

# Stability of partial Fourier matrices with clustered nodes\*

Dmitry Batenkov<sup>1</sup>, Laurent Demanet<sup>1</sup>, Gil Goldman<sup>2</sup>, and Yosef Yomdin<sup>2</sup>

<sup>1</sup>Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>2</sup>Department of Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel  
 {*batenkov,ldemanet*}@mit.edu, {*gil.goldman,yosef.yomdin*}@weizmann.ac.il

September 5, 2018

## Abstract

We prove sharp lower bounds for the smallest singular value of a partial Fourier matrix with arbitrary “off the grid” nodes (equivalently, a rectangular Vandermonde matrix with the nodes on the unit circle), in the case when some of the nodes are separated by less than the inverse bandwidth. The bound is polynomial in the reciprocal of the so-called “super-resolution factor”, while the exponent is controlled by the maximal number of nodes which are clustered together. This generalizes previously known results for the extreme cases when all of the nodes either form a single cluster, or are completely separated. We briefly discuss possible implications for the theory and practice of super-resolution under sparsity constraints.

**Keywords:** *Vandermonde matrix with nodes on the unit circle, prolate matrix, partial Fourier matrix, super-resolution, singular values, decimation*

**AMS 2010 Subject Classification:** *Primary:* 15A18; *Secondary:* 42A82, 65F22, 94A12

## 1 Introduction

### 1.1 Problem definition

Consider the  $s \times s$  matrix

$$\mathbb{G}(\mathbf{x}, \Omega) := \left[ \frac{\sin(\Omega(t_i - t_j))}{\Omega(t_i - t_j)} \right]_{1 \leq i, j \leq s}, \quad (1)$$

where  $\mathbf{x}$  is a vector of distinct nodes  $\mathbf{x} := (t_1, \dots, t_s)$  with  $t_j \in (-\frac{\pi}{2}, \frac{\pi}{2}]$ , and  $\Omega > 0$  is the normalized bandwidth. The scaling of the smallest eigenvalue<sup>1</sup>  $\lambda_{\min}(\mathbb{G})$  is of interest in applied

---

\*The research of DB and LD is supported in part by AFOSR grant FA9550-17-1-0316, NSF grant DMS-1255203, and a grant from the MIT-Skolkovo initiative. The research of GG and YY is supported in part by the Minerva Foundation.

<sup>1</sup>It is well-known that  $\mathbb{G}$  is positive-definite – for instance, because sinc is a positive-definite function.

harmonic analysis and in particular the theory of super-resolution, where this quantity controls the worst-case stability of recovering an atomic measure from bandlimited data (see Subsection 1.3 below). Since

$$\frac{\sin(\Omega t)}{\Omega t} = \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} \exp(i\omega t) d\omega = \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{k=-N}^N \exp\left(i \frac{k}{N} \Omega t\right),$$

we see that  $\mathbb{G}$  is the limit as  $N \rightarrow \infty$  of the matrix

$$\mathbb{G}_N := \frac{1}{2N} \left[ \mathcal{D}_N \left( \frac{\Omega(t_i - t_j)}{N} \right) \right]_{i,j},$$

where  $\mathcal{D}_N$  is the Dirichlet (periodic sinc) kernel

$$\mathcal{D}_N(\xi) := \sum_{k=-N}^N \exp(ik\xi) = \frac{\sin\left((N + \frac{1}{2})\xi\right)}{\sin \frac{\xi}{2}}. \quad (2)$$

For each  $N$ , let

$$\mathbb{V}_N(\mathbf{x}, \Omega) := \frac{1}{\sqrt{2N}} \left[ \exp\left(ik \frac{t_j \Omega}{N}\right) \right]_{k=-N, \dots, N}^{j=1, \dots, s} \quad (3)$$

be the rectangular  $(2N + 1) \times s$  Vandermonde matrix with complex nodes  $z_{j,N} = \exp(i\xi_{j,N})$  where  $\xi_{j,N} = \frac{t_j \Omega}{N}$ . Clearly  $\mathbb{V}_N^H \mathbb{V}_N = \mathbb{G}_N$ , and so  $\lambda_{\min}(\mathbb{G}_N) = \sigma_{\min}^2(\mathbb{V}_N)$ .

The question of lower bounds for  $\lambda_{\min}(\mathbb{G})$  (or, equivalently,  $\sigma_{\min}(\mathbb{V}_N)$ ) received much attention in the literature, see e.g. [3, 7, 19, 20, 15, 18, 5, 11].

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  $(-\pi, \pi]$ .

Given  $\mathbf{x}$  as above, we define the minimal separation (in the wrap-around sense) as

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

It is well-known that there are two very different scaling regimes for  $\lambda_{\min}$ , depending on the quantity which is frequently called the “super-resolution factor” (see Subsection 1.3 below)

$$\text{SRF} := \frac{1}{\Delta \Omega}.$$

If  $\text{SRF} < 1$  and  $s$  is fixed, the matrix  $\mathbb{G}$  is well-conditioned, and in fact it can be shown that in this case

$$\lambda_{\min} \approx (1 - \text{SRF}). \quad (4)$$

The case  $\text{SRF} > 1$  is somewhat more relevant to super-resolution applications, however all known results provide sharp bounds only in the particular case when all the nodes are clustered together, or approximately equispaced. In this setting we have the fast decay

$$\lambda_{\min}(\mathbb{G}) \approx (\Omega \Delta)^{2(s-1)}, \quad \Omega \Delta \ll 1. \quad (5)$$

For details on (4) and (5) see Section 2 below.

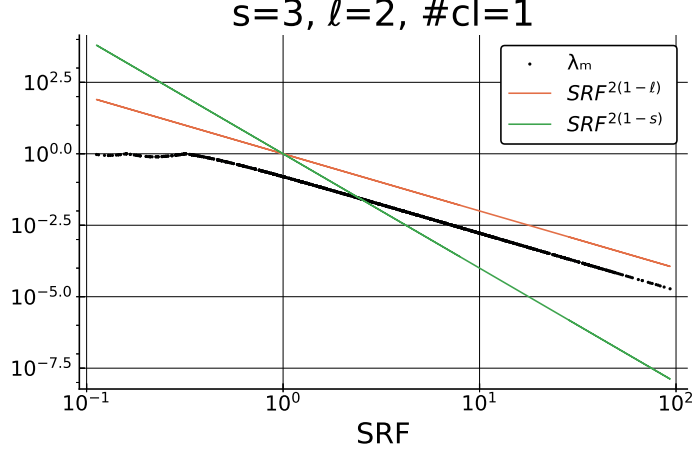


Figure 1: For different values of  $\Delta, \Omega$  we plot the quantity  $\lambda_m = \lambda_{\min}(\mathbb{G}(\mathbf{x}, \Omega))$  versus the super-resolution factor  $\text{SRF} = \frac{1}{\Delta\Omega}$ , where  $\mathbf{x} = (t_1 = \Delta, t_2 = 2\Delta, t_3 = \frac{\pi}{2})$  (i.e. a single cluster with  $s = 3$  and  $\ell = 2$ ). The correct scaling is seen to be  $\lambda_m \sim (\Delta\Omega)^{2(\ell-1)}$  rather than  $\lambda_m \sim (\Delta\Omega)^{2(s-1)}$ . See Section 4 for further details regarding the experimental setup. Note that the relationship breaks when  $\text{SRF} \leq O(1)$ , consistent with (4).

## 1.2 Main results

It turns out that the bound (5) is too pessimistic if only some of the nodes are known to be clustered. Consider for instance the configuration  $\mathbf{x} = (t_1 = \Delta, t_2 = 2\Delta, t_3 = \frac{\pi}{2})$ , then, as can be seen in Figure 1, we have in fact  $\lambda_{\min}(\mathbb{G}(\mathbf{x}, \Omega)) \approx (\Delta\Omega)^2$ , decaying much slower than  $(\Delta\Omega)^4$  – which would be the bound given by (5).

In this paper we bridge this theoretical gap. We consider the *partially clustered regime* where at most  $2 \leq \ell \leq s$  neighboring nodes can form a cluster (there can be several such clusters), with two additional parameters  $\rho, \tau$ , controlling the distance between the clusters and the uniformity of the distribution of nodes within the clusters.

**Definition 1.1.** The node vector  $\mathbf{x} = (t_1, \dots, t_s) \subset (-\frac{\pi}{2}, \frac{\pi}{2}]$  is said to form a  $(\Delta, \rho, s, \ell, \tau)$ -clustered configuration for some  $\Delta > 0$ ,  $2 \leq \ell \leq s$ ,  $\ell - 1 \leq \tau < \frac{\pi}{\Delta}$  and  $\rho \geq 0$ , if for each  $t_j$ , there exist at most  $\ell$  distinct nodes

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

such that the following conditions are satisfied:

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

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

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

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

Our main result is the following generalization of (5) for clustered configurations.

**Theorem 1.1.** *There exists a constant  $C_1 = C_1(s)$  such that for any  $4\tau\Delta \leq \rho$ , any  $\mathbf{x}$  forming a  $(\Delta, \rho, s, \ell, \tau)$ -clustered configuration, and any  $\Omega$  satisfying*

$$\frac{4\pi s}{\rho} \leq \Omega \leq \frac{\pi s}{\tau\Delta}, \quad (6)$$

we have

$$\sigma_{\min}(\mathbb{V}_N(\mathbf{x}, \Omega)) \geq C_1 \cdot (\Delta\Omega)^{\ell-1}, \quad \text{whenever } N > 2s^3 \left\lceil \frac{\Omega}{4s} \right\rceil; \quad (7)$$

$$\lambda_{\min}(\mathbb{G}(\mathbf{x}, \Omega)) \geq C_1^2 \cdot (\Delta\Omega)^{2(\ell-1)}. \quad (8)$$

The proof of Theorem 1.1 is presented in Subsection 3.3 below. It is based on the “decimation” technique, previously used in the context of super-resolution in [1, 2, 4, 5, 6] and references therein.

*Remark 1.1.* The same node vector  $\mathbf{x}$  can be regarded as a clustered configuration with different choices of the parameters  $(\ell, \rho, \tau)$ . For example, the vector  $\mathbf{x}$  from the beginning of this section (and also Figure 1) is both  $(\Delta, \frac{\pi}{2} - 2\Delta, 3, 2, 1)$ -clustered and  $(\Delta, \rho, 3, 3, \frac{\pi}{2\Delta} - 1)$ -clustered, with any  $\rho$ . To obtain as tight a bound as possible, one should choose the minimal  $\ell$  such that the condition (6) is satisfied for  $\Omega$  within the range of interest. For instance,  $\Omega$  might be too small if  $\rho$  is small enough, however by choosing  $\ell = s$  one is able to increase  $\rho$  without bound. See Figure 3 for a numerical example.

*Remark 1.2.* The constant  $C_1$  is given explicitly in (30), and it decays in  $s$  like  $\sim s^{-2s}$ . We do not know whether this rate can be substantially improved, however it is plausible that the best possible bound would scale like  $c^{-\ell}$  for some absolute constant  $c > 1$ .

For the case of finite  $N$ , one might be interested to consider the rectangular Vandermonde matrix  $\mathbb{V}_N$  without any reference to  $\Omega$ , i.e.

$$\mathbb{V}_N(\boldsymbol{\xi}) := \frac{1}{\sqrt{2N}} [\exp(ik\xi_j)]_{k=-N, \dots, N}^{j=1, \dots, s} \quad (9)$$

for some node vector  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_s)$ . Our next result is the analogue of (7) in this setting, albeit under an extra assumption that the nodes are restricted to the interval  $\frac{1}{s^2}(-\frac{\pi}{2}, \frac{\pi}{2}]$ .

**Corollary 1.1.** *There exists a constant  $C_2 = C_2(s)$  such that for any  $4\tau\Delta \leq \min(\rho, \frac{1}{s^2})$ , any  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_s) \subset \frac{1}{s^2}(-\frac{\pi}{2}, \frac{\pi}{2}]$  forming a  $(\Delta, \rho, s, \ell, \tau)$ -clustered configuration, and any  $N$  satisfying*

$$\max\left(\frac{4\pi s}{\rho}, 4s^3\right) \leq N \leq \frac{\pi s}{\tau\Delta}, \quad (10)$$

we have

$$\sigma_{\min}(\mathbb{V}_N(\boldsymbol{\xi})) \geq C_2 \cdot (N\Delta)^{\ell-1}. \quad (11)$$

*Proof.* Let us choose  $\tilde{\Omega} := \frac{N}{s^2}$  so that for all  $j = 1, \dots, s$  we have

$$\tilde{t}_j := \frac{N\xi_j}{\tilde{\Omega}} \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right].$$

Further define  $\tilde{\Delta} := s^2\Delta$ , and  $\tilde{\rho} := s^2\rho$ . We immediately obtain that the vector  $\tilde{\mathbf{x}} := (\tilde{t}_1, \dots, \tilde{t}_s)$  forms a  $(\tilde{\Delta}, \tilde{\rho}, s, \ell, \tau)$ -clustered configuration according to Definition 1.1, and the rectangular Vandermonde matrix  $\mathbb{V}_N(\boldsymbol{\xi})$  in (9) is precisely  $\mathbb{V}_N(\tilde{\mathbf{x}}, \tilde{\Omega})$ . Clearly,  $4\tau\tilde{\Delta} \leq s^2\rho = \tilde{\rho}$ , and also

$$\tilde{\Omega}s^2 = N \geq 4s^3 \implies \frac{\tilde{\Omega}}{4s} \geq 1 \implies \frac{2\tilde{\Omega}}{4s} > \left\lceil \frac{\tilde{\Omega}}{4s} \right\rceil \implies N = \tilde{\Omega}s^2 > 2s^3 \left\lceil \frac{\tilde{\Omega}}{4s} \right\rceil. \quad (12)$$

Using (10), we obtain precisely the conditions (6) with  $\tilde{\Omega}, \tilde{\rho}$  in place of  $\Omega, \rho$  respectively. Therefore the conditions of Theorem 1.1 are satisfied for  $\tilde{\mathbf{x}}, \tilde{\Omega}, \tilde{\rho}, \tilde{\Delta}, \tau$ , and so (11) follows immediately from (12) and (7), with  $C_2 = C_1$ .  $\square$

Returning back to Theorem 1.1, it turns out that the bound (8) is asymptotically optimal.

**Theorem 1.2.** *There exists an absolute constant  $\eta \ll 1$  and a constant  $C_3 = C_3(\ell)$  such that for any  $2 \leq \ell \leq s$  and any  $\Delta$  satisfying  $\Delta < \frac{\pi}{2(\ell-1)}$ , there exists a  $(\Delta, \rho', s, \ell, \tau')$ -clustered configuration  $\mathbf{x}_{\min}$  with  $s$  nodes and certain  $\rho', \tau'$  depending only on  $s, \ell$ , for which*

$$\lambda_{\min}(\mathbb{G}(\mathbf{x}_{\min}, \Omega)) \leq C_3 \cdot (\Delta\Omega)^{2(\ell-1)}, \quad \Delta\Omega < \eta.$$

The proof of Theorem 1.2 is presented in Subsection 3.4. Numerical experiments validating the above results are presented in Section 4.

### 1.3 Related work and discussion

Our main result has direct implications for the problem of super-resolution under sparsity constraints. For simplicity suppose that the nodes  $t_j$  must belong to the grid of step size  $\Delta$ . As demonstrated in [11, 18] and several other works, the minimax error rate for recovery of sparse point measures  $f(t) = \sum_{j=1}^s a_j \delta(t - t_j)$  from the bandlimited and inexact measurements  $\{\hat{f}(\omega) + e(\omega), |\omega| \leq \Omega, \|e\|_2 \leq \varepsilon\}$  is directly proportional to  $\varepsilon \cdot \min_{\mathbf{x}, |\mathbf{x}|=2s} \sqrt{\lambda_{\min}(\mathbb{G}(\mathbf{x}, \Omega))}$  where  $\mathbf{x}$  is any vector of length  $2s$ . Moreover, it is established in those works that without any further constraints on the support of  $f$ , the bound (5) holds and it is the best possible.

It is fairly straightforward to extend the results of [18] and [11] to our setting: if the support of  $f$  is known to be partially clustered (as in Definition 1.1), then the minimax error rate will satisfy

$$\inf_{\tilde{f}} \sup_f \sup_{\|e\| \leq \varepsilon} \|\tilde{f} - f\| \approx (\text{SRF})^{2\ell-1} \varepsilon \quad (13)$$

for any estimator  $\tilde{f}$  and the  $\ell_2$  norm  $\|\cdot\| = \|\cdot\|_2$ , and it will be attained by the intractable sparse  $\ell_0$ -minimization, with the additional restriction that the solutions should exhibit the appropriate clustered sparsity pattern instead of the unconstrained sparsity.

A different but closely related setting was considered in the seminal paper [12], where the measure  $f$  was assumed to have infinite number of spikes on a grid of size  $\Delta$ , with one spike per unit of time on average, but whose local complexity was constrained to have not more than  $R$  spikes per any interval of length  $R$ .  $R$  is called the ‘‘Rayleigh index’’, being the maximal number of spikes which

can be clustered together (a related notion of Rayleigh regularity was introduced in [23]). It was shown in [12] that the minimax recovery rate for such measures essentially scales like (13) where  $\ell$  is replaced with  $R$  (the work [12] had a small gap in the exponents between the lower and upper bounds, which was later closed in [11] for the finite sparse case). Our partial cluster model can therefore be regarded as the finite-dimensional version of these “sparsely clumped” measures with finite Rayleigh index, showing the same scaling of the error – polynomial in SRF and exponential in the “local complexity” of the signal.

If the grid assumption is relaxed, then one might wish to measure the accuracy of recovery  $\|\tilde{f} - f\|$  by comparing the locations of the recovered signal  $\tilde{f}$  with the true ones  $\{t_j\}$ . In this case, there are additional considerations which are required to derive the minimax rate, and it is possible to do so under the partial clustering assumptions. See [2, 6] for details, where we prove (13) in this scenario, for uniform bound on the noise  $\|e\| := \sup_{|\omega| \leq \Omega} |e(\omega)|$ . The extreme case  $\ell = s$  has been treated recently in [4, 5].

In the case of well-separated spikes (i.e. clusters of size  $\ell = 1$ ), a recent line of work using  $\ell_1$  minimization ([9, 8, 13, 10] and the great number of follow-up papers) has shown that the problem is stable and tractable.

Therefore, the partial clustering case is somewhat mid-way between the extremes  $\ell = 1$  and  $\ell = s$ , and while our results in this paper (and also in [6]) show that it is much more stable than in the unconstrained sparse case, it is an intriguing open question whether provably tractable solution algorithms exist.

Several candidate algorithms for sparse super-resolution are well-known – MUSIC, ESPRIT/matrix pencil, and variants; these have roots in parametric spectral estimation [27]. In recent years, the super-resolution properties of these algorithms are a subject of ongoing interest, see e.g. [14, 19, 25] and references therein. Smallest singular values of the partial Fourier matrices  $\mathbb{V}_N$ , for finite  $N$ , play a major role in these works, and therefore we hope that our results and techniques may be extended to analyze these algorithms as well.

## 2 Known bounds

### 2.1 Well-separated regime

Consider the well-separated case  $\Delta\Omega > 1$ , and let  $\mathbb{V}_N$  be as defined in (3), i.e. a rectangular Vandermonde matrix with nodes  $z_{j,N} = \exp(i\xi_{j,N})$  on the unit circle with  $\xi_{j,N} = t_j \frac{\Omega}{N}$ , so that  $\min_{i \neq j} |\xi_{i,N} - \xi_{j,N}| := \Delta_N > \frac{1}{N}$ .

Several more or less equivalent bounds on  $\sigma_{\min}(\mathbb{V}_N)$  are available in this case, using various results from analysis and number theory such as Ingham and Hilbert inequalities, large sieve inequalities and Selberg’s majorants [17, 20, 24, 3, 21, 22, 15, 7].

The tightest bound was obtained by Moitra in [20], where he showed that if  $N - 1 > \Delta_N^{-1}$  then

$$\sigma_{\min}(\sqrt{N}\mathbb{V}_N) \geq \sqrt{N - 1 - \frac{1}{\Delta_N}}.$$

In our setting, we have  $\Delta_N = \frac{\Delta\Omega}{N}$  and so as  $N \rightarrow \infty$  we obtain

$$\sigma_{\min}(\mathbb{V}_N) \geq \sqrt{1 - \frac{1}{N} - \frac{1}{N\Delta_N}} \rightarrow \sqrt{1 - \frac{1}{\Omega\Delta}},$$

which is exactly (4).

## 2.2 Single clustered regime

Let us now assume  $\text{SRF} > 1$ , i.e.  $\Delta\Omega < 1$  or, equivalently,  $\min_{i \neq j} |\xi_{i,N} - \xi_{j,N}| < \frac{1}{N}$ .

If all the nodes  $t_j$  are equispaced, say  $t_j = t_0 + j\Delta$ ,  $j = 1, \dots, s$ , then the matrix  $\mathbb{G}$  is the so-called *prolate matrix*, whose spectral properties are known exactly [28, 26]. Indeed, we have in this case

$$\mathbb{G}_{i,j} = \frac{\sin(\Omega(t_i - t_j))}{\Omega(t_i - t_j)} = \frac{\sin(\Omega\Delta(i - j))}{\Omega\Delta(i - j)} = \frac{\pi}{\Omega\Delta} \cdot \frac{\sin(2\pi W(i - j))}{\pi(i - j)}, \quad W := \frac{\Omega\Delta}{2\pi},$$

and therefore  $\mathbb{G} = \frac{\pi}{\Omega\Delta} \rho(s, W)$  where  $\rho(s, W)$  is the matrix defined in [26, eq. (21)]. The smallest eigenvalue of  $\rho(s, W)$ , denoted by  $\lambda_{s-1}(s, W)$  in the same paper, has the exact asymptotics for  $W$  small, given in [26, eqs. (64,65)]:

$$\lambda_{s-1}(s, W) = \frac{1}{\pi} (2\pi W)^{2s-1} C_4(s) (1 + O(W)), \quad C_4(s) := \frac{2^{2s-2}}{(2s-1) \binom{2s-2}{s-1}^3}, \quad (14)$$

which gives

$$\lambda_{\min}(\mathbb{G}) = C_4(s) (\Omega\Delta)^{2s-2} (1 + O(\Omega\Delta)), \quad \Omega\Delta \ll 1,$$

proving (5).

The same scaling was shown using Szego's theory of Toeplitz forms in [11] – see also Subsection 1.3. The authors showed that there exist  $C > 0$  and  $y^* > 0$  such that for  $\Omega\Delta < y^*$

$$\frac{C}{16} \left( \sin \frac{2\Omega\Delta}{\pi} \right)^{2s-2} \leq \lambda_{\min}(\mathbb{G}) \leq 16 \left( \sin \frac{2\Omega\Delta}{\pi} \right)^{2s-2}.$$

Essentially the same result was obtained in [18], where the authors considered partial discrete Fourier matrices

$$\Phi_{M,N,S} = \left[ \exp \left( -2\pi i \frac{mn}{N} \right) \right]_{m,n},$$

obtained from the un-normalized Discrete Fourier Transform matrix of size  $N \times N$  by taking the first  $M$  rows and an arbitrary set of  $S$  columns, with  $N \gg M$  and  $M \gg S$ . The authors showed that as  $M, N \rightarrow \infty$  with the ratio  $\frac{N}{M}$  fixed, we have the bound

$$\sigma_{\min}(\Phi_{M,N,S}) \approx \sqrt{M} \left( \frac{M}{N} \right)^{S-1},$$

which is attained for the configuration of consecutive  $S$  columns. In our equispaced setting, it is easy to see that the matrix  $\sqrt{N}\mathbb{V}_N$  for  $N$  large is precisely  $\Phi_{M',N',s}$  with  $M' = N$  and  $\frac{M'}{N'} = \Omega\Delta$ . Therefore the above result reduces to

$$\sigma_{\min}(\mathbb{V}_N) \approx (\Delta\Omega)^{s-1},$$

which is the same as (5).

### 3 Proofs

#### 3.1 Blowup

Here we introduce the uniform blowup of a node vector  $\mathbf{x} = (t_1, \dots, t_s)$  by a positive parameter  $\lambda$ , and study the effect of such a blowup mapping on the minimal wrap-around distance between the mapped nodes.

**Lemma 3.1.** *Let  $\mathbf{x}$  form a  $(\Delta, \rho, s, \ell, \tau)$  cluster, 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  $I \subset [\frac{\Omega}{2s}, \frac{\Omega}{s}]$  of total measure  $\frac{\Omega}{2s}\xi$  such that for every  $\lambda \in I$  the following holds for every  $t_j \in \mathbf{x}$ :*

$$\|\lambda y - \lambda t_j\|_{\mathbb{T}} \geq \lambda \Delta \geq \frac{\Delta \Omega}{2s}, \quad \forall y \in \mathbf{x}^{(j)} \setminus \{t_j\}; \quad (15)$$

$$\|\lambda y - \lambda t_j\|_{\mathbb{T}} \geq \frac{1-\xi}{s^2}\pi, \quad \forall y \in \mathbf{x} \setminus \mathbf{x}^{(j)}. \quad (16)$$

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

*Proof.* We begin with (15). Let  $\lambda \in [\frac{\Omega}{2s}, \frac{\Omega}{s}]$ , then  $\lambda \tau \Delta \leq \pi$  and since  $\|t_j - y\|_{\mathbb{T}} \leq \tau \Delta$  we immediately conclude that

$$\|\lambda t_j - \lambda y\|_{\mathbb{T}} = \lambda \|t_j - y\|_{\mathbb{T}} \geq \lambda \Delta.$$

To show (16), let  $\nu$  be the uniform probability measure on  $[\frac{\Omega}{2s}, \frac{\Omega}{s}]$ . Let  $t_j \in \mathbf{x}$  and  $y \in \mathbf{x} \setminus \mathbf{x}^{(j)}$  be fixed and put  $\delta := \|y - t_j\|_{\mathbb{T}}$ . For  $\lambda \in [\frac{\Omega}{2s}, \frac{\Omega}{s}]$ , let  $\gamma(\lambda) = \gamma^{(t_j, y)}(\lambda)$  be the random variable on  $\nu$ , defined by

$$\gamma^{(t_j, y)}(\lambda) := \|\lambda t_j - \lambda y\|_{\mathbb{T}}.$$

We now show that for any  $0 \leq \alpha \leq 1$

$$\nu \left\{ \gamma^{(t_j, y)}(\lambda) \leq \alpha \pi \right\} \leq 2\alpha. \quad (17)$$

Since  $\delta \geq \rho \geq \frac{4\pi s}{\Omega}$ , we can write  $\frac{\Omega}{2s} = \frac{2\pi}{\delta}(n + \zeta)$  where  $n \geq 1$  is an integer and  $0 \leq \zeta < 1$ . We break up the probability (17) as follows:

$$\begin{aligned} \nu \{ \gamma(\lambda) \leq \alpha \pi \} &= \sum_{k=1}^n \nu \left\{ \gamma(\lambda) \leq \alpha \pi \left| \lambda - \frac{\Omega}{2s} \in \frac{2\pi}{\delta} [k-1, k] \right. \right\} \nu \left\{ \lambda - \frac{\Omega}{2s} \in \frac{2\pi}{\delta} [k-1, k] \right\} \\ &\quad + \nu \left\{ \gamma(\lambda) \leq \alpha \pi \left| \lambda - \frac{\Omega}{2s} \in \frac{2\pi}{\delta} [n, n + \zeta] \right. \right\} \nu \left\{ \lambda - \frac{\Omega}{2s} \in \frac{2\pi}{\delta} [n, n + \zeta] \right\}. \end{aligned} \quad (18)$$

Now, consider the number  $a = y - t_j$ . As  $\lambda$  varies between  $\frac{\Omega}{2s} + \frac{2(k-1)\pi}{\delta}$  and  $\frac{\Omega}{2s} + \frac{2k\pi}{\delta}$ , the number  $\exp(i\lambda a)$  traverses the unit circle exactly once, and therefore the variable  $\gamma(\lambda)$  traverses the interval  $[0, \alpha \pi]$  exactly twice. Consequently,

$$\nu \left\{ \gamma(\lambda) \leq \alpha \pi \left| \lambda - \frac{\Omega}{2s} \in \frac{2\pi}{\delta} [k-1, k] \right. \right\} = \frac{2\alpha\pi}{2\pi} = \alpha.$$



Similarly, when  $\lambda$  varies between  $\frac{\Omega}{2s} + \frac{2\pi n}{\delta}$  and  $\frac{\Omega}{2s} + \frac{2\pi(n+\zeta)}{\delta}$ , we have

$$\nu \left\{ \gamma(\lambda) \leq \alpha\pi \left| \lambda - \frac{\Omega}{2s} \in \frac{2\pi}{\delta} [n, n + \zeta] \right. \right\} \leq \frac{2\alpha\pi}{2\pi\zeta} \leq \frac{\alpha}{\zeta}.$$

Overall,

$$\nu \{ \gamma(\lambda) \leq \alpha\pi \} \leq \alpha \frac{n}{n + \zeta} + \frac{\alpha}{\zeta} \frac{\zeta}{n + \zeta} = \alpha \frac{n + 1}{n + \zeta} \leq 2\alpha,$$

proving (17).

It is clear from the above that  $\{\lambda : \gamma(\lambda) \leq \alpha\pi\}$  is a union of intervals, each of length  $2\alpha\pi$ , repeating with the period of  $\frac{2\pi}{\delta}$ . Consequently the set  $\{\lambda \in [\frac{\Omega}{2s}, \frac{\Omega}{s}] : \gamma(\lambda) \leq \alpha\pi\}$  is a union of at most  $\lceil \frac{\Omega}{2s} \frac{\delta}{2\pi} \rceil$  intervals. Since  $\delta \leq \pi$  we have  $\lceil \frac{\Omega}{2s} \frac{\delta}{2\pi} \rceil \leq \lceil \frac{\Omega}{4s} \rceil$ , and so the set  $\{\lambda \in [\frac{\Omega}{2s}, \frac{\Omega}{s}] : \gamma^{(t_j, y)}(\lambda) \leq \alpha\pi\}$  is a union of at most  $\lceil \frac{\Omega}{4s} \rceil$  intervals.

Now we put  $\alpha_0 = \frac{1-\xi}{s^2}$  and apply (17) for every pair  $(t_j, y)$  where  $j = 1, \dots, s$  and  $y \in \mathbf{x} \setminus \mathbf{x}^{(j)}$ . By the union bound, we obtain

$$\nu \left\{ \exists t_j \exists y \in \mathbf{x} \setminus \mathbf{x}^{(j)} : \gamma^{(t_j, y)}(\lambda) \leq \alpha_0\pi \right\} \leq \sum_{t_j, y} 2\alpha_0 = 2 \binom{s}{2} \frac{1-\xi}{s^2} < 1 - \xi. \quad (19)$$

Fixing  $I$  as the complement of the above set,  $I = [\frac{\Omega}{2s}, \frac{\Omega}{s}] \setminus \{\lambda \in [\frac{\Omega}{2s}, \frac{\Omega}{s}] : \gamma(\lambda) \leq \alpha_0\pi\}$ , we have that  $I$  is of total measure greater or equal to  $\xi \frac{\Omega}{2s}$ , and for every  $\lambda \in I$  the estimate (16) holds. Clearly  $I^c$  is a union of at most  $\frac{s^2}{2} \lceil \frac{\Omega}{4s} \rceil$  intervals.  $\square$

Fix  $\xi = \frac{1}{2}$  and consider the set  $I$  given by Lemma 3.1. Let us also fix a finite and positive integer  $N$ , and consider the set of  $2N + 1$  equispaced points in  $[-\Omega, \Omega]$ :

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

**Proposition 3.1.** *If  $N > 2s^3 \lceil \frac{\Omega}{4s} \rceil$ , then  $P_N \cap I \neq \emptyset$ .*

*Proof.* By Lemma 3.1, the set  $I^c$  consists of  $K \leq \frac{s^2}{2} \lceil \frac{\Omega}{4s} \rceil$  intervals, and by (19) the total length of  $I^c$  is at most  $\frac{\Omega}{4s}$ . Denote the lengths of those intervals by  $d_1, \dots, d_K$ . The distance between neighboring points in  $P_N$  is  $\frac{\Omega}{N}$ , and therefore each interval contains at most  $\frac{d_j N}{\Omega} + 1$  points. Overall, the interval  $I^c$  contains at most

$$\sum_{j=1}^K \left( \frac{d_j N}{\Omega} + 1 \right) \leq \frac{\Omega}{4s} \frac{N}{\Omega} + K$$

points from  $P_N$ , and since the total number of points in  $[\frac{\Omega}{2s}, \frac{\Omega}{s}]$  is at least  $\frac{N}{2s}$ , we have

$$|P_N \cap I| \geq \frac{N}{2s} - \frac{N}{4s} - K \geq \frac{N}{4s} - \frac{s^2}{2} \lceil \frac{\Omega}{4s} \rceil > 0.$$

$\square$

### 3.2 Square Vandermonde matrices

Let  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_s)$  be a vector of  $s$  pairwise distinct complex numbers. Consider the square Vandermonde matrix

$$\mathbb{V}(\boldsymbol{\xi}) := \begin{bmatrix} 1 & 1 & \dots & 1 \\ \xi_1 & \xi_2 & \dots & \xi_s \\ \xi_1^2 & \xi_2^2 & \dots & \xi_s^2 \\ \vdots & \vdots & \ddots & \vdots \\ \xi_1^{s-1} & \xi_2^{s-1} & \dots & \xi_s^{s-1} \end{bmatrix}. \quad (20)$$

**Theorem 3.1** (Gautschi, [16]). *For a matrix  $A = (a_{i,j}) \in \mathbb{C}^{m \times n}$ , let  $\|A\|$  denote the  $\ell_\infty$  induced matrix norm*

$$\|A\|_\infty := \max_{1 \leq i \leq m} \sum_{1 \leq j \leq n} |a_{i,j}|.$$

Then we have

$$\|\mathbb{V}^{-1}(\boldsymbol{\xi})\|_\infty \leq \max_{1 \leq i \leq s} \prod_{j \neq i} \frac{1 + |\xi_j|}{|\xi_j - \xi_i|}. \quad (21)$$

**Proposition 3.2.** *Suppose that  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_s)$  is a vector of pairwise distinct complex numbers with  $|\xi_j| = 1$ ,  $j = 1, \dots, s$ , and let  $r \in \mathbb{R}$  be arbitrary. Let*

$$\mathbb{V}(\boldsymbol{\xi}, r) := \begin{bmatrix} \xi_1^r & \xi_2^r & \dots & \xi_s^r \\ \xi_1^{r+1} & \xi_2^{r+1} & \dots & \xi_s^{r+1} \\ \xi_1^{r+2} & \xi_2^{r+2} & \dots & \xi_s^{r+2} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_1^{r+s-1} & \xi_2^{r+s-1} & \dots & \xi_s^{r+s-1} \end{bmatrix}. \quad (22)$$

For  $1 \leq j < k \leq s$ , denote by  $\delta_{j,k}$  the angular distance between  $\xi_j$  and  $\xi_k$ :

$$\delta_{j,k} := \left| \text{Arg} \left( \frac{\xi_j}{\xi_k} \right) \right| = |\text{Arg}(\xi_j) - \text{Arg}(\xi_k) \pmod{(-\pi, \pi]}|.$$

Then

$$\sigma_{\min}(\mathbb{V}(\boldsymbol{\xi}, r)) \geq \frac{\pi^{1-s}}{\sqrt{s}} \min_{1 \leq j \leq s} \prod_{k \neq j} \delta_{j,k}. \quad (23)$$

*Proof.* Clearly, the matrix  $\mathbb{V}(\boldsymbol{\xi}, r)$  can be factorized as

$$\mathbb{V}(\boldsymbol{\xi}, r) = \mathbb{V}(\boldsymbol{\xi}, 0) \times \text{diag} \{ \xi_1^r, \dots, \xi_s^r \}.$$

Since  $\mathbb{V}(\boldsymbol{\xi}, 0) = \mathbb{V}(\boldsymbol{\xi})$  as in (20), using (21) we immediately have

$$\|\mathbb{V}^{-1}(\boldsymbol{\xi}, r)\|_\infty \leq 2^{s-1} \max_{1 \leq j \leq s} \prod_{k \neq j} |\xi_j - \xi_k|^{-1}. \quad (24)$$

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

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

and since for any  $\xi_j \neq \xi_k$

$$|\xi_j - \xi_k| = \left| 1 - \frac{\xi_j}{\xi_k} \right| = 2 \sin \left| \frac{1}{2} \text{Arg} \frac{\xi_j}{\xi_k} \right| = 2 \sin \left| \frac{\delta_{j,k}}{2} \right|,$$

we therefore obtain

$$\frac{2}{\pi} \delta_{j,k} \leq |\xi_j - \xi_k| \leq \delta_{j,k}. \quad (25)$$

Plugging (25) into (24) we have

$$\sigma_{\max}(\mathbb{V}^{-1}(\boldsymbol{\xi}, r)) \leq \sqrt{s} \|\mathbb{V}^{-1}(\boldsymbol{\xi}, r)\|_{\infty} \leq \sqrt{s} \pi^{s-1} \max_{1 \leq j \leq s} \prod_{k \neq j} \delta_{j,k}^{-1},$$

which is precisely (23).  $\square$

### 3.3 Proof of Theorem 1.1

We shall bound  $\sigma_{\min}(\mathbb{V}_N(\mathbf{x}, \Omega))$  defined as in (3) for sufficiently large  $N$ . For any subset  $R \subset \{-N, \dots, N\}$  let  $\mathbb{V}_{N,R}$ , be the submatrix of  $\mathbb{V}_N$  containing only the rows in  $R$ . By the Rayleigh characterization of singular values, it is immediately obvious that if  $\{-N, \dots, N\} = R_1 \cup \dots \cup R_P$  is any partition of the rows of  $\mathbb{V}_N$  then

$$\sigma_{\min}^2(\mathbb{V}_N) \geq \sum_{n=1}^P \sigma_{\min}^2(\mathbb{V}_{N,R_n}). \quad (26)$$

Let  $I$  be the set from Lemma 3.1 for  $\xi = \frac{1}{2}$ . By Proposition 3.1 we have that for all  $N > 2s^3 \lceil \frac{\Omega}{4s} \rceil$ ,  $I$  will contain a rational multiple of  $\Omega$  of the form  $\lambda_N = \frac{\Omega}{N} m$  for some  $m \in \mathbb{N}$ .

Consider the "new" nodes

$$u_{j,N} := t_j \frac{\Omega}{N} m = t_j \frac{\Omega}{N} \frac{\lambda_N N}{\Omega} = \lambda_N t_j, \quad j = 1, \dots, s. \quad (27)$$

Since  $\lambda_N \in I$ , we conclude by Lemma 3.1 that for every  $j = 1, \dots, s$

$$\|u_{j,N} - u_{k,N}\|_{\mathbb{T}} \geq \frac{1}{2s} (\Delta\Omega), \quad \forall t_k \in \mathbf{x}^{(j)} \setminus \{t_j\}; \quad (28)$$

$$\|u_{j,N} - u_{k,N}\|_{\mathbb{T}} \geq \frac{\pi}{2s^2}, \quad \forall t_k \in \mathbf{x} \setminus \mathbf{x}^{(j)}. \quad (29)$$

Since  $\lambda_N \leq \frac{\Omega}{s}$  it follows that  $ms \leq N$ . Now consider the particular interleaving partition of the rows  $\{-N, \dots, N\}$  by blocks  $R_{-m}, \dots, R_{-1}, R_0, R_1, \dots, R_m$  of  $s$  rows each, separated by  $m-1$  rows

between them (some rows might be left out):

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

For  $n = -m+1, \dots, m-1$ , each  $\mathbb{V}_{N, R_n}$  is a square Vandermonde-type matrix as in (22),

$$\mathbb{V}_{N, R_n} = \frac{1}{\sqrt{2N}} \mathbb{V}(\boldsymbol{\xi}, n),$$

with node vector

$$\boldsymbol{\xi} = \{e^{2u_{j,N}}\}_{j=1}^s,$$

where  $u_{j,N}$  are given by (27). We apply Proposition 3.2 with the crude bound obtained from (28) and (29) above:

$$\min_{1 \leq j \leq s} \prod_{k \neq j} \delta_{j,k} \geq \frac{1}{2^{s-1} s^{2s-2}} (\Delta\Omega)^{\ell-1}$$

and obtain

$$\sigma_{\min}(\mathbb{V}_{N, R_n}) \geq \frac{C_5(s)}{\sqrt{2N}} (\Delta\Omega)^{\ell-1}, \quad C_5(s) := \frac{1}{(2\pi)^{s-1} s^{2s-2} \sqrt{s}}.$$

Now we use (26) to aggregate the bounds on  $\sigma_{\min}$  for each square matrix  $\mathbb{V}_{N, R_n}$  and obtain

$$\lambda_{\min}(\mathbb{V}_N^H \mathbb{V}_N) = \sigma_{\min}^2(\mathbb{V}_N) \geq (2m-1) \frac{C_5^2}{2N} (\Delta\Omega)^{2(\ell-1)}.$$

Since  $m = \frac{\lambda_N N}{\Omega} \geq \frac{\Omega N}{2s\Omega} = \frac{N}{2s}$  and since by assumption  $N > 2s^3$ , we have that  $\frac{2m-1}{2N} \geq \frac{1}{4s}$  and so

$$\sigma_{\min}^2(\mathbb{V}_N) \geq \frac{C_5^2}{4s} (\Delta\Omega)^{2(\ell-1)}.$$

This proves (7) and (8) with

$$C_1(s) := \frac{1}{2(2\pi)^{s-1} s^{2s-1}}. \quad (30)$$

### 3.4 Proof of Theorem 1.2

Let  $\ell, s, \Delta, \Omega$  be fixed, with  $\Delta\Omega < \eta$ , where  $\eta$  will be specified during the proof below, and  $\Delta < \frac{\pi}{2(\ell-1)}$ . We shall exhibit a  $(\Delta, \rho', s, \ell, \tau')$ -clustered configuration  $\mathbf{x}_{\min}$  with certain  $\rho', \tau'$ , and a corresponding approximate minimal eigenvector  $\nu_{\min}$  of  $\mathbb{G}(\mathbf{x}_{\min}, \Omega)$ , for which the Rayleigh-Ritz quotient satisfies

$$R_{\mathbb{G}(\mathbf{x}_{\min}, \Omega)}(\nu_{\min}) := \frac{\langle \mathbb{G}(\mathbf{x}_{\min}, \Omega) \nu_{\min}, \nu_{\min} \rangle}{\langle \nu_{\min}, \nu_{\min} \rangle} \leq C_3 \cdot (\Delta\Omega)^{2(\ell-1)}, \quad (31)$$

for some constant  $C_3 = C_3(\ell)$ .

Define  $\mathbf{x}_{\ell,\Delta} = \{t_1, \dots, t_\ell\}$  to be the vector of  $\ell$  equispaced nodes separated by  $\Delta$ , i.e.  $t_j = j\Delta$ ,  $j = 1, \dots, \ell$ . Let  $\mathbb{G}^{(\ell,\ell)} = \mathbb{G}(\mathbf{x}_{\ell,\Delta}, \Omega)$  be the corresponding  $\ell \times \ell$  prolate matrix.

**Proposition 3.3.** *There exists an absolute constant  $0 < \eta_1 \ll 1$  and  $C_6 = C_6(\ell)$  such that whenever  $\Omega\Delta \leq \eta_1$ , we have*

$$\lambda_{\min} \left( \mathbb{G}^{(\ell,\ell)} \right) \leq C_6 \cdot (\Omega\Delta)^{2(\ell-1)}. \quad (32)$$

*Proof.* By Slepian's results [26] elaborated in Section 2, there exists a constant  $\eta' \ll 1$  for which (14) holds for all  $s$ , in particular for  $s = \ell$ , whenever  $W \leq \eta'$ , i.e. whenever  $\Omega\Delta \leq \eta_1 := 2\pi\eta'$ .  $\square$

We define  $\mathbf{x}_{\min}$  to be the extension of  $\mathbf{x}_{\ell,\Delta}$  such that the remaining  $s - \ell$  nodes are equidistributed between  $-\frac{\pi}{2}$  and 0, not including the endpoints. Under the assumptions on  $s, \ell, \Delta$  specified in Theorem 1.2, it is easy to check that the nodes  $t_1, \dots, t_\ell$  are between 0 and  $\frac{\pi}{2}$ , while the remaining nodes are separated at least by

$$\rho' := \frac{\pi}{2(s - \ell + 1)}. \quad (33)$$

Therefore,  $\mathbf{x}_{\min}$  is a particular  $(\Delta, \rho', s, \ell, \tau')$ -clustered configuration according to Definition 1.1, with  $\rho'$  given by (33) and  $\tau' := \ell - 1$ .

Now we construct the vector  $\nu_{\min}$ . Let  $\nu_0 \in \mathbb{R}^\ell$  be the unit-norm eigenvector of  $\mathbb{G}^{(\ell,\ell)}$  corresponding to the smallest eigenvalue  $\lambda_{\min}(\mathbb{G}^{(\ell,\ell)})$ . In fact,  $\nu_0$  is precisely the  $(\ell - 1)^{\text{st}}$  discrete prolate spheroidal sequence (DPSS)  $v_n^{(\ell-1)}(\ell, \frac{\Delta\Omega}{2\pi})$  as defined in [26, eq. (18)], index-limited to  $n = 0, 1, \dots, \ell - 1$ . Let

$$\nu_{\min} := [\nu_0; 0_{(s-\ell) \times 1}] \in \mathbb{R}^s.$$

By our choice of  $\nu_0$  we have  $\|\nu_{\min}\|_2 = 1$ . Now

$$\begin{aligned} R_{\mathbb{G}(\mathbf{x}_{\min}, \Omega)}(\nu_{\min}) &= \nu_{\min}^T \mathbb{G}(\mathbf{x}_{\min}, \Omega) \nu_{\min} \\ &= \lim_{N \rightarrow \infty} \|\mathbb{V}_N(\mathbf{x}_{\min}, \Omega) \nu_{\min}\|^2 \\ &= \lim_{N \rightarrow \infty} \|\mathbb{V}_N(\mathbf{x}_{\ell,\Delta}, \Omega) \nu_0\|^2 \\ &= \nu_0^T \mathbb{G}^{(\ell,\ell)} \nu_0 \\ &= \lambda_{\min} \left( \mathbb{G}^{(\ell,\ell)} \right). \end{aligned}$$

Using (32), this concludes the proof of (31) and of Theorem 1.2 with  $C_3 = C_6$  and  $\eta = \eta_1$ .

## 4 Numerical experiments

In order to validate Theorem 1.1, we computed  $\lambda_{\min}(\mathbb{G})$  for varying values of  $\Delta, \Omega, \ell, s$  and the actual clustering configurations. We checked two clustering scenarios:

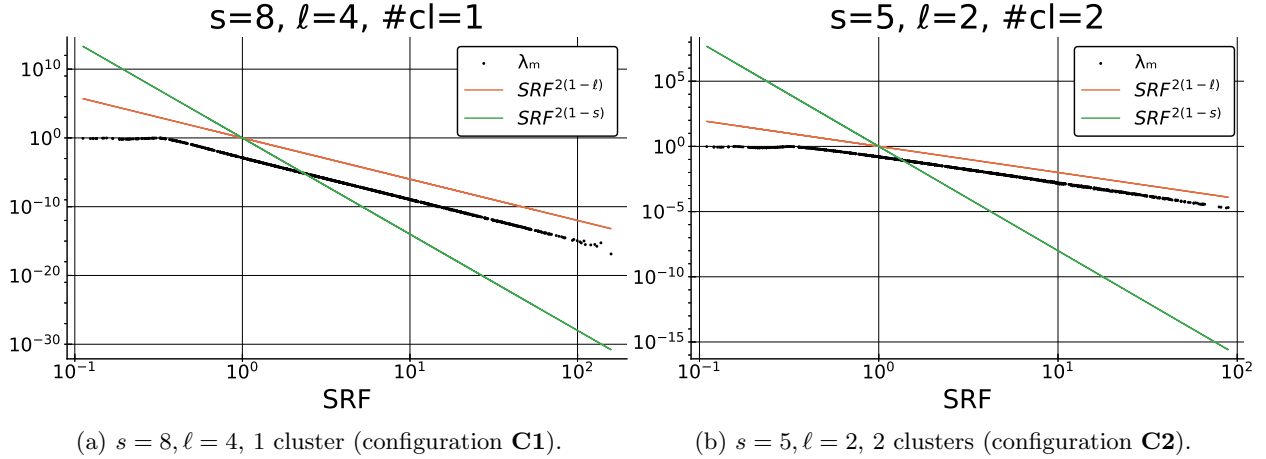


Figure 2: Decay rate of  $\lambda_{\min}$  as a function of SRF. Results of  $n = 1000$  random experiments with varying  $\Delta, \Omega$  are plotted versus the theoretical bound  $(\Omega\Delta)^{2(\ell-1)}$ . The curve  $(\Omega\Delta)^{2(s-1)}$  is shown for comparison. The bound stops to be accurate for  $\text{SRF} < O(1)$ .

**C1** A single equispaced cluster of size  $\ell$  in  $[\Delta, \ell\Delta]$ , with the rest of the nodes maximally separated and equidistributed in  $(-\frac{\pi}{2}, 0)$  (exactly as in the construction of  $\mathbf{x}_{\min}$  in Subsection 3.4). For example, in the case  $s = 8, \ell = 4$  (as in Figure 2a) we have  $t_j = j\Delta$  for  $j = 1, \dots, 4$ , and  $t_j = -\frac{\pi}{2} + (j-4)\frac{\pi}{10}$  for  $j = 5, \dots, 8$ .

**C2** Split the  $s$  nodes into two groups, and construct two single-clustered configurations as follows:

- (a)  $s_1 = \lfloor \frac{s}{2} \rfloor$  nodes, a single equispaced cluster of size  $\ell_1 = \ell$  in  $[\Delta, \ell\Delta]$ , and the rest of the  $s_1 - \ell_1$  nodes maximally separated and equidistributed in  $(\ell\Delta, \frac{\pi}{2})$ ;
- (b)  $s_2 = s - s_1$  nodes, a single equispaced cluster of size  $\ell_2 = \ell$  in  $[-\frac{\pi}{2} + \Delta, -\frac{\pi}{2} + \ell\Delta]$ , and the rest of the  $s_2 - \ell_2$  nodes maximally separated and equidistributed in  $(-\frac{\pi}{2} + \ell\Delta, 0)$ .

For example, in the case  $s = 5, \ell = 2$  (as in Figure 2b) we have  $t_1 = \Delta, t_2 = 2\Delta$  and  $t_3 = -\frac{\pi}{2} + \Delta, t_4 = -\frac{\pi}{2} + 2\Delta, t_5 = -\frac{\pi}{4} + \Delta$ .

In each experiment we fixed  $\ell, s$  and one of the scenarios above, and run  $n = 1000$  random tests for varying  $\Delta, \Omega$ . The results are presented Figure 2.

In another experiment (Figure 3), we fixed  $\Delta, \ell, s$  and changed  $\Omega$ . As expected, when  $\Omega$  became small enough, the left inequality in (6) was violated, and indeed we can see that in this case the asymptotic decay was  $\approx \text{SRF}^{2(1-s)}$ . See Remark 1.1 for further discussion.

To check Theorem 1.2, we added the computation of the approximate smallest eigenvector  $\nu_{\min}$  as defined in Subsection 3.4. We compared the exact  $\lambda_{\min}(\mathbb{G}(\mathbf{x}_{\min}, \Omega))$  with  $\lambda_{\min}(\mathbb{G}^{(\ell, \ell)})$ , and found them to be virtually indistinguishable, as is seen in Figure 4.

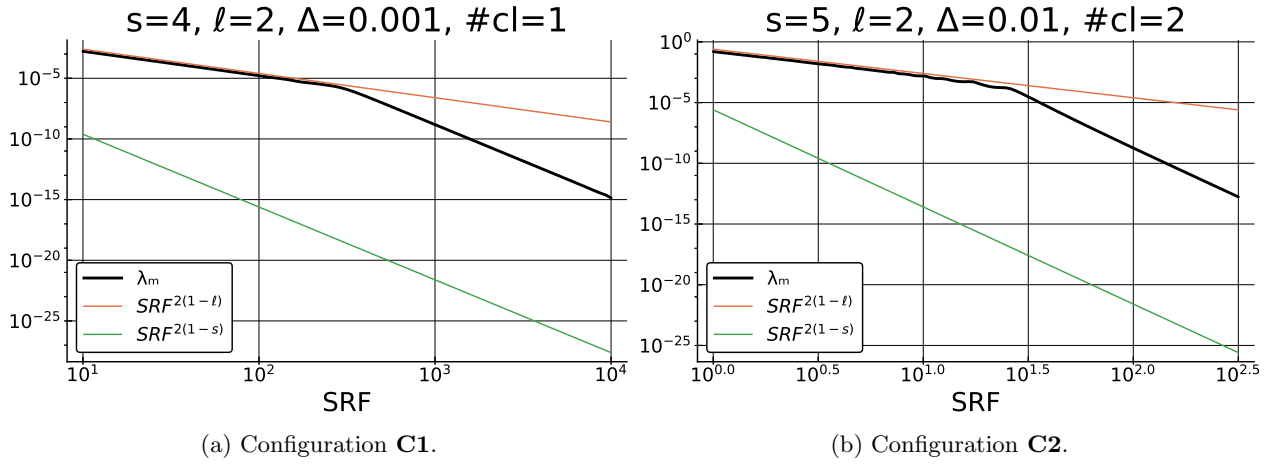


Figure 3: Breakdown of cluster structure. When  $\Omega$  is small enough, the assumptions of Theorem 1.1 are violated for certain  $\ell < s$ . As a result, the decay rate of  $\lambda_{\min}$  corresponds to the entire  $\mathbf{x}$  being a single cluster of size  $\ell = s$ .  $\Delta$  is kept fixed. See Remark 1.1.

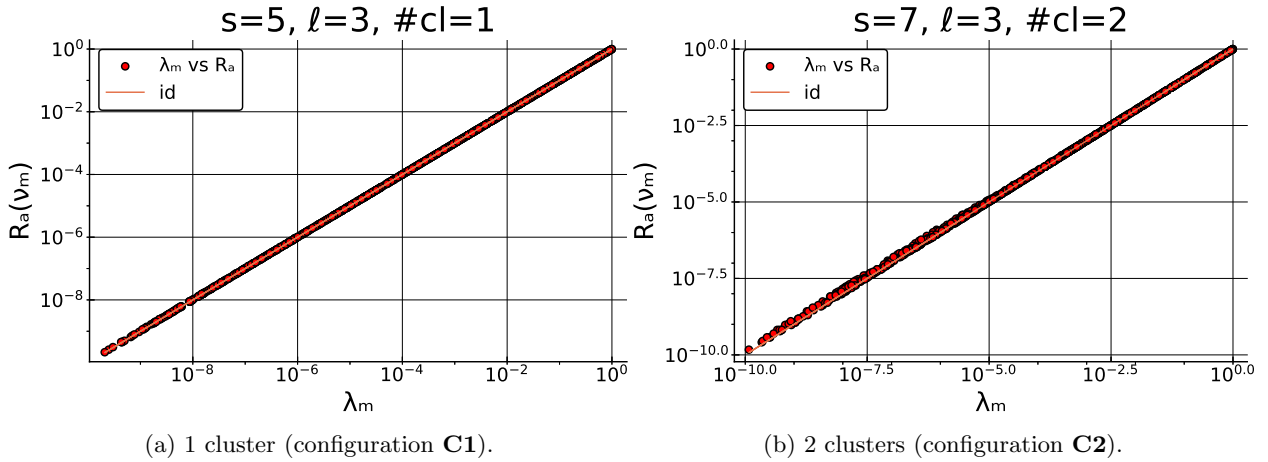


Figure 4: Minimal eigenvalue vs Rayleigh-Ritz quotient of the approximate minimal eigenvector. The values are virtually indistinguishable, confirming the tightness of the bound in Theorem 1.2.

## References

- [1] A. Akinshin, D. Batenkov, G. Goldman, and Y. Yomdin. Error amplification in solving Prony system with near-colliding nodes. *arXiv:1701.04058 [math]*, Jan. 2017.
- [2] A. Akinshin, D. Batenkov, and Y. Yomdin. Accuracy of spike-train Fourier reconstruction for colliding nodes. In *2015 International Conference on Sampling Theory and Applications (SampTA)*, pages 617–621, May 2015.
- [3] C. Aubel and H. Bölcskei. Vandermonde matrices with nodes in the unit disk and the large sieve. *Applied and Computational Harmonic Analysis*, Aug. 2017.
- [4] D. Batenkov. Accurate solution of near-colliding Prony systems via decimation and homotopy continuation. *Theoretical Computer Science*, 681:27–40, June 2017.
- [5] D. Batenkov. Stability and super-resolution of generalized spike recovery. *Applied and Computational Harmonic Analysis*, 45(2):299–323, Sept. 2018.
- [6] D. Batenkov, G. Goldman, and Y. Yomdin. Super-resolution of near-colliding point sources.
- [7] F. Bazán. Conditioning of rectangular Vandermonde matrices with nodes in the unit disk. *SIAM Journal on Matrix Analysis and Applications*, 21:679, 2000.
- [8] E. J. Candès and C. Fernandez-Granda. Super-Resolution from Noisy Data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, Dec. 2013.
- [9] E. J. Candès and C. Fernandez-Granda. Towards a Mathematical Theory of Super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, June 2014.
- [10] Y. de Castro and F. Gamboa. Exact reconstruction using Beurling minimal extrapolation. *Journal of Mathematical Analysis and Applications*, 395(1):336–354, Nov. 2012.
- [11] L. Demanet and N. Nguyen. The recoverability limit for superresolution via sparsity. 2014.
- [12] D. Donoho. Superresolution via sparsity constraints. *SIAM Journal on Mathematical Analysis*, 23(5):1309–1331, 1992.
- [13] V. Duval and G. Peyré. Exact Support Recovery for Sparse Spikes Deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, Oct. 2014.
- [14] A. Fannjiang. Compressive Spectral Estimation with Single-Snapshot ESPRIT: Stability and Resolution. *arXiv:1607.01827 [cs, math]*, July 2016.
- [15] P. Ferreira. Superresolution, the Recovery of Missing Samples, and Vandermonde Matrices on the Unit Circle. 1999.
- [16] W. Gautschi. On inverses of Vandermonde and confluent Vandermonde matrices. *Numerische Mathematik*, 4(1):117–123, 1962.
- [17] A. E. Ingham. Some trigonometrical inequalities with applications to the theory of series. *Mathematische Zeitschrift*, 41(1):367–379, Dec. 1936.



- [18] W. Li and W. Liao. Stable super-resolution limit and smallest singular value of restricted Fourier matrices. *arXiv:1709.03146 [cs, math]*, Sept. 2017.
- [19] W. Liao and A. Fannjiang. MUSIC for single-snapshot spectral estimation: Stability and super-resolution. *Applied and Computational Harmonic Analysis*, 40(1):33–67, Jan. 2016.
- [20] A. Moitra. Super-resolution, Extremal Functions and the Condition Number of Vandermonde Matrices. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, STOC '15, pages 821–830, New York, NY, USA, 2015. ACM.
- [21] H. L. Montgomery. *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*. American Mathematical Soc., 1994.
- [22] H. L. Montgomery and R. C. Vaughan. Hilbert’s Inequality. *Journal of the London Mathematical Society*, s2-8(1):73–82, May 1974.
- [23] V. I. Morgenshtern and E. J. Candes. Stable Super-Resolution of Positive Sources: The Discrete Setup. *Preprint*, 2014.
- [24] M. Negreanu and E. Zuazua. Discrete Ingham Inequalities and Applications. *SIAM Journal on Numerical Analysis*, 44(1):412–448, Jan. 2006.
- [25] D. Potts and M. Tasche. Error Estimates for the ESPRIT Algorithm. In D. A. Bini, T. Ehrhardt, A. Y. Karlovich, and I. Spitkovsky, editors, *Large Truncated Toeplitz Matrices, Toeplitz Operators, and Related Topics*, volume 259, pages 621–648. Springer International Publishing, Cham, 2017.
- [26] D. Slepian. Prolate spheroidal wave functions, fourier analysis, and uncertainty – V: The discrete case. *Bell System Technical Journal*, The, 57(5):1371–1430, May 1978.
- [27] P. Stoica and R. Moses. *Spectral Analysis of Signals*. Pearson/Prentice Hall, 2005.
- [28] J. Varah. The prolate matrix. *Linear Algebra and its Applications*, 187:269–278, July 1993.