\text{rank}(L)\|L\|_{\infty}^2} \\
\|L\|_{\infty} &\leq \sum_{i=1}^{\ell} \tau_i^{2\ell-1} + \sum_{i=1}^{\ell} (2\ell-1)\tau_i^{2\ell-2} \leq \ell(\tau^{2\ell-1} + (2\ell-1)\tau^{2\ell-2}) \leq 2\ell^2\tau^{2\ell-1} \\
\|x\|_2^2 &\geq \frac{(2\ell-1)!^2 \Delta^{2-4\ell}}{2\ell(2\ell^2\tau^{2\ell-1})^2} \Rightarrow \|x\|_2 \geq \frac{(2\ell-1)! \Delta^{1-2\ell}}{\sqrt{8\ell^5}\tau^{2\ell-1}} = C_3(\ell, \tau) \Delta^{1-2\ell}
\end{aligned}$$

Finally, for  $u$  defined in  $(\star)$ , we have

$$\begin{aligned}
u &= \Delta^{2\ell-1} \begin{pmatrix} I & 0 \\ 0 & \mathbf{Z} \end{pmatrix} X = \Delta^{2\ell-1} \mathbf{Z}x, \quad \mathbf{Z} = \text{diag}(z_1, \dots, z_{\ell}) \\
\|u\|_2 &= \Delta^{2\ell-1} \|\mathbf{Z}x\|_2 \|\mathbf{Z}^{-1}\|_2 \|\mathbf{Z}^{-1}\|_2^{-1} \\
&\geq \Delta^{2\ell-1} \|\mathbf{Z}\mathbf{Z}^{-1}x\|_2 \|\mathbf{Z}^{-1}\|_2^{-1} = \Delta^{2\ell-1} \|x\|_2 \|\mathbf{Z}^{-1}\|_2^{-1} \\
&\geq \frac{1}{\sqrt{2\ell}} \Delta^{2\ell-1} \Delta^{1-2\ell} C_3(\ell, \tau) = \tilde{C}_3(\ell, \tau)
\end{aligned}$$

We used in last inequality the following property:

$$\|\mathbf{Z}\|_2 \leq \sqrt{\text{rank}(\mathbf{Z})} \|\mathbf{Z}\|_{\infty}$$

## B ESPRIT Method

We provide the description of the matrix for completeness, see e.g. [12], [14].

**Definition B.1** (Hankel Matrix). *Let  $\mu(x) = \sum_{j=1}^s a_j \delta(x-t_j) + b_j \delta'(x-t_j)$ ,  $a_j, b_j \in \mathbb{C}$  and  $\mathbf{t} = (t_1, \dots, t_s)$ ,  $t_j \in \mathbb{T}$ , thus*

$$m_k := \hat{\mu}(k) = \sum_{j=1}^s a_j e^{-2\pi i k t_j} + 2\pi i k b_j e^{-2\pi i k t_j} = \sum_{j=1}^s a_j z_j^k + 2\pi i k b_j z_j^k$$

Then we define the  $C \times 2C$  Hankel matrix as follows:

$$H_C := \begin{pmatrix} m_0 & m_1 & \dots & m_{C-1} \\ m_1 & m_2 & \dots & m_C \\ \vdots & \vdots & \vdots & \vdots \\ m_{C-1} & m_C & \dots & m_{2C-2} \end{pmatrix}$$

where  $C := 2s$  (number of unknown coefficients).

The ESPRIT (and other subspace methods) relies on the following observations:

1. The range (column space) of both the data matrix  $H_C$  (B.1) and the confluent Vandermonde matrix  $\Phi := \Phi_{2C-1}$  2.1 are the same, namely  $H_C$  admits the following factorization:

$$H_C = \Phi B \Phi^T,$$

where  $B := \text{diag}(a_1, \dots, a_s, b_1, \dots, b_s)$ .

2. The matrix  $\Phi$  has the so-called rotational invariance property [14]:

$$\Phi^\uparrow = \Phi^\downarrow J$$

where  $\Phi^\uparrow$  denotes  $\Phi$  without the first row,  $\Phi^\downarrow$  denotes  $\Phi$  without the last row, and  $J$  is a block diagonal matrix whose  $i^{\text{th}}$  block is the  $2 \times 2$  Jordan block with the node  $z_i$  on the diagonal.

Suppose we know  $\Phi$ ; then the matrix  $J$  could be found by

$$J = \Phi_{\downarrow}^{\#} \Phi^\uparrow$$

(where  $\#$  denotes the Moore–Penrose pseudoinverse), and then the nodes  $z_j$  could be recovered as the eigenvalues of  $J$ .

Unfortunately,  $\Phi$  is unknown in advance, but suppose we had at our disposal a matrix  $W$  whose column space was identical to that of  $\Phi$ . In that case, we would have  $W = \Phi G$  for an invertible  $G$ , and consequently

$$W^\uparrow = W^\downarrow \Psi,$$

where

$$\Psi = G^{-1} J G,$$

which means that the eigenvalues of  $\Psi$  are also  $\{z_j\}$ . Such a matrix  $W$  can be obtained, for example, from the singular value decomposition (SVD) of the data matrix/covariance matrix. To summarize, the ESPRIT method for estimating  $\{z_j\}$ , as used in our experiments below, is as follows.

---

**Algorithm 1** ESPRIT method for recovering the nodes  $\{z_j\}$

---

**Require:** A rectangular  $C \times 2C$  Hankel matrix  $H_C$  built from the measurements.

**Ensure:** Recovered nodes  $\{z_j\}$ .

- 1: Compute the SVD  $H_C = W \Sigma V^T$
  - 2: Calculate  $\Psi = W_{\downarrow}^{\#} W^\uparrow$
  - 3: Set  $\{z_j\}$  to be the eigenvalues of  $\Psi$  with appropriate multiplicities (use, e.g., arithmetic means to estimate multiple nodes which are scattered by the noise).
- 

## References

- [1] D. Donoho, “Superresolution via sparsity constraints,” *SIAM Journal on Mathematical Analysis*, vol. 23, no. 5, pp. 1309–1331, 1992.
- [2] D. Batenkov, B. Diederichs, G. Goldman, and Y. Yomdin, “The spectral properties of Vandermonde matrices with clustered nodes,” *Linear Algebra and its Applications*, vol. 609, pp. 37–72, Jan. 2021, ISSN: 0024-3795. DOI: 10.1016/j.laa.2020.08.034.
- [3] D. Batenkov, G. Goldman, and Y. Yomdin, “Super-resolution of near-colliding point sources,” *Information and Inference: A Journal of the IMA*, vol. 10, no. 2, pp. 515–572, Jun. 2021. DOI: 10.1093/imaiai/iaaa005.

- [4] D. Batenkov, L. Demanet, G. Goldman, and Y. Yomdin, “Conditioning of Partial Nonuniform Fourier Matrices with Clustered Nodes,” *SIAM Journal on Matrix Analysis and Applications*, vol. 44, no. 1, pp. 199–220, Jan. 2020, ISSN: 0895-4798. DOI: 10/ggjwzb.
- [5] W. Li and W. Liao, “Stable super-resolution limit and smallest singular value of restricted Fourier matrices,” *Applied and Computational Harmonic Analysis*, vol. 51, pp. 118–156, 2020, ISSN: 1063-5203. DOI: 10.1016/j.acha.2020.10.004.
- [6] E. J. Candès and C. Fernandez-Granda, “Towards a Mathematical Theory of Super-resolution,” *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, Jun. 2014, ISSN: 1097-0312. DOI: 10.1002/cpa.21455.
- [7] L. Demanet and N. Nguyen, “The recoverability limit for superresolution via sparsity,” *arXiv preprint arXiv:1502.01385*, 2015. arXiv: 1502.01385.
- [8] M. Hockmann and S. Kunis, “Sparse super resolution is Lipschitz continuous,” *arXiv:2108.11925 [cs, math]*, Aug. 2021. arXiv: 2108.11925 [cs, math].
- [9] P. Liu and H. Zhang, “A Theory of Computational Resolution Limit for Line Spectral Estimation,” *IEEE Transactions on Information Theory*, vol. 67, no. 7, pp. 4812–4827, Jul. 2021, ISSN: 1557-9654. DOI: 10.1109/TIT.2021.3075149.
- [10] M. Petz, G. Plonka, and N. Derevianko, “Exact reconstruction of sparse non-harmonic signals from their Fourier coefficients,” *Sampling Theory, Signal Processing, and Data Analysis*, vol. 19, no. 1, p. 7, May 2021, ISSN: 2730-5724. DOI: 10.1007/s43670-021-00007-1.
- [11] A. Cuyt and W. Lee, “How to get high resolution results from sparse and coarsely sampled data,” *Applied and Computational Harmonic Analysis*, Oct. 2018, ISSN: 1063-5203. DOI: 10/ggb5cv.
- [12] D. Batenkov and Y. Yomdin, “On the accuracy of solving confluent Prony systems,” *SIAM J. Appl. Math.*, vol. 73, no. 1, pp. 134–154, 2013. DOI: 10.1137/110836584.
- [13] D. Batenkov, “Stability and super-resolution of generalized spike recovery,” *Applied and Computational Harmonic Analysis*, vol. 45, no. 2, pp. 299–323, Sep. 2018, ISSN: 1063-5203. DOI: 10.1016/j.acha.2016.09.004.
- [14] R. Badeau, G. Richard, and B. David, “High-resolution spectral analysis of mixtures of complex exponentials modulated by polynomials,” *IEEE transactions on signal processing*, vol. 54.4, pp. 1341–1350, 2006.
- [15] D. Batenkov and Y. Yomdin, “Algebraic Fourier reconstruction of piecewise smooth functions,” *Mathematics of Computation*, vol. 81, pp. 277–318, 2012. DOI: 10.1090/S0025-5718-2011-02539-1.
- [16] D. Batenkov, “Complete algebraic reconstruction of piecewise-smooth functions from Fourier data,” *Mathematics of Computation*, vol. 84, no. 295, pp. 2329–2350, 2015, ISSN: 0025-5718, 1088-6842. DOI: 10.1090/S0025-5718-2015-02948-2.
- [17] D. Batenkov and G. Goldman, “Single-exponential bounds for the smallest singular value of Vandermonde matrices in the sub-Rayleigh regime,” in *Applied and Computational Harmonic Analysis*, vol. 55, Nov. 2021, ISSN: 1063-5203. DOI: 10.1016/j.acha.2021.07.003. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1063520321000609> (visited on 08/05/2021).
- [18] W. Gautschi, “On inverses of vandermonde and confluent vandermonde matrices,” *Numerische Mathematik*, vol. 4.1, pp. 117–123, 1962. [Online]. Available: <https://doi.org/10.1007/BF01386302>.
- [19] Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, 2nd ed. New York: Cambridge University Press, 2013.
- [20] S. N. Bernstein, “Sur l’ordre de la meilleure approximation des fonctions continues par les polynômes de degré donné,” *Mémoires publiés par la Classe des Sciences de l’Académie de Belgique*, vol. 4, 1912.
- [21] T. Kailath and Richard H. Roy III, “Esprit—estimation of signal parameters via rotational invariance techniques,” *Optical Engineering*, vol. 29.4, pp. 296–313, 1990.

- [22] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, N.J. : Pearson/Prentice Hall, 2005.
- [23] W. Li, W. Liao, and A. Fannjiang, “Super-resolution limit of the esprit algorithm,” *IEEE transactions on information theory*, vol. 66.7, pp. 4593–4608, 2020.
- [24] W. Gautschi, “On inverses of vandermonde and confluent vandermonde matrices ii,” *Numerische Mathematik*, vol. 5, pp. 425–430, 1963.



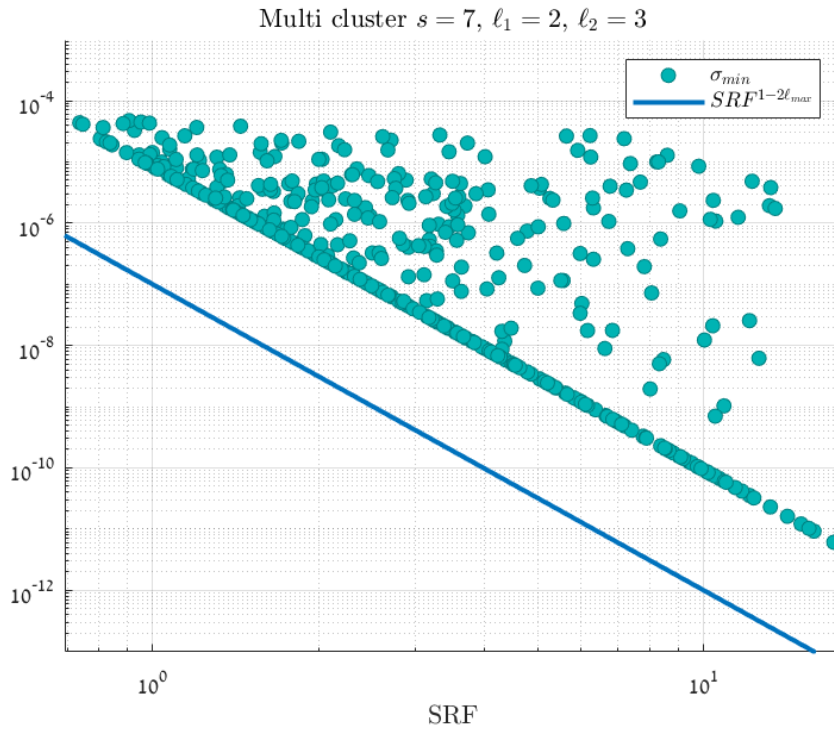
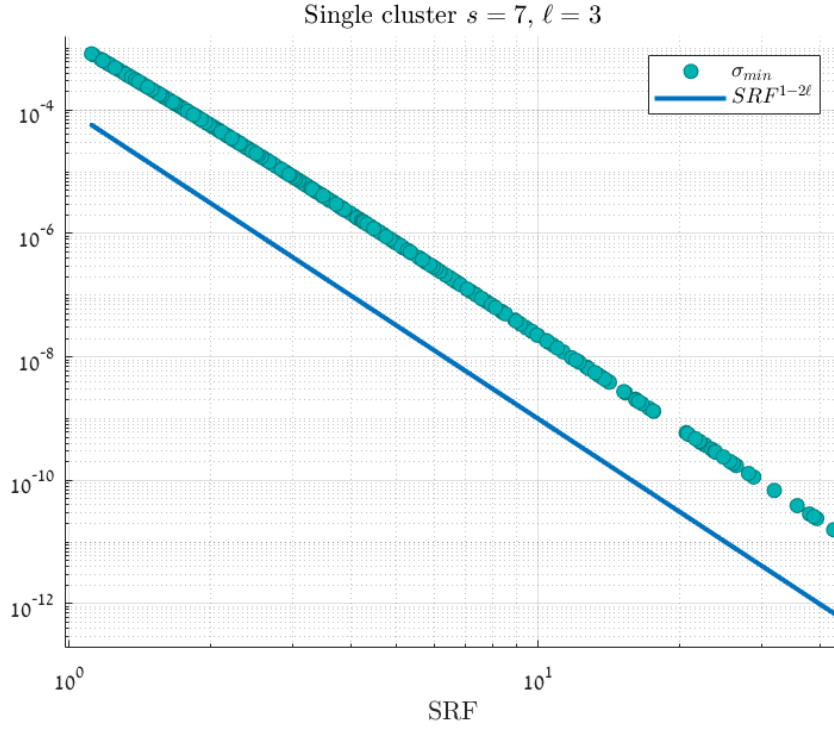


Figure 1: Decay rate of  $\sigma_{\min}$  as a function of  $SRF$ . Results of  $n = 1000$  random experiments with randomly chosen  $\Delta, N$  are plotted versus the theoretical bound  $SRF^{1-2\ell}$ .

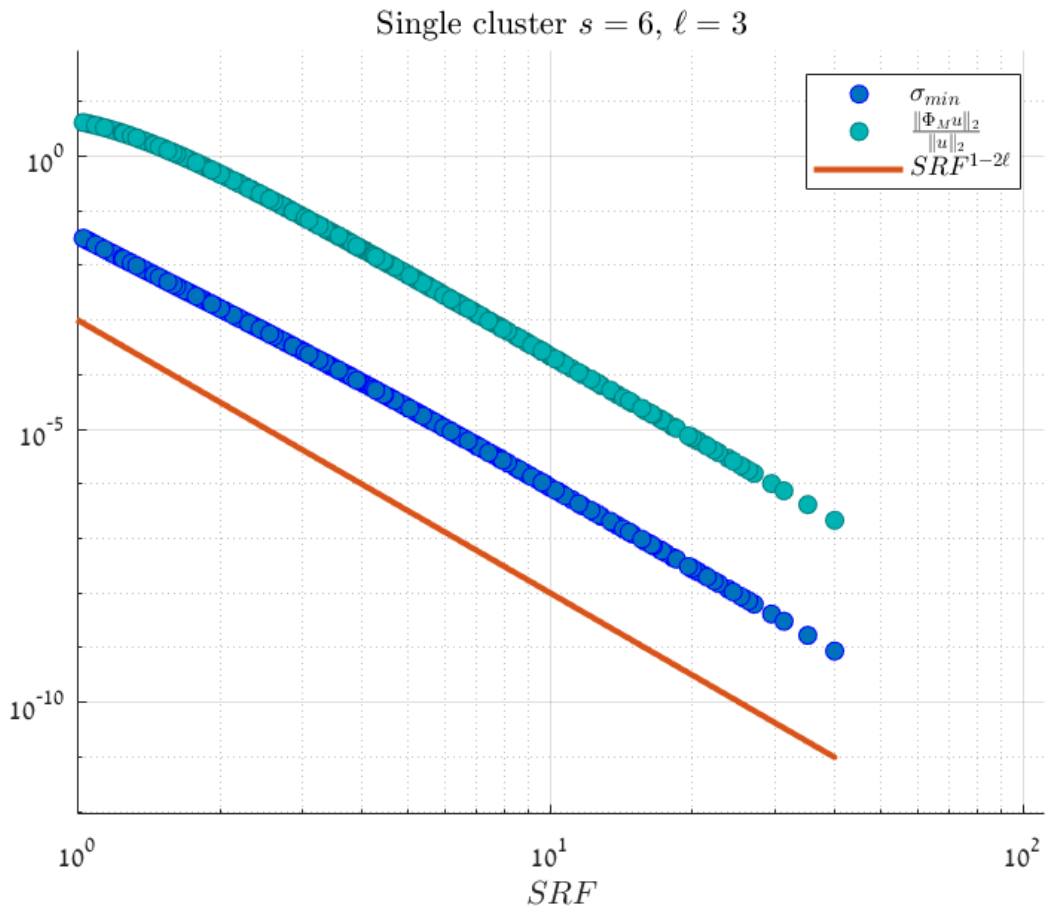


Figure 2: The Rayleigh quotient of the vector  $u$  defined in  $(\star)$  versus the minimal singular value of  $\Phi_M$ . We can see that they scale the same and differ by a constant.

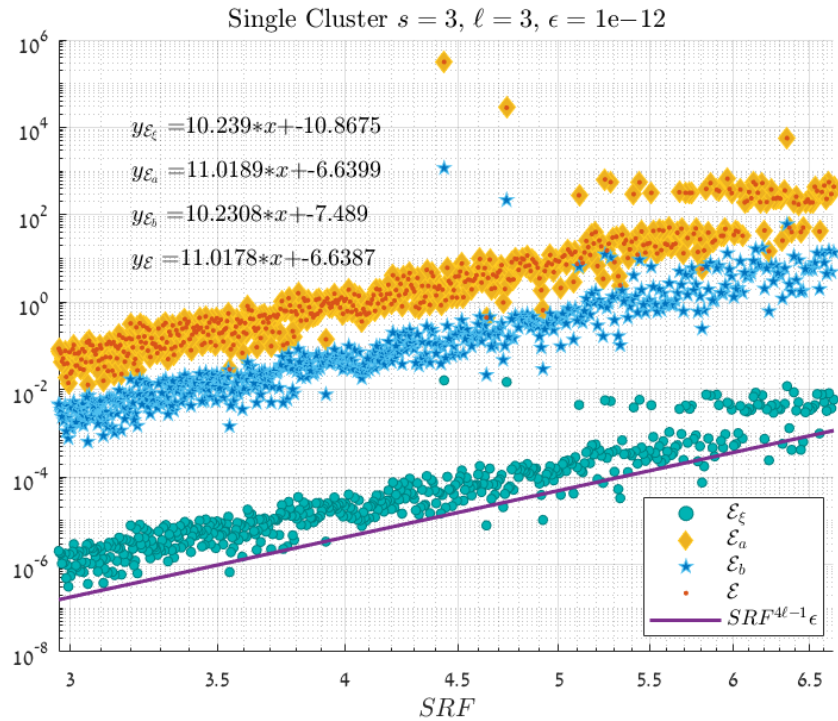
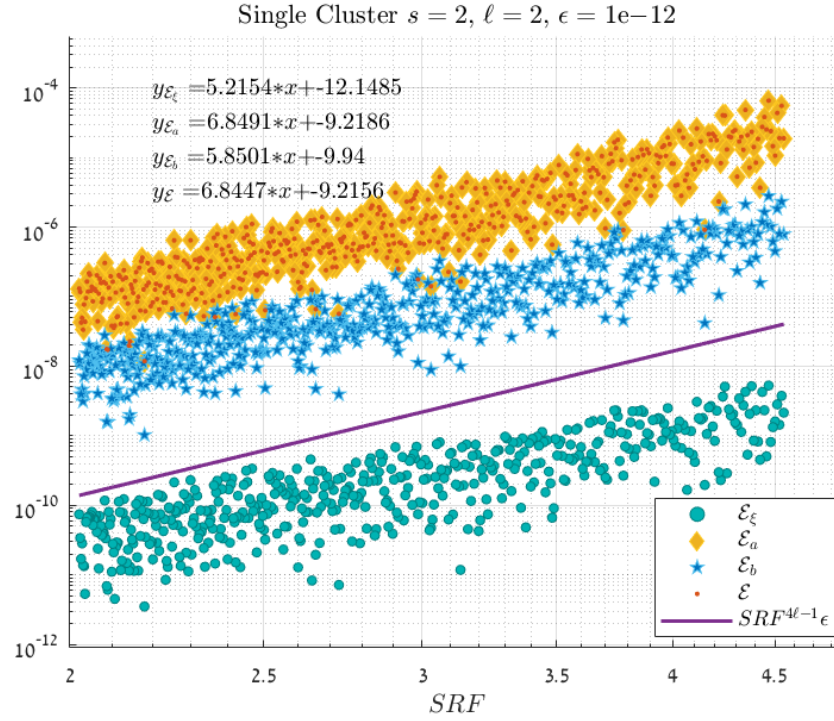


Figure 3: Accuracy of ESPRIT. Results of  $n=500$  random experiments with randomly chosen  $\Delta$  and fixed  $N$  are plotted versus the theoretical bound  $SRF^{4\ell-1}\epsilon$ .