THE SPECTRAL PROPERTIES OF VANDERMONDE MATRICES WITH CLUSTERED NODES

DMITRY BATENKOV, BENEDIKT DIEDERICHS, GIL GOLDMAN, AND YOSEF YOMDIN

Abstract. We study rectangular Vandermonde matrices $\mathbf{V}$ with $N+1$ rows and s irregularly spaced nodes on the unit circle, in cases where some of the nodes are "clustered" together – the elements inside each cluster being separated by at most $h \lesssim \frac{1}{N}$, and the clusters being separated from each other by at least $\theta \gtrsim \frac{1}{N}$. We show that any pair of column subspaces corresponding to two different clusters are nearly orthogonal: the minimal principal angle between them is at most

$$\frac{\pi}{2} - \frac{c_1}{N\theta} - c_2 N h,$$

for some constants c_1, c_2 depending only on the multiplicities of the clusters. As a result, spectral analysis of $\mathbf{V}_N$ is significantly simplified by reducing the problem to the analysis of each cluster individually. Consequently we derive accurate estimates for 1) all the singular values of $\mathbf{V}$, and 2) componentwise condition numbers for the linear least squares problem. Importantly, these estimates are exponential only in the local cluster multiplicities, while changing at most linearly with s.

1. Introduction

1.1. **Background.** For an ordered set of distinct nodes $\mathcal{X} = \{x_1, \ldots, x_s\}$ with $x_j \in (-\pi, \pi]$, and $N \geqslant s-1$, we consider the $(N+1) \times s$ Vandermonde matrix $\mathbf{V} = \mathbf{V}_N(\mathcal{X})$ with nodes $\{e^{\imath x_j}\}_{j=1}^{s}$, given by[1]

$$\mathbf{V}_N(\mathcal{X}) = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ e^{\imath x_1} & e^{\imath x_2} & \ldots & e^{\imath x_s} \\ e^{\imath 2 x_1} & e^{\imath 2 x_2} & \ldots & e^{\imath 2 x_s} \\ \vdots & \vdots & \vdots & \vdots \\ e^{\imath N x_1} & e^{\imath N x_2} & \ldots & e^{\imath N x_s} \end{bmatrix}. \tag{1.1}$$

Square and rectangular Vandermonde matrices have been studied quite extensively by numerical analysts due to their close relation to polynomial interpolation and approximation, quadrature and related topics, see e.g. [16, 9, 22, 23, 25, 24, 21, 11, 37, 43, 10] and references therein. The matrices $\mathbf{V}_N$ as in (1.1) have also received recent attention in the applied harmonic analysis community with relation to the problem of mathematical super-resolution [6, 4, 36, 5, 34, 33, 19, 18, 31, 30], where the magnitude of their smallest singular value controls the limit of stable recovery of point sources from bandlimited data. Similar connections exist in spectral estimation and direction of arrival problems, where $\mathbf{V}_N$ are closely related to data covariance matrices [32, 40, 42, 46].

While Vandermonde matrices with real nodes are known to be ill-conditioned (for instance, the condition number must grow exponentially in s, see [10, 11, 37] and references therein), the situation may be drastically different for complex nodes. Indeed, the columns of $\mathbf{V}_N$ become orthogonal when

2010 *Mathematics Subject Classification.* Primary 15A18, 65T40, 65F20.

Key words and phrases. Vandermonde matrices with nodes on the unit circle, nonuniform Fourier matrices, sub-Rayleigh resolution, singular values, super-resolution, subspace angles, condition number.

[1]Note a slight abuse of notation as $\mathbf{V}$ depends not only on $\mathcal{X}$, but also on the ordering of the nodes. Therefore, we always assume that the set of nodes comes with an arbitrary, but fixed, ordering.

$\mathcal{X}$ is a subset of the roots of unity of order $N+1$, but on the other hand may be arbitrary close to each other if two or more nodes collide. When the minimal distance[2] between any two nodes in $\mathcal{X}$ (denoted by η in this section) is larger than $\frac{1}{N}$, the matrix $\mathbf{V}_N(\mathcal{X})$ is known to be well-conditioned. The sharpest result in that direction was recently presented in [19], building upon earlier results [3, 36, 33, 8]. On the other hand, the spectral decomposition of $\mathbf{V}_N$ in the special case of equispaced and nearly colliding nodes (i.e. $x_j = x_0 + j\eta$ and $N\eta \ll 1$) can be derived from the seminal works on the spectral concentration problem by Slepian and co-workers, see [5, 39, 38] and references therein. However, the general case of irregularly spaced and partially colliding nodes is much less understood.

1.2. **Contributions.** In the line of recent works [1, 2, 5, 31, 30, 33, 34], in this paper we study the properties of $\mathbf{V}_N$ under the geometric assumption that the nodes $\{x_j\}$ form "clusters", each of which is contained in an interval of length at most $h \lesssim \frac{1}{N}$, while the inter-cluster distances are at least $\theta \gtrsim \frac{1}{N}$. Our main result (Theorem 2.1) establishes that the subspaces of $\mathbb{C}^{N+1}$ corresponding to each cluster (the so-called "cluster subspaces", see Definition 2.3 below) are nearly orthogonal. In more detail, we show that for large enough $N\theta$ and small enough Nh, the complementary subspace angle between each pair of cluster subspaces is at most $\frac{c_1}{N\theta} + c_2 Nh$ for some constants c_1, c_2 depending only on the multiplicities of (number of nodes in) the clusters. *As a result, spectral analysis of $\mathbf{V}_N$ is significantly simplified, reducing the problem to the analysis of each cluster separately* (see Theorem 2.2). To demonstrate this general principle, we establish the following results for the case that the points are approximately uniformly distributed in each cluster:

(1) We derive full asymptotic description of *all the singular values* of $\mathbf{V}_N$ (Theorems 2.3, 2.2).
(2) In the particular case where the size of all the clusters is of the same order h, the singular values of $N^{-1/2}\mathbf{V}_N$ have the following simple scales (up to constants):
$$(Nh)^0, \ldots., (Nh)^{\ell-1},$$
where ℓ is the maximal multiplicity of any cluster. Furthermore, the number of singular values scaling as $(Nh)^{j-1}$ is exactly equal to the number of clusters of multiplicity at least j (see Corollary 2.1).
(3) In Theorem 2.4 we obtain *componentwise* stability bounds of the linear least squares problem
$$\min_{\mathbf{a}} \|\mathbf{V}_N(\mathcal{X})\mathbf{a} - \mathbf{b}\|_2.$$
In particular, we show that the entries of $\mathbf{a}$ corresponding to the nodes of $\mathcal{X}$ inside a cluster of size h and multiplicity ℓ (i.e. h, ℓ may be different for different clusters), have condition numbers proportional to $(Nh)^{1-\ell}$, with the proportionality constant scaling *linearly* with s. In contrast, without prior geometric assumptions, all the entries have condition number on the scale of $(N\eta)^{1-s}$ (where η is the global minimal separation of the nodes).

1.3. **Related work and discussion.** The scaling $\sigma_j\left(\frac{1}{\sqrt{N}}\mathbf{V}_N\right) \approx (N\eta)^{j-1}$ for a single cluster can also be derived from [39, 32]. In the proof of Theorem 2.3 we use a particular technique based on Taylor expansion of the kernel matrix $\mathbf{V}_N^H\mathbf{V}_N$, used in [44]. It will be interesting to investigate the possibility of extending our result to more general types of matrices, for instance those considered in [32]. Another interesting question is to allow the nodes of the Vandermonde matrix to be in a small annulus containing the unit circle, as in [37].

Several previous works studied the behaviour of the minimal singular value of clustered Vandermonde matrices in the regime $N\eta \ll 1$. Below, positive constants that are independent of N, η are

[2] All distances are in the wrap-around sense, to be defined precisely below.

indicated by $c_1, c_2, \dots, c, c', \dots$. From Corollary 2.1 it directly follows that

$$\sigma_{\min}\left(\frac{1}{\sqrt{N}}\mathbf{V}_N\right) \geqslant c(N\eta)^{\ell-1}, \quad N\eta < c', \tag{1.2}$$

where, again, ℓ is the largest multiplicity. This scaling has been previously established in [5, 33, 31], by completely different techniques and under additional conditions. In the following, we briefly compare those results to ours.

- The bound (1.2) was first established in [5] in the regime $N\theta \geqslant c_1$. However, it was also required that the entire node set $\mathcal{X}$ be contained in an interval of length $\frac{1}{s^2}$. Compared with [5] we similarly require that $N\theta \geqslant c_2$, but the node set $\mathcal{X}$ is no longer restricted to such a tiny interval.
- In [31] (building upon [33]), (1.2) was shown to hold with $c' = 1$ but under further restriction of the form

$$N\theta > c_3(\gamma)(N\eta)^{-\gamma}, \tag{1.3}$$

  where $\gamma > 0$ can be arbitrarily small. However in this case $\lim_{\gamma\to 0} c_3(\gamma) = \infty$ and also $\lim_{\gamma\to 0} c(\gamma) = 0$ where c is the constant in (1.2). To make a comparison, let us fix θ, γ and consider what values of η are covered, first by our result: $\eta \in \left(0, c'N^{-1}\right]$ and then by [33, 31]: $\eta \in \left[c''N^{-(1+\gamma^{-1})}, N^{-1}\right]$. Note that:
  - The regime $\eta \in \left(0, c'N^{-1}\right]$ allows $\eta \to 0$ for a fixed N;
  - If N is sufficiently large then the regimes overlap, and all values of $\eta \in \left(0, N^{-1}\right]$ are either covered by the results of this paper or those of [33, 31].
- Our constant c in (1.2) is not explicit, while the authors of [31] managed to prove that under the condition (1.3) with $\gamma = \frac{\ell-1}{2}$, the constant $c(\gamma)$ is of order $C^{-\ell}$, for an absolute constant C (in [5] a much worse estimate $c \sim s^{-2s}$ was given). The scaling $c \sim C^{-\ell}$ can be shown to be optimal (up to the magnitude of the absolute constant C), see [31, Example 5.1]. Simulations suggest that (1.2) holds with $c \sim C^{-\ell}$ whenever $N\theta \geqslant c_4$, i.e. the clusters separation should only be large with respect to $\frac{1}{N}$, regardless of the relation between N and η. We plan to close this gap in the constant in a future publication.

In addition, our results have consequences for the analysis of super-resolution problem and algorithms, both on-grid and off-grid [6, 5, 20, 33, 34]. In this context, it should also be interesting to investigate low-rank approximation for the covariance matrices [11, 46].

We hope that using the cluster subspace orthogonality it will be possible to provide an accurate description of the *singular vectors*, in particular, their spectral concentration properties. These questions are important in e.g. time-frequency analysis and sampling of multiband signals [27].

1.4. **Organization of the paper.** In Section 2 we establish some notation and formulate our main results. In Section 3 we develop the necessary tools and prove Theorem 2.1. In Section 4 we analyze the case of a single cluster and prove Theorem 2.3. In Section 5 we analyze the multi-cluster setting and prove Theorems 2.2 and 2.4. In Section 6 we present results of numerical experiments validating our main results.

1.5. **Acknowledgements.** The research of GG and YY is supported in part by the Minerva Foundation.

2. Main results

2.1. **Notation.** For $x, y \in \mathbb{R}$, we denote the wrap-around distance

$$\Delta(x, y) = |\operatorname{Arg} \exp \imath(x - y)| = |x - y \mod (-\pi, \pi]| \in [0, \pi],$$

where for $z \in \mathbb{C} \backslash \{0\}$, $\operatorname{Arg}(z)$ is the principal value of the argument of z, taking values in $(-\pi, \pi]$. For a matrix $\mathbf{A}$, $\mathbf{A}^H$ denotes the Hermitian transpose of $\mathbf{A}$, and $\mathbf{A}^\dagger$ denotes the Moore-Penrose pseudoinverse [12] of $\mathbf{A}$. The k-th component of a vector $\mathbf{x}$ is denoted by $(\mathbf{x})_k$, and (i, j)-th entry of a matrix $\mathbf{A}$ is denoted by $(\mathbf{A})_{i,j}$. We use $\|\mathbf{A}\|$ for the spectral, $\|\mathbf{A}\|_F$ for the Frobenius and $\|\mathbf{A}\|_{\max}$ for the maximum norm of $\mathbf{A}$. We recall their definitions, as well as some relations we will frequently use, in Appendix A. We use the Landau symbols $\mathcal{O}$ for an asymptotic upper bound and Θ for asymptotically equal up to constants.

Definition 2.1 (Single cluster configuration). *The node set $\mathcal{X} = \{x_1, \ldots, x_s\} \subset (-\pi, \pi]$ is said to form*

- *an (h, s)-cluster if*
$$\forall x, y \in \mathcal{X}, x \neq y : \quad 0 < \Delta(x, y) \leqslant h;$$
- *an (h, τ, s)-cluster, for some $\tau > 0$, if*
$$\forall x, y \in \mathcal{X}, x \neq y : \quad \tau h \leqslant \Delta(x, y) \leqslant h.$$

Remark 2.1. *Clearly, an (h, τ, s) cluster is in particular an (h, s)-cluster, where in addition we assume that the nodes are approximately uniformly distributed within the cluster.*

Definition 2.2 (Multi-cluster configuration). *The node set $\mathcal{X} = \{x_1, \ldots, x_s\} \subset (-\pi, \pi]$ is said to form an $((h^{(j)}, s^{(j)})_{j=1}^M, \theta)$ (respectively, $((h^{(j)}, \tau^{(j)}, s^{(j)})_{j=1}^M, \theta)$)-clustered configuration if there exists an M-partition $\mathcal{X} = \biguplus_{j=1}^M \mathcal{C}^{(j)}$, such that for each $j \in \{1, \ldots, M\}$ the following conditions are satisfied:*

- *$\mathcal{C}^{(j)}$ is an $(h^{(j)}, s^{(j)})$ (respectively, an $(h^{(j)}, \tau^{(j)}, s^{(j)})$)-cluster;*
- *$\Delta(x, y) \geqslant \theta > 0, \quad \forall x \in \mathcal{C}^{(j)}, \ \forall y \in \mathcal{X} \backslash \mathcal{C}^{(j)}$.*

Below we write $C_k(s_1, s_2, \ldots, s_j)$ or $N_k(s_1, s_2, ..., s_j)$, for some indexes k, j and parameters $s_1, \ldots, s_j$, to indicate a constant that depends only on $s_1, \ldots, s_j$.

2.2. **Cluster subspace orthogonality.**

Definition 2.3 (Cluster subspace). *Let $\mathcal{X} = \{x_1, \ldots, x_s\} \subset (-\pi, \pi]$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_s$ denote the columns of the Vandermonde matrix $\mathbf{V}_N(\mathcal{X})$. We denote by $L(\mathcal{X})$ the subspace spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_s$, i.e.*

$$L(\mathcal{X}) := L(\mathcal{X}, N) = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_s\} \subset \mathbb{C}^{N+1}.$$

Definition 2.4 (Minimal principal angle). *For two subspaces $L_1, L_2 \subset \mathbb{C}^{N+1}$, the minimal principal angle $\angle_{\min}(L_1, L_2)$ between L_1 and L_2, taking values in $[0, \frac{\pi}{2}]$, is defined as*

$$\angle_{\min}(L_1, L_2) := \min_{v \in L_1 \backslash \{0\}, u \in L_2 \backslash \{0\}} \arccos\left(\frac{|\langle v, u\rangle|}{\|v\| \cdot \|u\|}\right).$$

Our first main result, proved in Section 3, reads as follows.

Theorem 2.1 (Cluster subspaces orthogonality). *Let $\mathcal{X}$ and $\mathcal{Y}$ form an $(h^{(1)}, s_1)$- and $(h^{(2)}, s_2)$-clusters, respectively, such that*

$$\Delta(x,y) \geqslant \theta \qquad\qquad \forall x \in \mathcal{X}, y \in \mathcal{Y}.$$

Put $h = \max\left(h^{(1)}, h^{(2)}\right)$. Then there exist positive constants C_1, C_2, C_3 and C_4, depending only on s_1 and s_2, such that for all N with $C_1 \leqslant N \leqslant \frac{C_2}{h}$ we have

$$\angle_{\min}(L(\mathcal{X},N), L(\mathcal{Y},N)) \geqslant \frac{\pi}{2} - \frac{C_4}{N\theta} - C_3 Nh. \tag{2.1}$$

Remark 2.2. *Note that Theorem 2.1 holds irrespective of the inner structure of each cluster.*

Remark 2.3. *Clearly, if $N > \max\left(C_1, \frac{4}{\pi}C_4\right) \cdot \max\left(1, \theta^{-1}\right)$ and $Nh < \min\left(C_2, \frac{\pi}{4C_3}\right)$, then $\angle_{\min}$ is guaranteed to be positive. So, Theorem 2.1 will always produce a nontrivial bound for sufficiently large N and sufficiently small Nh.*

2.3. **Full spectral description.** Now we establish *accurate estimates for all the singular values of* $\mathbf{V}$.

First, using the orthogonality result (Theorem 2.1), we show in Section 5.3 that the set of all the singular values of $\mathbf{V}$ equals, up to a small multiplicative perturbation, to *the union of the sets of singular values of the sub-matrices of* $\mathbf{V}$*, corresponding to the clusters.*

Theorem 2.2 (Multi-cluster Vandermonde matrix singular values). *Suppose that the node set $\mathcal{X} = \{x_1, \ldots, x_s\} \subset (-\pi, \pi]$ forms an $((h^{(j)}, s^{(j)})_{j=1}^M, \theta)$-clustered configuration, and consider the Vandermonde matrix $\mathbf{V}_N(\mathcal{X})$ and its sub-matrices formed by each cluster, $\mathbf{V}_N(\mathcal{C}^{(1)}), \ldots, \mathbf{V}_N(\mathcal{C}^{(M)})$. Let*

$$\sigma_1 \geqslant \ldots \geqslant \sigma_s$$

be the singular values of $\mathbf{V}_N(\mathcal{X})$ in non-increasing order. Further, let

$$\tilde{\sigma}_1 \geqslant \ldots \geqslant \tilde{\sigma}_s$$

be all the singular values of the sub-matrices $\{\mathbf{V}_N(\mathcal{C}^{(j)})\}$, also in non-increasing order.

Put $h = \max_j(h^{(j)})$. Then there exist positive constants C_5, C_6, C_7 and C_8, depending only on $s^{(1)}, \ldots, s^{(M)}$, such that for all N satisfying $\frac{C_5}{\theta} \leqslant N \leqslant \frac{C_6}{h}$ we have

$$\left(1 - \frac{C_7}{N\theta} - C_8 Nh\right)^{\frac{1}{2}} \tilde{\sigma}_j \leqslant \sigma_j \leqslant \left(1 + \frac{C_7}{N\theta} + C_8 Nh\right)^{\frac{1}{2}} \tilde{\sigma}_j, \qquad j = 1, \ldots, s. \tag{2.2}$$

Remark 2.4. *In the proof of Theorem 2.2 it is ensured that $1 - \frac{C_7}{N\theta} - C_8 Nh > 0$, see Proposition 5.1 and in particular* (5.11)*. Compare this with Remark 2.3.*

Thus, the analysis of the spectrum of $\mathbf{V}_N$ is reduced to looking at each cluster separately. To that effect, our next result (proved in Section 4) provides the decay rates for the singular values corresponding to a single cluster, assuming that the distribution of the nodes inside the cluster is approximately uniform.

Theorem 2.3 (Single cluster singular values). *Let $\mathcal{X}$ form an (h, τ, s)-cluster. Then there exist constants $C_9(\tau, s)$, $C_{10}(\tau, s)$ and $C_{11}(s)$, such that for all $N \geqslant s$ and $Nh \leqslant C_9$ we have*

$$C_{10} N^{\frac{1}{2}} (Nh)^{j-1} \leqslant \sigma_j\left(\mathbf{V}_N(\mathcal{X})\right) \leqslant C_{11} N^{\frac{1}{2}} (Nh)^{j-1}, \quad j = 1, \ldots, s. \tag{2.3}$$

Combining Theorems 2.2 and 2.3 provides complete scaling of all the singular values of $\mathbf{V}_N(\mathcal{X})$ for the case of approximately uniform clusters (i.e. $\min_j \tau^{(j)} \geqslant \tau > 0$). If we assume, in addition, that all the cluster sizes $h^{(j)}$ are of the same order (for simplicity we may take them to be equal to each other), then we have a particularly simple description of the spectrum of $\mathbf{V}_N$ as follows.

Corollary 2.1 (Entire spectrum)**.** *Let $\mathcal{X}$ form an $((h^{(j)},\tau^{(j)},s^{(j)})_{j=1}^M,\theta)$-clustered configuration, and furthermore suppose that $h^{(1)}=h^{(2)}=\cdots=h^{(M)}=h$. For each $j=1,2,\ldots,\max_j s^{(j)}$, define*

$$\ell_j := \#\{1 \leqslant k \leqslant M : j \leqslant s^{(k)}\}.$$

Then, there exist constants C_{12} and C_{13} such that for all $N\theta \geqslant C_{12}$ and $Nh \leqslant C_{13}$ there are precisely ℓ_j singular values of $\mathbf{V}_N(\mathcal{X})$ scaling like $\asymp N^{\frac{1}{2}}(Nh)^{j-1}$. All the constants in the statement depend only on $\left(s^{(j)},\tau^{(j)}\right)_{j=1}^M$.

2.4. **Accuracy of least squares problems.**

Definition 2.5 (Cluster index set)**.** *For a multi-cluster $\mathcal{X}=\{x_1,\ldots,x_s\}$, for each $1\leqslant j\leqslant M$, we define $C_j=C_j(\mathcal{X})$ to be the indices of all the nodes in cluster j, i.e.*

$$C_j(\mathcal{X}) = \left\{\ell : x_\ell \in \mathcal{C}^{(j)}\right\}.$$

Definition 2.6 (Least squares solution)**.** *Given a multi-cluster $\mathcal{X}$ as in Definition 2.2, $N>s$, and a vector $\mathbf{b}\in\mathbb{C}^{N+1}$, define*

$$\mathbf{a}(\mathcal{X},\mathbf{b}) := \arg\min_{\mathbf{a}} \|\mathbf{V}_N(\mathcal{X})\mathbf{a}-\mathbf{b}\|_2.$$

In Section 5.4 we prove the following result.

Theorem 2.4. *Suppose that the node set $\mathcal{X}=\{x_1,\ldots,x_s\}\subset(-\pi,\pi]$ forms an $((h^{(j)},\tau^{(j)},s^{(j)})_{j=1}^M,$ $\theta)$-clustered configuration, with $h=\max_{1\leqslant j\leqslant M}(h^{(j)})$. Let N satisfy $\frac{C_{14}}{\theta}\leqslant N\leqslant\frac{C_{15}}{h}$ for certain constants $C_{14}(s^{(1)},\ldots,s^{(M)})$ and $C_{15}(s^{(1)},\ldots,s^{(M)})$ to be specified in the proof. Next, let $\mathbf{a}_0\in\mathbb{C}^s$ be arbitrary and $\mathbf{b}_0=\mathbf{V}_N(\mathcal{X})\mathbf{a}_0$. Then for each $j\in\{1,2,\ldots,M\}$ there exists a constant $C_{16}(s^{(j)},\tau^{(j)})$ such that for all $\ell\in C_j(\mathcal{X})$ we have*

$$\left|\left(\mathbf{a}(\mathcal{X},\mathbf{b})-\mathbf{a}_0\right)_\ell\right| \leqslant C_{16}s\left(\frac{1}{Nh^{(j)}}\right)^{s^{(j)}-1}\|\mathbf{b}-\mathbf{b}_0\|_\infty. \tag{2.4}$$

The above result shows that for multi-cluster distributions of the nodes, the componentwise condition numbers are much more accurate than the standard condition number. Indeed, without any geometric assumptions on the node distribution, the classical condition number $\kappa(\mathbf{V}_N)=\frac{\sigma_{\max}(\mathbf{V}_N)}{\sigma_{\min}(\mathbf{V}_N)}$ is well-known to grow exponentially with s, see e.g. [11, 37] and references therein. In contrast, the errors in (2.4) grow exponentially with the multiplicities of each cluster, and only linearly in the overall number of nodes (compare also with [5, Corollary 3.10]).

Remark 2.5. *If one considers perturbations in $\mathcal{X}$ as well, the stability analysis becomes more complicated, see e.g.* [13, 26]*. The main point we would like to emphasize here is that the components of* $\mathbf{a}$ *have different condition numbers according to the multiplicity of the nodes of $\mathcal{X}$ in the corresponding cluster.*

3. Orthogonality of cluster subspaces

In this section we prove Theorem 2.1. To facilitate the reading, let us start with a short overview of the steps.

(1) In Section 3.1 we introduce several objects associated with an (h,s)-cluster $\mathcal{X}$: a particular basis for $L=L(\mathcal{X},N)$, called the "divided difference basis"; the limit space $\bar{L}$; and a certain basis for $\bar{L}$ called the "limit basis". We show that the spaces L and $\bar{L}$ are "close", in the sense that the divided difference basis vectors of L are close to the corresponding limit basis vectors of $\bar{L}$, up to order $\mathcal{O}(Nh)$.

(2) Next, in Subsection 3.2 we show that limit basis is well-conditioned (i.e. the smallest singular value of the corresponding matrix is effectively bounded from below for large enough N).
(3) Given two different clusters of the nodes, $\mathcal{X}$ with s_1 nodes and $\mathcal{Y}$ with s_2 nodes, in Subsection 3.3 we prove the following key property: the limit spaces $\bar{L}(\mathcal{X}, N)$, $\bar{L}(\mathcal{Y}, N)$, are nearly orthogonal, with $\angle_{\min}(\bar{L}_1, \bar{L}_2)$ being of order $\frac{\pi}{2} - \mathcal{O}\left(\frac{1}{N}\right)$.
(4) Combining the above results, in Subsection 3.4 we conclude that the angle between any $\mathbf{v} \in L_1$ and $\mathbf{u} \in L_2$ is at least $\frac{\pi}{2} - \mathcal{O}(Nh) - \mathcal{O}(\frac{1}{N})$, completing the proof.

3.1. **Spaces and bases.** Let $\mathcal{X}$ form an (h, s)-cluster as in Definition 2.1, fix $N \geqslant s-1$, and, as per Definition 2.3, let

$$L(\mathcal{X}) = L(\mathcal{X}, N) = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_s\} \subset \mathbb{C}^{N+1}, \quad \mathbf{V}_N(\mathcal{X}) = [\mathbf{v}_1, \ldots, \mathbf{v}_s].$$

We start by constructing a basis for $L(\mathcal{X})$, different from $\{\mathbf{v}_1, \ldots, \mathbf{v}_s\}$, which is given by certain divided differences of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_s$. For the reader's convenience we recall the definition of divided differences below, and list several of their propoerties in Lemma B.1 in Appendix B. For further references see e.g. [17] and [7, Section 6.2].

Definition 3.1 (Divided finite differences)**.** *Let an arbitrary sequence of points $(t_1, t_2, \ldots,)$ be given (repetitions are allowed). For any $n = 1, 2, \ldots,$, and for any smooth enough real-valued function f, defined at least on $t_1, \ldots, t_n$, the $n-1$-st divided difference $[t_1, \ldots, t_n]f$ is the n-th coefficient, in the Newton form, of the (uniquely defined) Hermite interpolation polynomial p, which agrees with f and its derivatives of appropriate order on $t_1, \ldots, t_n$, i.e. $[t_1, \ldots, t_n]f \equiv [t_1, \ldots, t_n]p$ where*

$$f^{(\ell)}(t_j) = p^{(\ell)}(t_j): \quad 1 \leqslant j \leqslant n,\ 0 \leqslant \ell < d_j := \#\{i : \ t_i = t_j\};$$
$$p(t) \equiv \sum_{j=1}^{n} \{[t_1, \ldots, t_j]p\} \prod_{k=1}^{j-1} (t - t_j).$$

Divided differences of complex-valued functions are defined by applying them to the real and the imaginary parts separately:

$$[t_1, \ldots, t_n](f + \imath g) = [t_1, \ldots, t_n]f + \imath[t_1, \ldots, t_n]g.$$

For a vector of functions $\mathbf{f} = (f_1, \ldots, f_m)$, we denote

$$[t_1, \ldots, t_n]\mathbf{f} := ([t_1, \ldots, t_n]f_1, \ldots, [t_1, \ldots, t_n]f_m) \in \mathbb{C}^m.$$

Denote by

$$\mathbf{v}_N(x) = \left(1, e^{\imath x}, e^{2\imath x}, \ldots, e^{N\imath x}\right)$$

the vector of $N+1$ exponential functions. Note that by (B.2) the standard basis for $L(\mathcal{X}, N)$ can be simply written as

$$\mathbf{v}_j \equiv [x_j]\mathbf{v}_N(x), \quad j = 1, \ldots, s. \tag{3.1}$$

Definition 3.2. *Given $\mathcal{X} = \{x_1, \ldots, x_s\} \subset (-\pi, \pi]$ and a positive integer N, define, for every $j \in \{1, \ldots, s\}$ the following vector:*

$$\mathbf{w}_j := (j-1)![x_1, \ldots, x_j]\mathbf{v}_N(x), \qquad \tilde{\mathbf{w}}_j := \frac{\mathbf{w}_j}{\|\mathbf{w}_j\|}. \tag{3.2}$$

The ordered set $\mathcal{W} = \mathcal{W}(\mathcal{X}) = \{\mathbf{w}_1, \ldots, \mathbf{w}_s\}$ is called the divided difference basis *to $L(\mathcal{X})$, and the ordered set $\widetilde{\mathcal{W}} = \widetilde{\mathcal{W}}(\mathcal{X}) = \{\tilde{\mathbf{w}}_1, \ldots, \tilde{\mathbf{w}}_s\}$ is called the normalized divided difference basis to $L(\mathcal{X})$.*

Definition 3.3 (Limit basis and limit space)**.** *Given* $\zeta \in (-\pi, \pi]$ *and positive integers* N, s*, define, for every* $j \in \{1, \ldots, s\}$ *the following vector:*

$$\mathbf{u}_j := (j-1)![\underbrace{\zeta, \ldots, \zeta}_{j \text{ times}}]\mathbf{v}_N(x), \qquad \tilde{\mathbf{u}}_j = \frac{\mathbf{u}_j}{\|\mathbf{u}_j\|}. \tag{3.3}$$

The ordered set $\mathcal{U}(\zeta, N, s) := \{\mathbf{u}_1, \ldots, \mathbf{u}_s\} \subset \mathbb{C}^{N+1}$ *is called the* (ζ, N, s) *limit basis, and the ordered set* $\widetilde{\mathcal{U}} = \widetilde{\mathcal{U}}(\zeta, N, s) = \{\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_s\}$ *is called the normalized* (ζ, N, s) *limit basis.*
We further define the limit space at ζ *as*

$$\bar{L}(\zeta, N, s) := \operatorname{span}\{\mathcal{U}(\zeta, N, s)\} \subset \mathbb{C}^{N+1}.$$

Definition 3.4 (Cluster limit space)**.** *For* $\mathcal{X}$ *forming an* (h, s)*-cluster we define the cluster limit space* $\bar{L}(\mathcal{X}, N)$ *by*

$$\bar{L}(\mathcal{X}, N) := \bar{L}(x_1, N, s) \subset \mathbb{C}^{N+1}.$$

Remark 3.1. *The choice of the point* x_1 *in the definition of* $\bar{L}(\mathcal{X}, N)$ *is arbitrary, in the sense that all subsequent results hold true if we replace* x_1 *in Definition 3.4 with another node* $x_j \in \mathcal{X}$*, or, even more generally, with any other point* $x \in [\min_j x_j, \max_j x_j]$*. Furthermore, note that* $\bar{L}(\mathcal{X}, N)$ *depends neither on the cluster size* h *nor on the relative positions of the points inside the cluster.*

Below we establish several properties of the sets $\mathcal{W}$, $\widetilde{\mathcal{W}}$, $\mathcal{U}$, $\widetilde{\mathcal{U}}$ and the relationships between them, which will be used in the rest of the section, towards the proof of Theorem 2.1 in Subsection 3.4.

Proposition 3.1. *Let* $\mathcal{X}$ *form an* (h, s)*-cluster, and let* $N \geqslant s - 1$*.*

(1) $\operatorname{span}\{\mathcal{W}(\mathcal{X})\} = L(\mathcal{X})$*.*
(2) Each $\mathbf{u}_j \in \mathcal{U}(\zeta, N, s)$ *is explicitly given by*

$$\mathbf{u}_{j,k} = \frac{\mathrm{d}^{j-1}}{\mathrm{d}x^{j-1}} e^{\imath kx}\Big|_{\zeta} = (\imath k)^{j-1} e^{\imath k\zeta}, \qquad j = 1, \ldots, s, \; k = 0, \ldots, N, \tag{3.4}$$

(3) $\mathcal{U}(\zeta, N, s)$ *is a linearly independent set, i.e. it is indeed a basis for* $\bar{L}(\zeta, N, s)$*.*
(4) Putting $\zeta = x_1$*, then*

$$\lim_{h \to 0} \mathbf{w}_j = \mathbf{u}_j \in \mathcal{U}(x_1, N, s), \qquad j \in \{1, \ldots, s\}.$$

(5) With $C_{17} = C_{17}(s) := \frac{1}{\sqrt{2s-1}}$*, we have*

$$C_{17} N^{j-\frac{1}{2}} \leqslant \|\mathbf{u}_j\| \leqslant N^{j-\frac{1}{2}}, \quad \mathbf{u}_j \in \mathcal{U}(\zeta, N, s), \; j \in \{1, \ldots, s\}. \tag{3.5}$$

(6) With $C_{18} := 2\sqrt{2}$*, we have*

$$\|\tilde{\mathbf{u}}_j - \tilde{\mathbf{w}}_j\| \leqslant C_{18} N h, \quad \tilde{\mathbf{u}}_j \in \widetilde{\mathcal{U}}(x_1, N, s), \; \tilde{\mathbf{w}}_j \in \widetilde{\mathcal{W}}(\mathcal{X}), \; j \in \{1, \ldots, s\}. \tag{3.6}$$

Proof. In the proofs below, we use results from Appendices B and C.

(1) Using the extended form (B.3) together with (3.1), we see that each vector $\mathbf{w}_j$ is a linear combination of the original basis vectors $\mathbf{v}_1, \ldots, \mathbf{v}_j$, i.e.

$$\mathbf{w}_j = \sum_{k=1}^{j} \frac{1}{\prod_{k \neq j}(x_j - x_k)} \mathbf{v}_k.$$

Hence the set $\mathbf{w}_1, \ldots, \mathbf{w}_s$ is given by a triangular transformation, with non-zero coefficients, of the original basis $\mathbf{v}_1, \ldots, \mathbf{v}_s$, and therefore forms another basis to the subspace $L(\mathcal{X})$.

(2) Follows from (B.5) and (3.3).

(3) Using the previous explicit formula for $\mathbf{u}_j$, the matrix $[\mathbf{u}_1, \ldots, \mathbf{u}_s] \in \mathbb{C}^{(N+1)\times s}$ is, up to a diagonal factor, the Pascal-Vandermonde matrix and is known to be full rank (see e.g. [4, Section 4.1]).
(4) Directly follows from the definitions and the continuity property (B.1).
(5) For each $j \in \{1, \ldots, s\}$ we have that $\|\mathbf{u}_j\| = \sqrt{\sum_{k=0}^{N} k^{2(j-1)}}$. Using (C.2) we have

$$\frac{N^{j-\frac{1}{2}}}{\sqrt{2s-1}} \leqslant \frac{N^{j-\frac{1}{2}}}{\sqrt{2j-1}} \leqslant \|\mathbf{u}_j\| \leqslant N^{j-\frac{1}{2}},$$

which then proves (3.5).
(6) First we have that

$$\begin{aligned}\|\tilde{\mathbf{u}}_j - \tilde{\mathbf{w}}_j\| = \left\| \frac{\mathbf{u}_j}{\|\mathbf{u}_j\|} - \frac{\mathbf{w}_j}{\|\mathbf{w}_j\|} \right\| &= \frac{1}{\|\mathbf{u}_j\|} \left\| \mathbf{u}_j - \mathbf{w}_j + \left(1 - \frac{\|\mathbf{u}_j\|}{\|\mathbf{w}_j\|}\right) \mathbf{w}_j \right\| \\ &\leqslant \frac{1}{\|\mathbf{u}_j\|} \left(\|\mathbf{u}_j - \mathbf{w}_j\| + \big|\|\mathbf{u}_j\| - \|\mathbf{w}_j\|\big|\right) \\ &\leqslant 2\frac{\|\mathbf{u}_j - \mathbf{w}_j\|}{\|\mathbf{u}_j\|}.\end{aligned} \tag{3.7}$$

Let $k \in \{0, 1, \ldots, N\}$. By assumption, $|x_\ell - x_m| \leqslant h$ for all $\ell, m \in \{1, \ldots, j\}$. Using (3.4), (3.2), and applying (B.4) to the real and the imaginary parts of $e^{\imath kx}$, we obtain $\xi_1, \xi_2 \in [\min_\ell x_\ell, \max_\ell x_\ell]$ such that

$$\begin{aligned}|\mathbf{u}_{j,k} - \mathbf{w}_{j,k}|^2 &= \left| \frac{\mathrm{d}^{j-1}}{\mathrm{d}x^{j-1}} e^{\imath kx}\Big|_{x_1} - (j-1)![x_1, \ldots, x_j] e^{\imath kx} \right|^2 \\ &= \left| \frac{\mathrm{d}^{j-1}}{\mathrm{d}x^{j-1}} \cos(kx)\Big|_{x_1} - \frac{\mathrm{d}^{j-1}}{\mathrm{d}x^{j-1}} \cos(kx)\Big|_{\xi_1} \right|^2 + \left| \frac{\mathrm{d}^{j-1}}{\mathrm{d}x^{j-1}} \sin(kx)\Big|_{x_1} - \frac{\mathrm{d}^{j-1}}{\mathrm{d}x^{j-1}} \sin(kx)\Big|_{\xi_2} \right|^2 \\ &\leqslant 2k^{2j}h^2,\end{aligned}$$

where the last line is obtained by applying the standard mean value theorem to $\cos^{(j-1)}(kx)$ and $\sin^{(j-1)}(kx)$, respectively.

Now $|\mathbf{u}_{j,k} - \mathbf{w}_{j,k}| \leqslant \sqrt{2}k^j h$, which implies that

$$\|\mathbf{u}_j - \mathbf{w}_j\| \leqslant \sqrt{2}h \left(\sum_{k=1}^{n} k^{2j}\right)^{1/2}.$$

Using $\|\mathbf{u}_j\| = \sqrt{\sum_{k=0}^{N} k^{2(j-1)}}$ and (3.7), we obtain that

$$\|\tilde{\mathbf{u}}_j - \tilde{\mathbf{w}}_j\| \leqslant 2\sqrt{2}h \left(\frac{\sum_{k=1}^{N} k^{2j}}{\sum_{k=1}^{N} k^{2(j-1)}}\right)^{1/2} \leqslant 2\sqrt{2}hN.$$

□

3.2. **Conditioning of the limit basis.** Given $\zeta \in (-\pi, \pi]$, and $N \geqslant s-1$, consider the normalized limit basis $\widetilde{\mathcal{U}} = \widetilde{\mathcal{U}}(\zeta, N, s)$ to the limit space $\bar{L}(\zeta, N, s) \subset \mathbb{C}^{N+1}$. While for any fixed N we have seen that $\widetilde{\mathcal{U}}$ is linearly independent, in this section we will furthermore establish that for sufficiently large N the corresponding condition number is bounded from below by a constant which does not depend on N.

For each ζ as above, let $\mathbf{U}(\zeta, N, s) \in \mathbb{C}^{(N+1)\times s}$ denote the matrix with columns $\{\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_s\}$:

$$\mathbf{U}(\zeta, N, s) = [\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_s] = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ e^{\imath\zeta} & (\imath)e^{\imath\zeta} & \ldots & (\imath)^{s-1}e^{\imath\zeta} \\ e^{\imath 2\zeta} & (\imath 2)e^{\imath 2\zeta} & \ldots & (\imath 2)^{s-1}e^{\imath 2\zeta} \\ \vdots & \vdots & \vdots & \vdots \\ e^{\imath N\zeta} & (\imath N)e^{\imath N\zeta} & \ldots & (\imath N)^{s-1}e^{\imath N\zeta} \end{bmatrix} \cdot \operatorname{diag}(\|\mathbf{u}_1\|^{-1}, \ldots, \|\mathbf{u}_s\|^{-1}).$$

Proposition 3.2. *Given a positive integer s and $\zeta \in (-\pi, \pi]$, there exist a monotonically increasing constant $N_1 = N_1(s)$ and a monotonically decreasing constant $\Xi = \Xi(s) > 0$, such that for any $N \geqslant N_1$,*

$$\sigma_{\min}\left(\mathbf{U}(\zeta, N, s)\right) \geqslant \Xi.$$

Moreover, $\Xi = \sqrt{\frac{\lambda_{\min}(\bar{\mathbf{H}}_s)}{2}}$, where $\bar{\mathbf{H}}_s$ is the normalized $s \times s$ Hilbert matrix defined in (3.12) *below.*

Proof. First we extract the $\imath^{j-1}$ factor from each column by putting

$$\tilde{\mathbf{U}}_N = \mathbf{U}(\zeta, N, s) \cdot \operatorname{diag}(\imath^{-0}, \ldots, \imath^{-(s-1)}),$$

and clearly $\sigma_{\min}\left(\tilde{\mathbf{U}}_N\right) = \sigma_{\min}\left(\mathbf{U}(\zeta, N, s)\right)$. We therefore continue with $\tilde{\mathbf{U}}_N$.

Next we consider the Gramian matrix $\tilde{\mathbf{U}}_N^H \tilde{\mathbf{U}}_N$, and we have that

$$\left[\tilde{\mathbf{U}}_N^H \tilde{\mathbf{U}}_N\right]_{j,l} = \frac{\sum_{k=0}^N k^{j+l-2}}{\|\mathbf{u}_j\| \cdot \|\mathbf{u}_l\|} = \frac{\sum_{k=0}^N k^{j+l-2}}{\sqrt{\left(\sum_{k=0}^N k^{2(j-1)}\right)\left(\sum_{k=0}^N k^{2(l-1)}\right)}}. \tag{3.8}$$

By combining (3.8) and (C.1) we get[3]

$$\left[\tilde{\mathbf{U}}_N^H \tilde{\mathbf{U}}_N\right]_{j,l} = \frac{\frac{N^{j+l-1}}{j+l-1} + \mathcal{O}(N^{j+l-2})}{\sqrt{\left(\frac{N^{2j-1}}{2j-1} + \mathcal{O}(N^{2j-2})\right)\left(\frac{N^{2l-1}}{2l-1} + \mathcal{O}(N^{2l-2})\right)}} = \frac{\sqrt{2j-1}\sqrt{2l-1}}{j+l-1} + \mathcal{O}\left(\frac{1}{N}\right), \tag{3.9}$$

where the $\mathcal{O}$ asymptotic notation here and throughout the rest of the proof will always refer to N. Define the inner product $\langle\cdot,\cdot\rangle_H$ and the corresponding norm $\|\cdot\|_H$, over the vector space of polynomials of degree smaller than s, as

$$\langle P, Q\rangle_H = \int_0^1 P(x)\overline{Q(x)}\mathrm{d}x. \tag{3.10}$$

Then by (3.9) we can write $\tilde{\mathbf{U}}_N^H \tilde{\mathbf{U}}_N$ as follows

$$\tilde{\mathbf{U}}_N^H \tilde{\mathbf{U}}_N = \bar{\mathbf{H}}_s + \mathbf{E}_s, \tag{3.11}$$

where $\mathbf{E}_s, \bar{\mathbf{H}}_s$ are Hermitian matrices, the entries of $\mathbf{E}_s$ are $\mathcal{O}\left(\frac{1}{N}\right)$, and $\bar{\mathbf{H}}_s$ is the normalized Hilbert matrix (see e.g [15, 41]),

$$[\bar{\mathbf{H}}_s]_{j,l} = \left\langle \frac{x^{j-1}}{\|x^{j-1}\|_H}, \frac{x^{l-1}}{\|x^{l-1}\|_H} \right\rangle_H = \int_0^1 \frac{x^{j-1}}{\|x^{j-1}\|_H}\frac{x^{l-1}}{\|x^{l-1}\|_H}\mathrm{d}x. \tag{3.12}$$

$\bar{\mathbf{H}}_s$ is the Gramian matrix with respect to the inner products of the form (3.10), of the normalized monomial basis $\frac{1}{\|1\|_H}, \ldots, \frac{x^{s-1}}{\|x^{s-1}\|_H}$. Therefore, it is non-degenerate, and its smallest eigenvalue $\lambda_{\min}(\bar{\mathbf{H}}_s)$ is bounded from below by a positive constant depending only on s.

[3]Notice also that an explicit bound for the $\mathcal{O}(N)$ terms can be obtained from Faulhaber's formula given in (C.1).

On the other hand, since the entries of $\mathbf{E}_s$ are $\mathcal{O}\left(\frac{1}{N}\right)$ and $\mathbf{E}_s$ is Hermitian (but not necessarily PSD), we have that

$$\lambda_{\min}(\mathbf{E}_s) \geqslant -\|\mathbf{E}_s\| \geqslant -\|\mathbf{E}_s\|_F \geqslant s\mathcal{O}\left(\frac{1}{N}\right). \tag{3.13}$$

Using (3.13) we set $N_1(s)$ to be such that for all $N \geqslant N_1$, $\lambda_{\min}(\mathbf{E}_s) \geqslant -\frac{\lambda_{\min}(\bar{\mathbf{H}}_s)}{2}$. Furthermore, we increase $N_1(s)$ to be as least as large as $N_1(1), \ldots, N_1(s-1)$. Then using (3.11) and Weyl's perturbation inequality, for all $N \geqslant N_1$,

$$\sigma_{\min}(\tilde{\mathbf{U}}_N) = \sqrt{\lambda_{\min}\left(\tilde{\mathbf{U}}_N^H \tilde{\mathbf{U}}_N\right)} \geqslant \sqrt{\lambda_{\min}\left(\bar{\mathbf{H}}_s\right) + \lambda_{\min}\left(\mathbf{E}\right)} \geqslant \sqrt{\frac{\lambda_{\min}\left(\bar{\mathbf{H}}_s\right)}{2}}.$$

The last claim that $\Xi(s)$ is decreasing follows as deleting the last row and column of $\bar{\mathbf{H}}_{s+1}$ gives $\bar{\mathbf{H}}_s$. Thus, $\lambda_{\min}(\bar{\mathbf{H}}_{s+1}) \leqslant \lambda_{\min}(\bar{\mathbf{H}}_s)$ by the minimax principle. ☐

Remark 3.2. *Clearly,* $\sigma_{\min}\left(\mathbf{U}(\zeta, N, s)\right)$ *does not depend on* ζ *and we just proved that*

$$\sigma_{\min}\left(\mathbf{U}(\zeta, N, s)\right) \to \sqrt{\lambda_{\min}(\bar{\mathbf{H}}_s)}, \quad N \to \infty.$$

Since $\mathbf{U}(\zeta, N, s)$ *is injective whenever* $N \geqslant s-1$ *(recall Proposition 3.1), we could replace* $N_1(s)$ *by* $s-1$*. Then, however, we would lose the control over* $\Xi(s)$*. Here we can give an asymptotic lower bound for* $\Xi(s)$ *as follows. The asymptotic behavior of* $\lambda_{\min}(\mathbf{H}_s)$ *where* $\mathbf{H}_s$ *is the unnormalized Hilbert matrix is known to be*

$$\lambda_{\min}(\mathbf{H}_s) \in \Theta\left(\sqrt{s}(1+\sqrt{2})^{-4s}\right), \quad s \to \infty,$$

see [45]*, equation (3.35). By normalization we lose at most another factor of* s*, ending up with a lower bound of order* $(1+\sqrt{2})^{-2s}/\sqrt{s}$ *for* $\Xi(s)$*.*

3.3. **Near-orthogonality of the limit spaces.**

Proposition 3.3. *For any two distinct points* $\zeta_1, \zeta_2 \in (-\pi, \pi]$ *and positive integers* s_1, s_2*, there exists a positive constant* $C_{19}(s_1, s_2)$*, such that for any* $N \geqslant \max(s_1, s_2) - 1$*:*

$$|\langle \mathbf{z}_1, \mathbf{z}_2 \rangle| \leqslant \frac{C_{19}}{\Delta\left(\zeta_1, \zeta_2\right) N}; \quad \mathbf{z}_\ell \in \widetilde{\mathcal{U}}\left(\zeta_\ell, N, s_\ell\right), \ \ \ell = 1, 2. \tag{3.14}$$

Proof. Consider the following limit vectors (as they appear in Definition 3.3):

$$\mathcal{U}\left(\zeta_\ell, N, s_\ell\right) = \left\{\mathbf{u}_j^{(\ell)}\right\}_{j=1}^{s_\ell}, \qquad \ell = 1, 2;$$
$$\widetilde{\mathcal{U}}\left(\zeta_\ell, N, s_\ell\right) = \left\{\tilde{\mathbf{u}}_j^{(\ell)}\right\}_{j=1}^{s_\ell}, \qquad \ell = 1, 2.$$

Consider an arbitrary any pair of normalized limit vectors $\mathbf{z}_1 = \tilde{\mathbf{u}}_p^{(1)}$ and $\mathbf{z}_2 = \tilde{\mathbf{u}}_q^{(2)}$, $p = 1, \ldots, s_1$, $q = 1, \ldots, s_2$. By (3.5) we have:

$$\begin{aligned} \frac{1}{\sqrt{2s_1 - 1}} N^{p-\frac{1}{2}} &\leqslant \|\mathbf{u}_p^{(1)}\| \leqslant N^{p-\frac{1}{2}}, \\ \frac{1}{\sqrt{2s_2 - 1}} N^{q-\frac{1}{2}} &\leqslant \|\mathbf{u}_q^{(2)}\| \leqslant N^{q-\frac{1}{2}}. \end{aligned} \tag{3.15}$$

By (3.4) we have

$$\langle \mathbf{u}_p^{(1)}, \mathbf{u}_q^{(2)} \rangle = \imath^{p+q-2}(-1)^{q-1} \sum_{k=0}^{N} k^{p+q-2} z^k,$$

with $z = e^{\imath(\zeta_1-\zeta_2)}$. Notice that since $\zeta_1 \neq \zeta_2$, we have $z \neq 1$. Now using Lemma D.1 we get that

$$\left|\langle \mathbf{u}_p^{(1)}, \mathbf{u}_q^{(2)} \rangle\right| = \left|\sum_{k=0}^{N} k^{p+q-2} z^k\right| \leqslant \frac{2}{|1-z|} N^{p+q-2}. \tag{3.16}$$

Finally we have that

$$|1-z| = |1-e^{\imath(\zeta_1-\zeta_2)}| \geqslant \frac{2}{\pi}\Delta(\zeta_1,\zeta_2),$$

which together with (3.16) gives that

$$\left|\langle \mathbf{u}_p^{(1)}, \mathbf{u}_q^{(2)} \rangle\right| \leqslant \frac{\pi}{\Delta(\zeta_1,\zeta_2)} N^{p+q-2}. \tag{3.17}$$

Combining (3.15) with (3.17) we obtain:

$$|\langle \mathbf{z}_1, \mathbf{z}_2 \rangle| = \frac{|\langle \mathbf{u}_p^{(1)}, \mathbf{u}_q^{(2)} \rangle|}{\|\mathbf{u}_p^{(1)}\|\|\mathbf{u}_q^{(2)}\|} \leqslant \frac{\sqrt{(2s_1-1)(2s_2-1)}\pi N^{p+q-2}}{N^{p-\frac{1}{2}} N^{q-\frac{1}{2}} \Delta(\zeta_1,\zeta_2)},$$

finishing the proof with $C_{19}(s_1,s_2) = \pi\sqrt{(2s_1-1)(2s_2-1)}$. □

3.4. **Proof of Theorem 2.1.** Let $\mathcal{X}, s_1, h^{(1)}, \mathcal{Y}, s_2, h^{(2)}, \theta, h$ be as stated in Theorem 2.1. Let N be such that $C_1 \leqslant N \leqslant \frac{C_2}{h}$, where $C_1 = C_1(s_1,s_2)$, $C_2 = C_2(s_1,s_2)$, will be specified within the proof.

Let $\mathbf{v} \in L_1 = L(\mathcal{X}, N)$ and $\mathbf{u} \in L_2 = L(\mathcal{Y}, N)$ be two unit vectors. We will show that

$$|\langle \mathbf{v}, \mathbf{u} \rangle| \leqslant \frac{C_{20}(s_1,s_2)}{N\theta} + C_{21}(s_1,s_2) N h, \tag{3.18}$$

where $C_{20} = C_{20}(s_1,s_2)$ and $C_{21} = C_{21}(s_1,s_2)$ will be specified during the proof. Now, (3.18) implies that

$$\angle_{\min}(L_1,L_2) \geqslant \frac{\pi}{2} - \frac{\pi}{2}\frac{C_{20}(s_1,s_2)}{N\theta} - \frac{\pi}{2} C_{21}(s_1,s_2) N h,$$

thus proving (2.1) with $C_4 = \frac{\pi}{2} C_{20}$ and $C_3 = \frac{\pi}{2} C_{21}$.

Recalling Definitions 3.2 and 3.3, let

$$\begin{aligned}
\widetilde{\mathcal{U}}(x_1,N,s_1) &:= \left\{\tilde{\mathbf{u}}_1^{(1)}, \ldots, \tilde{\mathbf{u}}_{s_1}^{(1)}\right\}, & \mathbf{U}_1 &:= [\tilde{\mathbf{u}}_1^{(1)}, \ldots, \tilde{\mathbf{u}}_{s_1}^{(1)}];\\
\widetilde{\mathcal{U}}(y_1,N,s_2) &:= \left\{\tilde{\mathbf{u}}_1^{(2)}, \ldots, \tilde{\mathbf{u}}_{s_2}^{(2)}\right\}, & \mathbf{U}_2 &:= [\tilde{\mathbf{u}}_1^{(2)}, \ldots, \tilde{\mathbf{u}}_{s_2}^{(2)}];\\
\widetilde{\mathcal{W}}(\mathcal{X}) &:= \left\{\tilde{\mathbf{w}}_1^{(1)}, \ldots, \tilde{\mathbf{w}}_{s_1}^{(1)}\right\}, & \mathbf{W}_1 &:= [\tilde{\mathbf{w}}_1^{(1)}, \ldots, \tilde{\mathbf{w}}_{s_1}^{(1)}];\\
\widetilde{\mathcal{W}}(\mathcal{Y}) &:= \left\{\tilde{\mathbf{w}}_1^{(2)}, \ldots, \tilde{\mathbf{w}}_{s_2}^{(2)}\right\}, & \mathbf{W}_2 &:= [\tilde{\mathbf{w}}_1^{(2)}, \ldots, \tilde{\mathbf{w}}_{s_2}^{(2)}].
\end{aligned}$$

Furthermore, put

$$\mathbf{W}_1 = \mathbf{U}_1 + \mathbf{E}_1, \qquad \mathbf{W}_2 = \mathbf{U}_2 + \mathbf{E}_2. \tag{3.19}$$

Then by (3.6) and Lemma A.1,

$$\|\mathbf{E}_j\| \leqslant \|\mathbf{E}_j\|_F \leqslant \sqrt{s} C_{18} N h, \quad j \in \{1,2\} \tag{3.20}$$

for $s = \max(s_1,s_2)$. In addition, since the columns of $\mathbf{U}_1$ and $\mathbf{U}_2$ have unit length, we also have that

$$\|\mathbf{U}_j\| \leqslant \|\mathbf{U}_j\|_F \leqslant \sqrt{s}, \quad j \in \{1,2\}. \tag{3.21}$$

We now represent $\mathbf{v}$ using the basis $\mathbf{W}_1$ and $\mathbf{u}$ using the basis $\mathbf{W}_2$ as follows:

$$\mathbf{v} = \mathbf{W}_1 \mathbf{a}, \qquad \mathbf{u} = \mathbf{W}_2 \mathbf{b}.$$

Then using (3.20) and (3.21), and assuming $Nh < 1$, we get that

$$\begin{aligned}|\langle \mathbf{v}, \mathbf{u}\rangle| &= \left|\mathbf{b}^H \mathbf{W}_2^H \mathbf{W}_1 \mathbf{a}\right| \\ &= \left|\mathbf{b}^H (\mathbf{U}_2 + \mathbf{E}_2)^H (\mathbf{U}_1 + \mathbf{E}_1)\mathbf{a}\right| \\ &\leqslant |\mathbf{b}^H \mathbf{U}_2^H \mathbf{U}_1 \mathbf{a}| + |\mathbf{b}^H \mathbf{U}_2^H \mathbf{E}_1 \mathbf{a}| + |\mathbf{b}^H \mathbf{E}_2^H \mathbf{U}_1 \mathbf{a}| + |\mathbf{b}^H \mathbf{E}_2^H \mathbf{E}_1 \mathbf{a}| \\ &\leqslant |\mathbf{b}^H \mathbf{U}_2^H \mathbf{U}_1 \mathbf{a}| + \|\mathbf{b}\|\|\mathbf{a}\| s C_{18} Nh + \|\mathbf{b}\|\|\mathbf{a}\| s C_{18} Nh + \|\mathbf{b}\|\|\mathbf{a}\| s C_{18}^2 (Nh)^2 \\ &\leqslant |\mathbf{b}^H \mathbf{U}_2^H \mathbf{U}_1 \mathbf{a}| + C_{22}(s)\|\mathbf{b}\|\|\mathbf{a}\| Nh.\end{aligned} \tag{3.22}$$

By Proposition 3.2 we have that

$$\sigma_{\min}(\mathbf{U}_1) \geqslant \Xi(s), \qquad \sigma_{\min}(\mathbf{U}_2) \geqslant \Xi(s), \tag{3.23}$$

provided $N \geqslant N_1(s)$, where Ξ and N_1 are the constants defined[4] in Proposition 3.2.
Assume that N and h satisfy

$$Nh \leqslant \min\left(1, \frac{\Xi(s)}{2\sqrt{s}C_{18}}\right) = C_{23}(s). \tag{3.24}$$

Then, using (3.19), (3.20), (3.24) and (3.23), and applying the standard singular value perturbation bound, we get

$$\begin{aligned}\sigma_{\min}(\mathbf{W}_1) &= \sigma_{\min}(\mathbf{U}_1 + \mathbf{E}_1) \geqslant \sigma_{\min}(\mathbf{U}_1) - \|\mathbf{E}_1\| \geqslant \frac{\Xi(s)}{2}, \\ \sigma_{\min}(\mathbf{W}_2) &= \sigma_{\min}(\mathbf{U}_2 + \mathbf{E}_2) \geqslant \sigma_{\min}(\mathbf{U}_2) - \|\mathbf{E}_2\| \geqslant \frac{\Xi(s)}{2}.\end{aligned} \tag{3.25}$$

By (3.25) we conclude that

$$\begin{aligned}\|\mathbf{a}\| &\leqslant \|\mathbf{W}_1^\dagger\|\|\mathbf{v}\| \leqslant 2\Xi^{-1}(s), \\ \|\mathbf{b}\| &\leqslant \|\mathbf{W}_2^\dagger\|\|\mathbf{u}\| \leqslant 2\Xi^{-1}(s).\end{aligned} \tag{3.26}$$

Now combining (3.22), (3.26) and (3.14), we get that

$$\begin{aligned}|\langle \mathbf{v}, \mathbf{u}\rangle| &\leqslant |\mathbf{b}^H \mathbf{U}_2^H \mathbf{U}_1 \mathbf{a}| + 4C_{22}\Xi^{-2}(s)Nh \\ &\leqslant \frac{4s\Xi^{-2}(s)C_{19}(s_1, s_2)}{N\theta} + 4C_{22}\Xi^{-2}(s)Nh,\end{aligned}$$

which proves (3.18) with $C_{20} = 4s\Xi^{-2}(s)C_{19}(s_1, s_2)$, and $C_{21} = 4C_{22}\Xi^{-2}(s)$.
Collecting the assumptions we have made along the proof, regarding the range of N for which the intermediate claims hold, we required:

- $N \leqslant \frac{C_{23}}{h}$, this assumption was used in (3.24) in order to establish (3.26) and (3.22).
- $N \geqslant N_1(s)$, this assumption was used to establish (3.23).

Therefore, we have proved Theorem 2.1 with C_4 and C_3 as above and with $C_1 = N_1(s)$ and $C_2 = C_{23}$. □

[4]Here we used the fact the $N_1(s)$ is increasing in s and $\Xi(s)$ is decreasing in s (see Proposition 3.2).

4. Proof of Theorem 2.3

Given nodes $\mathcal{X}=\{x_1,\dots,x_s\}$ and $N=2M$ where M is an integer, define

$$\tilde{\mathbf{V}}_N(\mathcal{X}):=\frac{1}{\sqrt{N}}\mathbf{V}_N\times\operatorname{diag}\left\{e^{-\imath M x_j}\right\}_{j=1,\dots,s}=\frac{1}{\sqrt{2M}}\left[\exp\left(\imath k x_j\right)\right]_{k=-M,\dots,M}^{j=1,\dots,s}.$$

Also, let

$$\mathbf{G}_N=\mathbf{G}_N(\mathcal{X}):=\tilde{\mathbf{V}}_N(\mathcal{X})^H\tilde{\mathbf{V}}_N(\mathcal{X})=\frac{1}{2M}\left[\mathcal{D}_M\left(x_i-x_j\right)\right]_{i,j},$$

where $\mathcal{D}_M$ is the Dirichlet kernel of order M:

$$\mathcal{D}_M(t):=\sum_{k=-M}^{M}\exp(\imath k t)=\begin{cases}\frac{\sin\left(\left(M+\frac{1}{2}\right)t\right)}{\sin\frac{t}{2}} & t\notin 2\pi\mathbb{Z}\\ 2M+1 & \text{else.}\end{cases}$$

Therefore,

$$\sigma_j\left(\mathbf{V}_N\left(\mathcal{X}\right)\right)=N^{\frac{1}{2}}\sigma_j\left(\tilde{\mathbf{V}}_N\left(\mathcal{X}\right)\right)=\sqrt{N\lambda_j\left(\mathbf{G}_N\right)},\quad j=1,\dots,s. \tag{4.1}$$

Put $\varepsilon:=Mh=\frac{Nh}{2}$. Further, put $\{y_1,\dots,y_s\}=\mathcal{Y}:=\frac{1}{h}\mathcal{X}$. Then we have

$$\mathbf{G}_N=\frac{1}{2M}\left[\mathcal{D}_M\left(\varepsilon\frac{(y_i-y_j)}{M}\right)\right]_{1\leqslant i,j\leqslant s},$$

where $\tau\leqslant|y_i-y_j|\leqslant 1$ for $i\neq j$.

The following is essentially a variation of [44, Theorem 8], suitable for our setting.

Denote by $\boldsymbol{\mathcal{D}}=\boldsymbol{\mathcal{D}}(\mathcal{Y})$ the distance matrix $\boldsymbol{\mathcal{D}}=\left[y_i-y_j\right]_{i,j}$, and $\boldsymbol{\mathcal{D}}^k=\left[(y_i-y_j)^k\right]_{i,j}$ the element-wise powers of $\boldsymbol{\mathcal{D}}$.

Next, define for $m=0,1,\dots,s-1$ the $(m+1)\times s$ Vandermonde matrices

$$\mathbf{P}_m=\mathbf{P}_m(\mathcal{Y})=\left[y_j^k\right]_{k=0,\dots,m}^{j=1,\dots,s}.$$

Since the elements of $\{y_j\}$ are pairwise different, the matrices $\mathbf{P}_m$ have full row rank (equal to $m+1$). Thus, $\dim\ker\mathbf{P}_m=s-1-m$, and furthermore, with $\mathbf{P}_{-1}:=0$,

$$\{\mathbf{0}\}=\ker\mathbf{P}_{s-1}\subset\ker\mathbf{P}_{s-2}\subset\cdots\subset\ker\mathbf{P}_0\subset\ker\mathbf{P}_{-1}\equiv\mathbb{R}^s.$$

The following key result is precisely the well-known Micchelli lemma.

Lemma 4.1 (Lemma 3.1 in [35]). *Let $m=0,1,\dots,s-1$. If $\mathbf{a}\in\ker\mathbf{P}_{m-1}$ then*

$$(-1)^m\mathbf{a}^T\boldsymbol{\mathcal{D}}^{2m}\mathbf{a}\geqslant 0, \tag{4.2}$$

while equality holds if and only if $\mathbf{a}\in\ker\mathbf{P}_m$.

It can be readily checked that the Taylor expansion of the normalized Dirichlet kernel at the origin is

$$\frac{1}{2M}\mathcal{D}_M\left(\frac{t}{M}\right)=\sum_{k=0}^{\infty}(-1)^k\frac{F(M,k)}{(2k)!}t^{2k},\quad F(M,k):=\frac{1}{2M^{2k+1}}\sum_{m=-M}^{M}m^{2k}.$$

Note that by (C.2) we have

$$F(M,k)\in\left[\frac{1}{2k+1},1\right]. \tag{4.3}$$

Consider the following quadratic form

$$\mathbf{a}^T \mathbf{G}_N \mathbf{a} = \sum_{k=0}^{\infty} (-1)^k \frac{F(M,k)\varepsilon^{2k}}{(2k)!} \mathbf{a}^T \boldsymbol{\mathcal{D}}^{2k} \mathbf{a}. \tag{4.4}$$

Let $m \in \{0, 1, \ldots, s-1\}$. If $\mathbf{a} \in \ker \mathbf{P}_{m-1}$ and $\|\mathbf{a}\| = 1$, then by Theorem A.1 (i.e. (A.2) applied to $L = \ker \mathbf{P}_{m-1}$)

$$\lambda_{m+1}(\mathbf{G}_N) \leqslant \sum_{k=m}^{\infty} \frac{\varepsilon^{2k}}{(2k)!} \left| \mathbf{a}^T \boldsymbol{\mathcal{D}}^{2k} \mathbf{a} \right|. \tag{4.5}$$

For every $k \in \mathbb{N}$, the entries of $\boldsymbol{\mathcal{D}}^{2k}$ are bounded from above by 1, therefore Lemma A.1 yields

$$|\mathbf{a}^T \boldsymbol{\mathcal{D}}^{2k} \mathbf{a}| \leqslant \|\boldsymbol{\mathcal{D}}^{2k}\| \leqslant \|\boldsymbol{\mathcal{D}}^{2k}\|_F \leqslant s.$$

Now suppose that $\varepsilon < 1$, then clearly

$$\lambda_{m+1}(\mathbf{G}_N) \leqslant \sum_{k=m}^{\infty} \frac{\varepsilon^{2k}}{(2k)!} |\mathbf{a}^T \boldsymbol{\mathcal{D}}^{2k} \mathbf{a}| \leqslant se\varepsilon^{2m} \tag{4.6}$$

Let $\ker \mathbf{P}_{m-1} = \ker \mathbf{P}_m \oplus \mathcal{M}_m$ (i.e. $\mathcal{M}_m$ is the orthogonal complement of $\ker \mathbf{P}_m$ in $\ker \mathbf{P}_{m-1}$). Clearly $\dim \mathcal{M}_m = 1$. Now if $\mathbf{a} \in \mathcal{M}_m$, $\|\mathbf{a}\| = 1$ then (4.2) holds with strict inequality. Applying (A.1) with $L = \bigoplus_{k=0}^{m} \mathcal{M}_k$ to (4.4) in this case we have

$$\lambda_{m+1}(\mathbf{G}_N) \geqslant \min_{\mathbf{a} \in \mathcal{M}_m,\ \|\mathbf{a}\|=1} \left\{ \frac{F(M,m)\varepsilon^{2m}}{(2m)!} \left| \mathbf{a}^T \boldsymbol{\mathcal{D}}^{2m} \mathbf{a} \right| - \sum_{k=m+1}^{\infty} \frac{\varepsilon^{2k}}{(2k)!} \left| \mathbf{a}^T \boldsymbol{\mathcal{D}}^{2k} \mathbf{a} \right| \right\}. \tag{4.7}$$

Define

$$\mathcal{Y}(\tau, s) := \{\mathcal{Y} = \{y_1, \ldots, y_s\} : \quad \tau \leqslant |y_i - y_j| \leqslant 1 \text{ for } i \neq j\}.$$

Next, the following minimum exists:

$$C_{24} = C_{24}(\tau, m, s) := \min_{\mathcal{Y} \in \mathcal{Y}(\tau,s)} \ \min_{\mathbf{a} \in \mathcal{M}_m(\mathcal{Y}),\ \|\mathbf{a}\|=1} \left| \mathbf{a}^T \boldsymbol{\mathcal{D}}^{2m}(\mathcal{Y}) \mathbf{a} \right| > 0.$$

Substituting (4.6) with $m+1$ in place of m into (4.7) and applying (4.3) we have

$$\lambda_{m+1}(\mathbf{G}_N) \geqslant \varepsilon^{2m} \left(\frac{C_{24}}{(2m+1)!} - se\varepsilon^2 \right).$$

Thus for $Nh = 2\varepsilon < \min\left(2, \sqrt{\frac{2C_{24}}{se(2m+1)!}}\right)$ we obtain the lower bound

$$\lambda_{m+1}(\mathbf{G}_N) \geqslant C_{25}\varepsilon^{2m}, \quad C_{25} = \frac{C_{24}}{2(2m+1)!}. \tag{4.8}$$

Combining (4.1), (4.6) and (4.8), we conclude that (2.3) holds with

$$C_9(\tau, s) := \min\left(2, \frac{2}{\sqrt{se}} \cdot \min_{0 \leqslant m < s} C_{25}(\tau, m, s)^{\frac{1}{2}}\right),$$
$$C_{10}(\tau, s) := \min_{0 \leqslant m < s} C_{25}(\tau, m, s)^{\frac{1}{2}},$$
$$C_{11}(s) := \sqrt{se}.$$

5. Spectral properties of multi-cluster Vandermonde matrices

5.1. **Singular values of nearly orthogonal spaces.** In this section we consider an $N \times s$ matrix $\mathbf{A}$ whose columns are partitioned as $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_M]$, and with the blocks $\mathbf{A}_j$ having the following property: Let L_j be the subspace spanned be the columns of the sub-matrix $\mathbf{A}_j$. We consider the case where the minimal principle angle between each pair of subspaces L_j, L_k is large:

$$\angle_{\min}(L_j, L_k) \geqslant \frac{\pi}{2} - \alpha,$$

for some "small enough" α.

We show below that in this case, the singular values of $\mathbf{A}$ are given by a multiplicative perturbation of the singular values of all the sub-matrices $\mathbf{A}_j$, the size of the multiplicative factor of the perturbation γ is $\sqrt{1-s\alpha} \leqslant \gamma \leqslant \sqrt{1+s\alpha}$.

Lemma 5.1. *Let $\mathbf{A} \in \mathbb{C}^{N\times s}$, $N \geqslant s$, such that $\mathbf{A}$ is given in the following block form*

$$\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_M],$$

with $\mathbf{A}_j \in \mathbb{C}^{N\times s_j}$ and $\sum_{j=1}^{M} s_j = s$. Let $L_j \subset \mathbb{C}^N$ be the subspace spanned by the columns of the sub-matrix $\mathbf{A}_j$. Assume that for all $1 \leqslant j,k \leqslant M$, $j \neq k$, and $0 \leqslant \alpha \leqslant \frac{1}{s}$,

$$\angle_{\min}(L_j, L_k) \geqslant \frac{\pi}{2} - \alpha. \tag{5.1}$$

Then the following hold.

(1) For each $j = 1, \dots, M$, let $\mathbf{A}_j = \mathbf{Q}_j\mathbf{R}_j$ be the QR-decomposition of $\mathbf{A}_j$, where $\mathbf{Q}_j \in \mathbb{C}^{N\times s_j}$ has orthonormal columns, and $\mathbf{R}_j \in \mathbb{C}^{s_j\times s_j}$ is upper triangular. Write

$$\mathbf{A} = \mathbf{Q}\mathbf{R} \equiv [\mathbf{Q}_1, \dots, \mathbf{Q}_M]\operatorname{diag}(\mathbf{R}_1, \dots, \mathbf{R}_M), \tag{5.2}$$

where $\operatorname{diag}(\mathbf{R}_1, \dots, \mathbf{R}_M) \in \mathbb{C}^{s\times s}$ is a block diagonal matrix whose diagonal blocks are $\mathbf{R}_1, \dots, \mathbf{R}_M$. Then

$$\sqrt{1-s\alpha} \leqslant \sigma_{\min}(\mathbf{Q}) \leqslant \sigma_{\max}(\mathbf{Q}) \leqslant \sqrt{1+s\alpha}. \tag{5.3}$$

(2) Let

$$\sigma_1 \geqslant \dots \geqslant \sigma_s$$

be the ordered collection of all the singular values of $\mathbf{A}$, and let

$$\tilde{\sigma}_1 \geqslant \dots \geqslant \tilde{\sigma}_s$$

be the ordered collection of all the singular values of the sub-matrices $\{\mathbf{A}_j\}$. Then

$$\sqrt{1-s\alpha}\,\tilde{\sigma}_j \leqslant \sigma_j \leqslant \sqrt{1+s\alpha}\,\tilde{\sigma}_j \qquad j = 1, \dots, s. \tag{5.4}$$

Proof. First we argue that the singular values of $\mathbf{R} = \operatorname{diag}(\mathbf{R}_1, \dots, \mathbf{R}_M)$ are given by $\tilde{\sigma}_1 \geqslant \dots \geqslant \tilde{\sigma}_s$.

Indeed we have that the singular values of $\mathbf{R}$ are given by the union of the singular values of its diagonal blocks $\mathbf{R}_1, \dots, \mathbf{R}_M$. From the other hand, for each j, the singular values of $\mathbf{A}_j$ are equal to the singular values of $\mathbf{R}_j$ (and this is true since $\mathbf{Q}_j$ is an orthogonal matrix). Therefore the singular values of $\mathbf{R}$, ordered according to their magnitude, are exactly

$$\tilde{\sigma}_1 \geqslant \dots \geqslant \tilde{\sigma}_s. \tag{5.5}$$

Put $\mathbf{Q} = [\mathbf{Q}_1, \dots, \mathbf{Q}_M]$. Next we show that $\sigma_{\max}(\mathbf{Q}) \leqslant \sqrt{1+s\alpha}$ and that $\sigma_{\min}(\mathbf{Q}) \geqslant \sqrt{1-s\alpha}$.

We write the Gramian matrix

$$\mathbf{Q}^H\mathbf{Q} = [\mathbf{Q}_1^H; \dots; \mathbf{Q}_M^H][\mathbf{Q}_1, \dots, \mathbf{Q}_M] = \begin{bmatrix} \mathbf{Q}_1^H\mathbf{Q}_1 & \mathbf{Q}_1^H\mathbf{Q}_2 & \dots & \mathbf{Q}_1^H\mathbf{Q}_M \\ \mathbf{Q}_2^H\mathbf{Q}_1 & \mathbf{Q}_2^H\mathbf{Q}_2 & \dots & \mathbf{Q}_2^H\mathbf{Q}_1 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{Q}_M^H\mathbf{Q}_1 & \mathbf{Q}_M^H\mathbf{Q}_2 & \dots & \mathbf{Q}_M^H\mathbf{Q}_M \end{bmatrix}.$$

The off-diagonal blocks are made out of inner products of unit vectors from different subspaces L_j. By (5.1), for each pair of unit vectors $\mathbf{v} \in L_j$ and $\mathbf{u} \in L_k$, $j \neq k$,

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leqslant \cos\left(\frac{\pi}{2} - \alpha\right) = \sin(\alpha) \leqslant \alpha.$$

Therefore the absolute value of each entry in the off-diagonal blocks is less than α.

On the other hand, for each $j = 1, \dots, M$, the diagonal block $\mathbf{Q}_j^H\mathbf{Q}_j = \mathbf{I}_{s_j}$, where $\mathbf{I}_{s_j}$ is the $s_j \times s_j$ identity matrix. We can therefore write $\mathbf{Q}^H\mathbf{Q}$ as

$$\mathbf{Q}^H\mathbf{Q} = \mathbf{I}_s + \mathbf{E}, \tag{5.6}$$

where $\mathbf{I}_s$ is the $s \times s$ identity matrix, and $\mathbf{E} \in \mathbb{C}^{s\times s}$ is an Hermitian matrix, the absolute value of each one of its entries is bounded by α. Therefore

$$\lambda_{\min}(\mathbf{E}) \geqslant -\|\mathbf{E}\| \geqslant -\|\mathbf{E}\|_F \geqslant -s\alpha, \tag{5.7}$$

$$\lambda_{\max}(\mathbf{E}) \leqslant \|\mathbf{E}\| \leqslant \|\mathbf{E}\|_F \leqslant s\alpha. \tag{5.8}$$

Now using Weyl's perturbation inequality on the perturbation (5.6) and the bounds (5.7) and (5.8), we have that

$$\sigma_{\min}(\mathbf{Q}) = \sqrt{\lambda_{\min}(\mathbf{Q})} \geqslant \sqrt{1 - s\alpha}, \tag{5.9}$$

$$\sigma_{\max}(\mathbf{Q}) = \sqrt{\lambda_{\max}(\mathbf{Q})} \leqslant \sqrt{1 + s\alpha}. \tag{5.10}$$

We conclude that according to (5.2), (5.5), (5.9) and (5.10), $\mathbf{A}$ can be written as follows:

$$\mathbf{A} = \mathbf{Q}\mathbf{R},$$

where the minimal and maximal singular values of $\mathbf{Q}$ are bounded as in (5.9) and (5.10), and the singular values of $\mathbf{R}$ are exactly $\tilde{\sigma}_1 \geqslant \dots \geqslant \tilde{\sigma}_s$. The proof of Lemma 5.1 is then completed by invoking Lemma A.2 with $\mathbf{C} = \mathbf{A}$, $\mathbf{B} = \mathbf{Q}$ and $\mathbf{A} = \mathbf{R}$ (on the left side are the matrices of Lemma A.2). □

5.2. **Multi-cluster subspace angles.**

Proposition 5.1 (Multi-cluster subspace angles). *Suppose that $\mathcal{X} = \{x_1, \dots, x_s\} \subset (-\pi, \pi]$ forms an $((h^{(j)}, s^{(j)})_{j=1}^M, \theta)$-clustered configuration, and put $h = \max_j(h^{(j)})$. Then there exist constants C_{26}, C_{27}, C_{28} and C_{29}, depending only on $s^{(1)}, \dots, s^{(M)}$, such that for $\frac{C_{26}}{\theta} \leqslant N \leqslant \frac{C_{27}}{h}$ we have*

$$\angle_{\min}(L(\mathcal{C}^{(j)}, N), L(\mathcal{C}^{(k)}, N)) \geqslant \frac{\pi}{2} - \alpha, \quad \alpha := \frac{C_{28}}{N\theta} + C_{29}Nh \leqslant \frac{1}{s}, \qquad 1 \leqslant j < k \leqslant M. \tag{5.11}$$

Proof. (5.11) immediately follows from Theorem 2.1 with

$$C_{28} = \max_{1 \leqslant j < k \leqslant M} C_4(s^{(j)}, s^{(k)}), \tag{5.12}$$

$$C_{29} = \max_{1 \leqslant j < k \leqslant M} C_3(s^{(j)}, s^{(k)}), \tag{5.13}$$

$$C_{26} := \max\left(\pi \left\{\max_{1 \leqslant j < k \leqslant M} C_1(s^{(j)}, s^{(k)})\right\}, 2sC_{28}\right), \tag{5.14}$$

$$C_{27} := \min\left(\left\{\min_{1 \leqslant j < k \leqslant M} C_2(s^{(j)}, s^{(k)})\right\}, \frac{1}{2sC_{29}}\right). \tag{5.15}$$

Here C_1, C_2, C_4 and C_3, are the constants specified in Theorem 2.1. □

5.3. **Proof of Theorem 2.2.** Let $\mathcal{X}$ form an $((h^{(j)}, s^{(j)})_{j=1}^{M}, \theta)$-clustered configuration and put $h = \max_j(h^{(j)})$. Let $\sigma_1 \geqslant \ldots \geqslant \sigma_s$ and $\tilde{\sigma}_1 \geqslant \ldots \geqslant \tilde{\sigma}_s$ be as specified in Theorem 2.2. Without loss of generality we assume that $\mathbf{V}_N(\mathcal{X})$ is organized in block form, according to the clusters, as follows:

$$\mathbf{V}_N(\mathcal{X}) = \left[\mathbf{V}_N(\mathcal{C}^{(1)}), \ldots, \mathbf{V}_N(\mathcal{C}^{(M)})\right]. \tag{5.16}$$

By Proposition 5.1 we have the estimate (5.11) for

$$\frac{C_{26}}{\theta} \leqslant N \leqslant \frac{C_{27}}{h},$$

when furthermore $\alpha \leqslant \frac{1}{s}$. Now we invoke Lemma 5.1 with $\mathbf{A} = \mathbf{V}_N(\mathcal{X})$, $\mathbf{A}_j = \mathbf{V}_N(\mathcal{C}^{(j)})$ and α as above and get that

$$(1 - s\alpha)^{\frac{1}{2}}\, \tilde{\sigma}_j \leqslant \sigma_j \leqslant (1 + s\alpha)^{\frac{1}{2}}\, \tilde{\sigma}_j \qquad\qquad j = 1, \ldots, s,$$

thus proving Theorem 2.2 with $C_5 = C_{26}$, $C_6 = C_{27}$, $C_7 = C_{28}s$ and $C_8 = C_{29}s$.

5.4. **Proof of Theorem 2.4.** Let $\mathbf{A} \in \mathbb{C}^{m \times n}$ be a matrix. We use the following notations:

- $\|\mathbf{A}\|_{k,1}$ for the ℓ_1 norm of the k-th row of $\mathbf{A}$, i.e.

$$\|\mathbf{A}\|_{k,1} := \sum_{\ell=1}^{n} \left|(\mathbf{A})_{k,\ell}\right|, \quad k \in \{1, \ldots, m\};$$

- $\|\mathbf{A}\|_{k,\max}$ for the maximum norm of the k-th row of $\mathbf{A}$, i.e.

$$\|\mathbf{A}\|_{k,\max} := \max_{1 \leqslant \ell \leqslant n} \left|(\mathbf{A})_{k,\ell}\right|, \quad k \in \{1, \ldots, m\}.$$

Lemma 5.2. *Let* $\mathbf{A} = \mathbf{BC} \in \mathbb{C}^{m \times n}$, *where* $\mathbf{C} \in \mathbb{C}^{p \times n}$. *Then*

$$\|\mathbf{A}\|_{k,1} \leqslant \sqrt{pn}\|\mathbf{B}\|_{k,\max}\|\mathbf{C}\|_F. \tag{5.17}$$

Proof. We have

$$\begin{aligned}
\|\mathbf{A}\|_{k,1} = \sum_{\ell=1}^{n} \left|(\mathbf{A})_{k,\ell}\right| &= \sum_{\ell=1}^{n} \left|\sum_{j=1}^{p} (\mathbf{B})_{k,j} (\mathbf{C})_{j,\ell}\right| \\
&\leqslant \|\mathbf{B}\|_{k,\max} \sum_{j,\ell} \left|(\mathbf{C})_{j,\ell}\right| \\
&\leqslant \sqrt{pn}\|\mathbf{B}\|_{k,\max}\|\mathbf{C}\|_F,
\end{aligned}$$

where the last transition is just the Hölder's inequality. □

First we prove the following estimate.

Proposition 5.2 (Pseudoinverse row norms). *Suppose that the node set $\mathcal{X} = \{x_1, \ldots, x_s\} \subset (-\pi, \pi]$ forms an $((h^{(j)}, \tau^{(j)}, s^{(j)})_{j=1}^{M}, \theta)$-clustered configuration, with $h = \max_{1 \leqslant j \leqslant M}(h^{(j)})$. Then there exist constants $C_{30}(s^{(1)}, \ldots, s^{(M)})$ and $C_{31}(s^{(1)}, \ldots, s^{(M)})$ such that for all $\frac{C_{30}}{\theta} \leqslant N \leqslant \frac{C_{31}}{h}$ and each $j \in \{1, 2, \ldots, M\}$ there exists a constant $C_{32}(s^{(j)}, \tau^{(j)})$ such that for all $\ell \in C_j(\mathcal{X})$ (recall Definition 2.5) we have*

$$\|\mathbf{V}_N^{\dagger}(\mathcal{X})\|_{\ell,1} \leqslant C_{32} s \left(\frac{1}{N h^{(j)}}\right)^{s^{(j)}-1}. \tag{5.18}$$

Proof. Again, applying Proposition 5.1 and then Lemma 5.1 to the matrix $\mathbf{V}_N(\mathcal{X})$ assumed to be in the block form (5.16), we obtain the block QR-decomposition (5.2) $\mathbf{V}_N(\mathcal{X}) = \mathbf{Q}\mathbf{R}$. Since $\mathbf{Q} \in \mathbb{C}^{(N+1)\times s}$ has full column rank, and $\mathbf{R} \in \mathbb{C}^{s \times s}$ is invertible, we have

$$\mathbf{V}_N^{\dagger} = \mathbf{R}^{-1}\mathbf{Q}^{\dagger}. \tag{5.19}$$

Using (5.3), we have

$$\|\mathbf{Q}^{\dagger}\| = \sigma_{\max}\left(\mathbf{Q}^{\dagger}\right) = \sigma_{\min}^{-1}(\mathbf{Q}) \leqslant (1 - s\alpha)^{-\frac{1}{2}}, \tag{5.20}$$

where $\alpha = \frac{C_{28}}{N\theta} + C_{29} N h$ as provided by Proposition 5.1.

On the other hand, each one of the blocks $\mathbf{R}_j$ has its singular values exactly equal to the singular values of $\mathbf{V}_N\left(\mathcal{C}^{(j)}\right)$. By Theorem 2.3, the smallest one scales like $N^{\frac{1}{2}}\left(N h^{(j)}\right)^{s^{(j)}-1}$, and therefore for some constant $C_{33}(s^{(j)}, \tau^{(j)})$ we have, using Lemma A.1,

$$\|\mathbf{R}_j^{-1}\|_{\max} \leqslant \|\mathbf{R}_j^{-1}\| = \sigma_{\min}^{-1}(\mathbf{R}_j) \leqslant C_{33} \frac{1}{\sqrt{N}} \left(\frac{1}{N h^{(j)}}\right)^{s^{(j)}-1}.$$

Let $\ell \in C_j(\mathcal{X})$, then by (5.17) applied to (5.19), (5.20) and the fact that $\mathbf{Q}^{\dagger}$ is of rank s, we have that

$$\begin{aligned} \|\mathbf{V}_N^{\dagger}\|_{\ell,1} &\leqslant \sqrt{sN} \|\mathbf{R}_j^{-1}\|_{\max} \|\mathbf{Q}^{\dagger}\|_F \\ &\leqslant s C_{33} (1 - s\alpha)^{-\frac{1}{2}} \left(\frac{1}{N h^{(j)}}\right)^{s^{(j)}-1}. \end{aligned}$$

Here, we used Lemma A.1 once again. If $N\theta > 4sC_{28}$ and $Nh < \frac{1}{4sC_{29}}$ we have $\alpha < \frac{1}{2s}$ and consequently $(1 - s\alpha)^{-\frac{1}{2}} < \sqrt{2}$. This completes the proof of (5.18) with $C_{30}(s^{(1)}, \cdots, s^{(M)}) = \max(C_{26}, 4sC_{28})$, $C_{31}(s^{(1)}, \ldots, s^{(M)}) = \min\left(C_{27}, \frac{1}{4sC_{29}}\right)$ and $C_{32}(s^{(j)}, \tau^{(j)}) = \sqrt{2} C_{33}$. ☐

Proof of Theorem 2.4. From Definition 2.6 we clearly have $\mathbf{a}(\mathcal{X}, \mathbf{b}) = \mathbf{V}_N^{\dagger}(\mathcal{X})\,\mathbf{b}$, and since $\mathbf{a}_0 = \mathbf{V}_N^{\dagger}(\mathcal{X})\mathbf{b}_0$, we obtain by (5.18)

$$\begin{aligned} |(\mathbf{a} - \mathbf{a}_0)_{\ell}| = \left|\left(\mathbf{V}_N^{\dagger}(\mathbf{b} - \mathbf{b}_0)\right)_{\ell}\right| &\leqslant \|\mathbf{V}_N^{\dagger}(\mathcal{X})\|_{\ell,1} \|\mathbf{b} - \mathbf{b}_0\|_{\infty} \\ &\leqslant C_{32} s \left(\frac{1}{N h^{(j)}}\right)^{s^{(j)}-1} \|\mathbf{b} - \mathbf{b}_0\|_{\infty}. \end{aligned}$$

This completes the proof of (2.4) with $C_{14} = C_{30}$, $C_{15} = C_{31}$ and $C_{16} = C_{32}$. ☐

6. Numerical experiments

In this section we provide basic numerical evidence supporting our main results. All calculations were performed using Julia 1.1 with standard packages, in double precision floating point (and sometimes multi-precision).

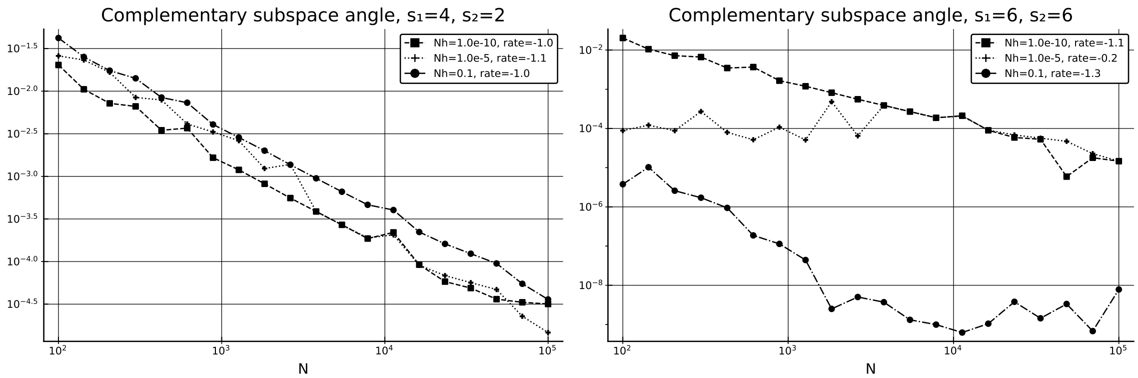

FIGURE 1. Complementary subspace angle β. Nh and θ fixed, varying N.

6.1. **Cluster subspace angles.** We use the notation from Theorem 2.1. In order to compute the minimal principal angle $\angle_{\min}(L_1, L_2)$ between $L_1 = L(\mathcal{X}, N)$ and $L_2 = L(\mathcal{Y}, N)$, we use the standard SVD-based algorithm (see e.g. [14, 29]) which is numerically stable for large angles. We then compute the *complementary angle*

$$\beta\left(L_1, L_2\right) := \frac{\pi}{2} - \angle_{\min}(L_1, L_2).$$

In the experiments, the two clusters were chosen to consist of equispaced nodes with the same cluster size $h^{(1)} = h^{(2)} = h$, and with a prescribed distance θ between the closest nodes.

According to the estimate (2.1) from Theorem 2.1, we have the bound

$$\beta\left(L_1, L_2\right) \leqslant \frac{C_4}{N\theta} + C_3 Nh. \tag{6.1}$$

In the first set of experiments, we kept the value of θ fixed, and changed N, h simultaneously so that the product Nh remained fixed. We chose 3 different values $Nh = 10^{-10}, 10^{-5}, 0.1$. The dependence of β on N is presented in Figure 1. For $s^{(1)} = 4, s^{(2)} = 2$ (left plot) the asymptotic decay $\beta \sim \frac{1}{N}$ is clearly seen, with smaller values of β corresponding to smaller values of Nh. This suggests that $\frac{C_4}{N\theta}$ is indeed the dominant term with respect to $C_3 Nh$ in (6.1). However, for $s^{(1)} = s^{(2)} = 6$ (right plot) we see that when $Nh = 10^{-5}$, the value of β is relatively constant, $\approx 10^{-5}$, while the value of β for $Nh = 0.1$ decays relatively slowly with N (for $Nh = 10^{-10}$ we still have $\beta \sim \frac{1}{N}$). This suggests that the value of Nh is indeed important for controlling the subspace angle in the regime $Nh \ll 1$, however for $Nh = \mathcal{O}(1)$ there must be other factors.

In the second set of experiments, we kept the values of N and θ fixed, while changing h. We chose again 3 different values $\theta = 0.01, 0.1, 1$. The dependence of β on h (or Nh) in this case is shown in Figure 2. Notice that for small enough Nh we indeed see that β approaches a positive value which is proportional to $\frac{1}{N\theta}$, i.e. the dominant role is played by the cluster separation. For increasing values of Nh, the actual cluster subspaces move further away from the limit spaces and therefore this regime is not covered by our theory. However, apparently also in this case β remains small, but this must be due to other factors.

6.2. **Spectra of multi-cluster Vandermonde matrices.** We normalize the Vandermonde matrix and compute the spectra of the matrices $\frac{1}{\sqrt{N}}\mathbf{V}_N(\mathcal{X})$. For each experiment we choose the values of N and h randomly from within a prescribed range. In the multi-cluster setting, we construct 2 nontrivial clusters of same size $h^{(1)} = h^{(2)} = h$, and add zero or more well-separated nodes (so if $s^{(1)} = 2$, $s^{(2)} = 3$ and $s = 7$, there are 2 clusters of multiplicity 1). The values of $\sigma_j(\frac{1}{\sqrt{N}}\mathbf{V}_N(\mathcal{X}))$ for different $\mathcal{X}$ are plotted in Figure 3.

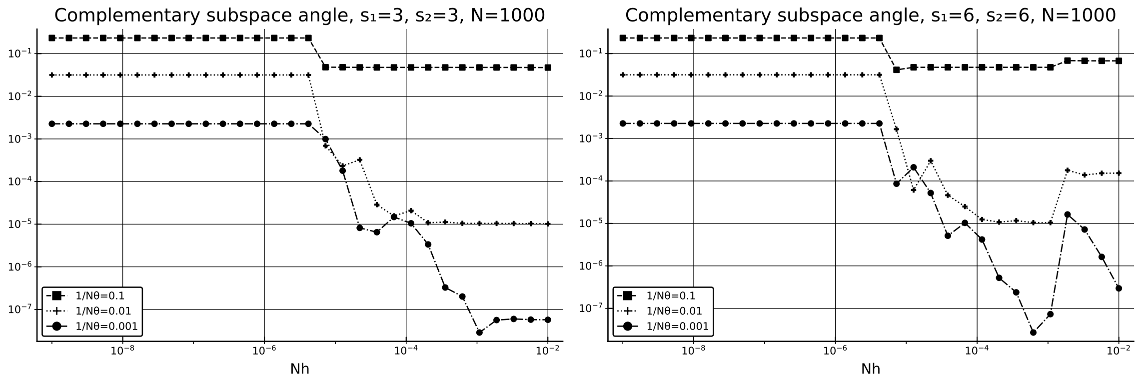

FIGURE 2. Complementary subspace angle β. N, θ fixed, varying h.

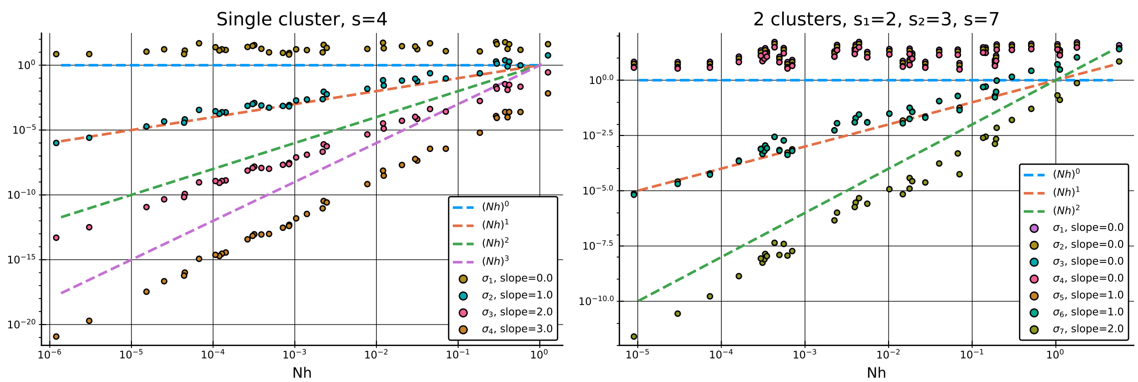

FIGURE 3. Singular values of $\frac{1}{\sqrt{N}}\mathbf{V}_N$ as a function of Nh. **(left)** A single cluster with $s = 4$. **(right)** 2 nontrivial clusters (and 4 overall), multiplicities $= 2, 1, 3, 1$. The slopes of the lines, computed by a linear fit, are written in the respective legend labels.

There is good agreement with Theorem 2.3 and Corollary 2.1. The proportionality constant is seen to be not too large. Furthermore, the minimal value of Nh for which the bounds hold (corresponding to the constants C_9 and C_6 in Theorems 2.3 and 2.2, respectively) is apparently reasonably high.

6.3. **Least squares accuracy.** Here we solve the least squares problem as in Theorem 2.4. In each experiment, we choose N, h, ε uniformly at random within prescribed ranges. This in particular defines $\mathcal{X}$ and $\mathbf{V}_N(\mathcal{X})$. We then choose the entries of the vectors $\mathbf{a}_0$ and $\mathbf{f}$ to be uniformly randomly distributed in $[0, 1]$. Then we put $\mathbf{b}_0 = \mathbf{V}_N(\mathcal{X})\mathbf{a}_0$ and $\mathbf{b} = \mathbf{b}_0 + \varepsilon\mathbf{f}$. We then compute $\mathbf{a} = \mathbf{a}(\mathcal{X}, \mathbf{b})$ as in Definition 2.6. Finally, we set

$$\delta a_\ell := \frac{|(\mathbf{a} - \mathbf{a}_0)_\ell|}{\|\mathbf{b} - \mathbf{b}_0\|_\infty}.$$

We then repeat the experiment multiple times, and plot δa_ℓ for all $\ell = 1, \ldots, s$ as a function of Nh. The results are presented in Figure 4. There is a good agreement with the estimate (2.4) for each component.

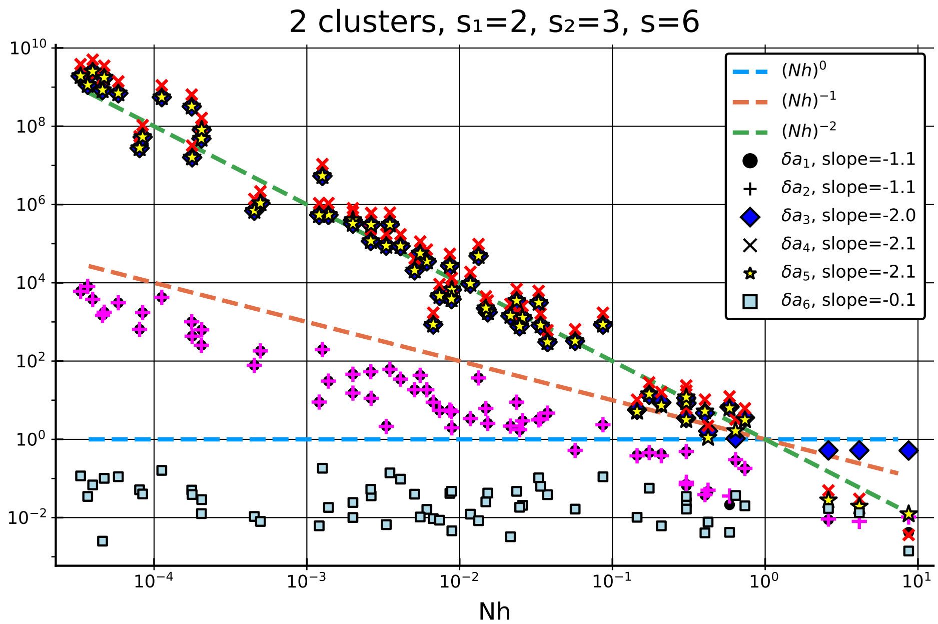

FIGURE 4. Accuracy of least squares reconstruction. 2 nontrivial clusters (and 3 overall), multiplicities $= 2, 3, 1$. The slopes of the lines, computed by a linear fit, are written in the respective legend labels.

Appendix A. Matrix analysis

Here we collect well-known facts from matrix analysis.

Definition A.1 (Matrix Norms). *Let $\mathbf{A} \in \mathbb{C}^{m\times n}$.*

(1) We denote the spectral norm by

$$\|\mathbf{A}\| = \max_{\|\mathbf{a}\|=1} \|\mathbf{A}\mathbf{a}\| = \sigma_{\max}(\mathbf{A}).$$

If $\mathbf{A}$ is Hermitian, we clearly have

$$\|\mathbf{A}\| = \max_{\lambda \text{ eigenvalue of } \mathbf{A}} |\lambda|.$$

(2) The maximum norm is given by

$$\|\mathbf{A}\|_{\max} = \max_{j,k} |(\mathbf{A})_{j,k}|.$$

Note that this is not the operator norm induced by the ∞-norm and not a sub-multiplicative matrix norm.

(3) The Frobenius norm, defined as

$$\|\mathbf{A}\|_F = \left(\sum_{j,k} |(\mathbf{A})_{j,k}|^2 \right)^{\frac{1}{2}}.$$

The following relations are standard and we use them frequently.

Lemma A.1 (Relations of Matrix Norms). *For any $\mathbf{A} \in \mathbb{C}^{m\times n}$ we have*

$$\begin{aligned} \|\mathbf{A}\|_{\max} &\leqslant \|\mathbf{A}\|, \\ \|\mathbf{A}\| &\leqslant \|\mathbf{A}\|_F, \\ \|\mathbf{A}\|_F &\leqslant \sqrt{r}\|\mathbf{A}\|, \qquad \text{if } \mathbf{A} \text{ is of rank } r. \end{aligned}$$

Theorem A.1 (Theorem 4.2.6 in [28], Courant-Fischer minimax principle). *Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be Hermitian and have eigenvalues $\lambda_1(\mathbf{A}) \geqslant \lambda_2(\mathbf{A}) \geqslant \cdots \geqslant \lambda_n(\mathbf{A})$. Let L denote a linear subspace of $\mathbb{C}^n$. Then, for $j = 1, \ldots, n$:*

$$\lambda_j(\mathbf{A}) = \max_{L:\dim L=j} \min_{\|\mathbf{v}\|=1, \mathbf{v}\in L} \mathbf{v}^H\mathbf{A}\mathbf{v} \tag{A.1}$$

$$= \min_{L:\dim L=n-j+1} \max_{\|\mathbf{v}\|=1, \mathbf{v}\in L} \mathbf{v}^H\mathbf{A}\mathbf{v}. \tag{A.2}$$

For any matrix $\mathbf{C} \in \mathbb{C}^{m\times n}$, we will refer to its singular values in a decreasing order and list them as

$$\sigma_{\max}(\mathbf{C}) = \sigma_1(\mathbf{C}) \geqslant \ldots \geqslant \sigma_{\min(m,n)}(\mathbf{C}) = \sigma_{\min}(\mathbf{C}).$$

Lemma A.2. *For $m \geqslant n$, Let $\mathbf{C} \in \mathbb{C}^{m\times n}$, $\mathbf{B} \in \mathbb{C}^{m\times m}$ and $\mathbf{A} \in \mathbb{C}^{m\times n}$ such that $\mathbf{C} = \mathbf{B}\mathbf{A}$. Then the singular values of $\mathbf{C}$ are given by a multiplicative perturbation of the singular values of $\mathbf{A}$ as follows:*

$$\sigma_{\min}(\mathbf{B})\sigma_j(\mathbf{A}) \leqslant \sigma_j(\mathbf{C}) \leqslant \sigma_{\max}(\mathbf{B})\sigma_j(\mathbf{A}), \qquad 1 \leqslant j \leqslant n. \tag{A.3}$$

Proof. The claim is a direct consequence of Theorem A.1. Indeed, let $j \in \{1, \ldots, n\}$, then by applying (A.1) to the matrix $\mathbf{C}^H\mathbf{C}$ we have

$$\sigma_j(\mathbf{C}) = \max_{L:\, \dim(L)=j} \min_{\mathbf{v}\in L:\, \|\mathbf{v}\|=1} \|\mathbf{B}\mathbf{A}\mathbf{v}\| \leqslant \sigma_{\max}(\mathbf{B}) \max_{L:\, \dim(L)=j} \min_{\mathbf{v}\in L:\, \|\mathbf{v}\|=1} \|\mathbf{A}\mathbf{v}\| = \sigma_{\max}(\mathbf{B})\sigma_j(\mathbf{A}),$$

$$\sigma_j(\mathbf{C}) = \max_{L:\,\dim(L)=j} \min_{\mathbf{v}\in L:\,\|\mathbf{v}\|=1} \|\mathbf{BAv}\| \geqslant \sigma_{\min}(\mathbf{B}) \max_{L:\,\dim(L)=j} \min_{\mathbf{v}\in L:\,\|\mathbf{v}\|=1} \|\mathbf{Av}\| = \sigma_{\min}(\mathbf{B})\sigma_j(\mathbf{A}).$$

□

Appendix B. Divided differences

Recall Definition 3.1. The following properties can be found in e.g. [17], see also [7, Section 6.2].

Lemma B.1. *The functionals* $[t_1, \ldots, t_n]$ *satisfy the following.*

(1) $[t_1, \ldots, t_n]f$ *is a symmetric function of* $(t_1, \ldots, t_n)$.
(2) $[t_1, \ldots, t_n]f$ *is a continuous function of* $(t_1, \ldots, t_n)$, *i.e.*

$$\lim_{(t_1,\ldots,t_n)\to(u_1,\ldots,u_n)} [t_1, \ldots, t_n]f = [u_1, \ldots, u_n]f. \tag{B.1}$$

(3) The numbers $[t_1, \ldots, t_n]f$ *can be computed by the recursive rule*

$$[t_1, \ldots, t_n]f = \begin{cases} f(t_1) & n = 1; \\ \frac{[t_2,\ldots,t_n]f-[t_1,\ldots,t_{n-1}]f}{t_n-t_1} & t_1 \neq t_n; \\ \lim_{\xi\to t_n}\left\{\frac{\mathrm{d}}{\mathrm{d}\xi}\left([\xi, t_2, \ldots, t_{n-1}]f\right)\right\} & t_1 = t_n. \end{cases} \tag{B.2}$$

(4) (Chakalov's Expansion): for $\xi \in \{t_j\} := T$, *denote by* $\#\xi := \#\{j : t_j = \xi\}$ *the number of occurences of* ξ *in the sequence* $(t_1, \ldots, t_n)$, *then*

$$[t_1, \ldots, t_n]f = \sum_{\xi\in T} \sum_{0\leqslant j<\#\xi} A_{\xi,\ell} f^{(\ell)}(\xi),$$

where the coefficients $A_{\xi,\ell}$ *are defined by the partial fraction decomposition*

$$\prod_{j=1}^{n} (t - t_j)^{-1} \equiv \sum_{\xi\in T} \sum_{\ell=0}^{\#\xi-1} \frac{A_{\xi,\ell}\,\ell!}{(t - \xi)^{\ell+1}},$$

or directly by Cauchy's residue theorem:

$$A_{\xi,\ell} = \frac{1}{\#\xi - 1 - \ell} \lim_{z\to\xi} \left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^{\#\xi-1-\ell} \left\{\frac{(z - \xi)^{\ell+1}}{\prod_{j=1}^{n}(z - t_j)}\right\}.$$

(5) In particular, if all $\{t_j\}$*'s are distinct, then*

$$[t_1, \ldots, t_n]f = \sum_{j=1}^{n} \frac{f(t_j)}{\prod_{k\neq j}(t_j - t_k)}. \tag{B.3}$$

(6) (Mean value theorem) Let $t_1, \ldots, t_n \in \mathbb{R}$ *and put* $I := [\min_\ell t_\ell, \max_\ell t_\ell]$. *Then*

$$[t_1, \ldots, t_n]f = \frac{f^{(n-1)}(\xi)}{(n - 1)!}, \quad \xi \in I. \tag{B.4}$$

(7) From the above, in particular,

$$[\underbrace{t, t, \ldots, t}_{n\ \textit{times}}]f = \frac{f^{(n-1)}(t)}{(n - 1)!}. \tag{B.5}$$

Appendix C. Power sums

Lemma C.1. *For a positive integer p, the sum of the p^{th} powers of the first $N+1$ non-negative integers is given by Faulhaber's formula*

$$\sum_{k=0}^{N} k^p = \frac{N^{p+1}}{p+1} + \frac{1}{2}N^p + \sum_{k=2}^{p} \frac{B_k}{k!}(p)_{k-1}N^{p-k+1} = \frac{N^{p+1}}{p+1} + \mathcal{O}(N^p), \tag{C.1}$$

where B_k are the Bernoulli numbers, and $(p)_{k-1}$ is the falling factorial, $(p)_{k-1} = \dfrac{p!}{(p-k+1)!}$. We also have the following non-asymptotic bounds:

$$\frac{N^{p+1}}{p+1} = \int_0^N x^p \mathrm{d}x \leqslant \sum_{k=0}^{N} k^p \leqslant N^{p+1}. \tag{C.2}$$

Appendix D. Trigonometric cancellation

Lemma D.1. *For each $z \neq 1$ with $|z| = 1$ and for each $m \in \mathbb{N}$ we have*

$$\left| \sum_{k=0}^{N} k^m z^k \right| \leqslant \frac{2}{|1-z|} N^m. \tag{D.1}$$

Proof. Let us notice that a "naive" upper bound

$$\left| \sum_{k=0}^{N} k^m z^k \right| \leqslant \sum_{k=0}^{N} k^m \sim N^{m+1}$$

is not sufficient for our purposes. To get the order of N^m we have to take into account cancellations in the sum (D.1) as follows:

$$(1-z)\sum_{k=0}^{N} k^m z^k = -N^m z^{N+1} + \sum_{k=0}^{N} k^m z^k - \sum_{k=0}^{N-1} k^m z^{k+1} = -N^m z^{N+1} + \sum_{k=1}^{N} (k^m - (k-1)^m) z^k.$$

Then by the triangle inequality

$$|1-z| \left| \sum_{k=0}^{N} k^m z^k \right| \leqslant N^m + \sum_{k=1}^{N} k^m - (k-1)^m = 2N^m.$$

This completes the proof of Lemma D.1. ☐

References

[1] 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.

[2] A. Akinshin, G. Goldman, and Y. Yomdin. Geometry of error amplification in solving Prony system with near-colliding nodes. *arXiv:1701.04058 [math]*, Jan. 2017.

[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. Stability and super-resolution of generalized spike recovery. *Applied and Computational Harmonic Analysis*, 45(2):299–323, Sept. 2018.

[5] D. Batenkov, L. Demanet, G. Goldman, and Y. Yomdin. Conditioning of partial nonuniform Fourier matrices with clustered nodes. *arXiv:1809.00658 [cs, math]*, Sept. 2018.

[6] D. Batenkov, G. Goldman, and Y. Yomdin. Super-resolution of near-colliding point sources. *arXiv:1904.09186 [math]*, Apr. 2019.

[7] D. Batenkov and Y. Yomdin. Geometry and Singularities of the Prony mapping. *Journal of Singularities*, 10:1–25, 2014.

[8] F. Bazán. Conditioning of rectangular Vandermonde matrices with nodes in the unit disk. *SIAM Journal on Matrix Analysis and Applications*, 21:679, 2000.

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

[10] B. Beckermann and E. B. Saff. The sensitivity of least squares polynomial approximation. In *Applications and Computation of Orthogonal Polynomials*, pages 1–19. Springer, 1999.

[11] B. Beckermann and A. Townsend. Bounds on the Singular Values of Matrices with Displacement Structure. *SIAM Review*, 61(2):319–344, Jan. 2019.

[12] A. Ben-Israel and T. N. E. Greville. *Generalized Inverses: Theory and Applications*. Number 15 in CMS Books in Mathematics. Springer, New York, 2nd ed edition, 2003.

[13] Å. Björck. Component-wise perturbation analysis and error bounds for linear least squares solutions. *BIT*, 31(2):237–244, 1991.

[14] Å. Björck and G. H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of computation*, 27(123):579–594, 1973.

[15] M.-D. Choi. Tricks or treats with the Hilbert matrix. *The American Mathematical Monthly*, 90(5):301–312, 1983.

[16] A. Cordova, W. Gautschi, and S. Ruscheweyh. Vandermonde matrices on the circle: Spectral properties and conditioning. *Numerische Mathematik*, 57(1):577–591, Dec. 1990.

[17] C. deBoor. Divided differences. *Surveys in Approximation Theory*, 1:46–69, 2005. [Online article at] http://www.math.technion.ac.il/sat.

[18] L. Demanet and N. Nguyen. The recoverability limit for superresolution via sparsity. 2014.

[19] B. Diederichs. Well-Posedness of Sparse Frequency Estimation. *arXiv:1905.08005 [math]*, May 2019.

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

[21] A. Eisinberg, P. Pugliese, and N. Salerno. Vandermonde matrices on integer nodes: The rectangular case. *Numerische Mathematik*, 87(4):663–674, Feb. 2001.

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

[23] W. Gautschi. On inverses of Vandermonde and confluent Vandermonde matrices. II. *Numerische Mathematik*, 5(1):425–430, 1963.

[24] W. Gautschi. Norm estimates for inverses of Vandermonde matrices. *Numerische Mathematik*, 23(4):337–347, 1974.

[25] W. Gautschi. On inverses of Vandermonde and confluent Vandermonde matrices III. *Numerische Mathematik*, 29(4):445–450, 1978.

[26] N. J. Higham. A survey of componentwise perturbation theory in numerical linear algebra. In *Proceedings of Symposia in Applied Mathematics*, volume 48, pages 49–77, 1994.

[27] J. A. Hogan and J. D. Lakey. *Duration and Bandwidth Limiting: Prolate Functions, Sampling, and Applications*. Springer Science & Business Media, Dec. 2011.

[28] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge ; New York, 2nd ed edition, 2012.

[29] A. V. Knyazev and M. E. Argentati. Principal angles between subspaces in an A-based scalar product: Algorithms and perturbation estimates. *SIAM Journal on Scientific Computing*, 23(6):2008–2040, 2002.

[30] S. Kunis and D. Nagel. On the condition number of Vandermonde matrices with pairs of nearly-colliding nodes. *arXiv:1812.08645 [math]*, Dec. 2018.

[31] S. Kunis and D. Nagel. On the smallest singular value of multivariate Vandermonde matrices with clustered nodes. *arXiv:1907.07119 [cs, math]*, July 2019.

[32] H. B. Lee. Eigenvalues and eigenvectors of covariance matrices for signals closely spaced in frequency. *IEEE Transactions on signal processing*, 40(10):2518–2535, 1992.

[33] W. Li and W. Liao. Stable super-resolution limit and smallest singular value of restricted Fourier matrices. *arXiv:1709.03146 [cs, math]*, Sept. 2017.

[34] W. Li, W. Liao, and A. Fannjiang. Super-resolution limit of the ESPRIT algorithm. *arXiv:1905.03782 [cs, math]*, May 2019.

[35] C. A. Micchelli. Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constructive Approximation*, 2(1):11–22, Dec. 1986.

[36] 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.

[37] V. Y. Pan. How Bad Are Vandermonde Matrices? *SIAM Journal on Matrix Analysis and Applications*, 37(2):676–694, Jan. 2016.
[38] D. Slepian. A numerical method for determining the eigenvalues and eigenfunctions of analytic kernels. *SIAM Journal on Numerical Analysis*, 5(3):586–600, 1968.
[39] 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.
[40] P. Stoica and R. Moses. *Spectral Analysis of Signals*. Pearson/Prentice Hall, 2005.
[41] J. Todd. The condition of the finite segments of the Hilbert matrix. *Contributions to the solution of systems of linear equations and the determination of eigenvalues*, 39:109–116, 1954.
[42] T. E. Tuncer and B. Friedlander. *Classical and Modern Direction-of-Arrival Estimation*. Academic, London, 2009. OCLC: 299711034.
[43] E. E. Tyrtyshnikov. How bad are Hankel matrices? *Numerische Mathematik*, 67(2):261–269, Mar. 1994.
[44] A. J. Wathen and S. Zhu. On spectral distribution of kernel matrices related to radial basis functions. *Numerical Algorithms*, 70(4):709–726, Dec. 2015.
[45] H. S. Wilf. *Finite sections of some classical inequalities*, volume 52. Springer Science & Business Media, 2012.
[46] Z. Yang, J. Li, P. Stoica, and L. Xie. Chapter 11 - Sparse methods for direction-of-arrival estimation. In R. Chellappa and S. Theodoridis, editors, *Academic Press Library in Signal Processing, Volume 7*, pages 509–581. Academic Press, Jan. 2018.

Department of Applied Mathematics, School of Mathematical Sciences, Tel Aviv University, P.O. Box 39040, Tel Aviv 6997801, Israel

E-mail address: dbatenkov@tauex.tau.ac.il

University of Passau and Fraunhofer IIS Research Group Knowledge Based Image Processing, Passau, Germany

E-mail address: benedikt.diederichs@uni-passau.de

Department of Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel

E-mail address: gil.goldman@weizmann.ac.il

Department of Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel

E-mail address: yosef.yomdin@weizmann.ac.il