New perspectives on Fourier matrices

Weilin Li
Department of Mathematics
City University of New York – City College
NY, USA
wli6@ccny.cuny.edu

***Abstract*—We outline new lower bounds for the smallest singular value of univariate and multivariate non-harmonic Fourier matrices. If the node set has sufficiently small local sparsity at an appropriate scale, then the smallest singular value is primarily determined by the local multiscale geometry of the node set. This illustrates an implicit localization phenomenon of the Fourier transform. We highlight some important conceptual and technical advancements that lead to these bounds, such as the use of refined polynomial interpolation methods and sparsity decomposition.**

I. Introduction

For any finite set $X = \{x_k\}_{k=1}^s \subseteq \mathbb{T} := \mathbb{R}/\mathbb{Z}$, a *non-harmonic Fourier matrix* of size $m \times s$ is defined as

$$\Phi := \Phi(m, X) := \left[e^{-2\pi i j x_k}\right]_{j=0,1,\ldots,m-1,\, k=1,2,\ldots,s}.$$

We call X the *node set*, which is completely arbitrary. This definition generalizes the discrete Fourier transform (DFT) matrix, which consists of orthogonal (harmonic) columns and corresponds to the case where $m = s$ and $X = \{k/m\}_{k=0}^{s-1}$.

Fourier matrices are classical objects that are connected to various topics, such as polynomial interpolation/regression, Fourier analysis, and exponential sums. Numerical applications require good understanding of their extreme singular values. Two motivational examples include super-resolution [10], [17] and nonuniform discrete Fourier transforms [11], [20]

Generally speaking, older papers concentrated on square matrices, $m = s$. Aside from special cases like the DFT matrix, there is a consensus that Fourier and Vandermonde matrices are oftentimes highly ill-conditioned (e.g., condition number that grows quickly in s). This is supported by both empirical and theoretical work, such as [12], [9], [6], [7].

While square Fourier and Vandermonde matrices can be brittle and numerically difficult to work with, tall Fourier matrices where $m \gtrsim s$ are more stable. To understand the main questions and direction of more recent research on this topic, it may be helpful to consider the following examples.

For $t \in \mathbb{T}$, let $|t| := \min_{n\in\mathbb{Z}} |t-n|$ and define the *minimum separation* of X as

$$\Delta(X) := \min_{j\neq k} |x_j - x_k|.$$

For a given number of rows m, we say a set X is *well-separated* if $\Delta(X) > 1/m$. It was shown in [1] (a slight improvement over [18]) that if X is well-separated, then

$$\sqrt{m - \Delta^{-1}} \leq \sigma_{\min}(\Phi) \leq \sigma_{\max}(\Phi) \leq \sqrt{m + \Delta^{-1}}. \quad \text{(I.1)}$$

Note inequality (I.1) implicitly assumes that $m > s$ through the requirement that $\Delta(X) > 1/m$. It tells us that a large class of tall Fourier matrices has uniformly bounded condition number, in stark contrast to the square setting.

To bypass the well-separated assumption, some geometric models were considered. We say $X \subseteq \mathbb{T}$ consists of *separated clumps* if there are disjoint sets $U_1, \ldots, U_r$ called *clumps*, such that each one has diameter on the order of $1/m$ and

$$X = U_1 \cup U_2 \cup \cdots \cup U_r.$$

It was shown in [17] (see the proceeding [16] for a summary) that if $m \geq s^2$, the largest clump has cardinality λ, and the clumps are at least $C\sqrt{s\lambda^5 s m^{-3}\Delta^{-1}}$ apart from each other, then

$$\sigma_{\min}(\Phi) \gtrsim \sqrt{\frac{m}{\lambda}}\,(cm\Delta)^{\lambda-1}. \quad \text{(I.2)}$$

Inequality (I.2) is sharp in m and Δ. It communicates a *localization* effect of the Fourier transform: even though Φ depends on all elements in X, the lower bound is only determined by on the local properties of X, as captured by λ. There has been significant research over the past several years on related problems and there are variations of this inequality, see [17], [4], [5], [13], [2].

While estimate (I.2) extends the class of X for which there is an accurate bound for $\sigma_{\min}(\Phi)$, it is not entirely satisfactory. Since $\sigma_{\min}(\Phi) > 0$ for any $X \subseteq \mathbb{T}$ and $m \geq |X|$, in principle, one should be able to quantify its extreme singular values without resorting to special cases. For instance, under what conditions on X, aside from clumps, do we expect an inequality like (I.2) to hold?

Having explained the requisite background, we state the main goals of this proceeding. In Section II, we explain the main results in [14]. This paper settles the problem of finding a lower bound for $\sigma_{\min}(\Phi)$ that not only holds for arbitrary X, but also reduces to (I.1) and (I.2) (modulo quantities that do not depend on m and Δ) as special cases. In Section III, we briefly describe the main ideas and new techniques developed in [14], which are based on polynomial methods and sparsity decompositions. Finally, in Section IV, we explain the results in [15], which extends these techniques to higher dimensions and provides universal lower bounds for the smallest singular value of multivariate Fourier matrices.

II. A UNIVERSAL ESTIMATE AND MULTISCALE BEHAVIOR

In order to provide a lower bound for $\sigma_{\min}(\Phi)$ that is applicable to all X, we need a general yet natural notion of a "complexity" of X. This is done through two definitions.

Let $\tau \in (0, \frac{1}{2}]$ and $I(x,\tau) \subseteq \mathbb{T}$ be the closed (periodic) interval of length 2τ centered at x. The τ *local sparsity* of X is defined as

$$\nu := \nu(\tau, X) := \max_{k=1,\dots,s} \big|X \cap I(x_k, \tau)\big|.$$

It is the maximum number of τ-neighbors any $x_k \in X$ has, including itself. We say a finite set $X \subseteq \mathbb{T}$ satisfies the (m,τ) *density criterion* if

$$\frac{3\nu(\tau, X)}{\tau} \leq m.$$

Notice that $\nu(\tau, X)/(2\tau)$ is precisely the maximum density of X at scale τ, so this requirement roughly states that X is not too dense at this scale. If $m \geq 6s$, then the density criterion is trivially satisfied for $\tau = 1/2$. For various interesting sets, we can set τ to be smaller. For example, if $\Delta(X) > 3/m$, then the density criterion is satisfied for $\tau = \Delta(X)$.

Let us first state a simplified version of [14, Theorem 1] where all explicit constants have been suppressed and further simplified.[1]

Theorem 1. *For any $m \geq 6s$ and $X = \{x_k\}_{k=1}^s \subseteq \mathbb{T}$. There are explicit $C > 0$ and $c \in (0,1)$ such that if X satisfies the (m,τ) density criterion for some τ, then*

$$\sigma_{\min}(\Phi) \geq C\, c^{\nu} \sqrt{\frac{m}{s\nu}} \min_{1\leq k\leq s} \left\{ \prod_{0<|x_j - x_k| \leq \frac{\tau}{2}} \frac{|x_j - x_k|}{\tau} \right\}.$$

Theorem 1 expresses a localization result whereby if X has sufficiently low density at scale τ, then $\sigma_{\min}(\Phi)$ is controlled by product term is taken only over $0 < |x_j - x_k| \leq \tau/2$. This is a localization phenomenon, and the only requirements are that the density criterion holds and that Φ has at least 6 times more rows than columns.

Theorem 1 highlights how $\sigma_{\min}(\Phi)$ depends on the multiscale structure of X. The coarsest scale is $\tau/2$ and interactions between x_j and x_k that are further apart than $\tau/2$ only influence $\sigma_{\min}(\Phi)$ through ν. Finer scales are defined as $2^{-\ell}\tau$ for $\ell \geq 2$. The theorem shows that interactions between elements in X at finer scales dominate the behavior of $\sigma_{\min}(\Phi)$.

Separated clumps is an example of a low density set. [14, Corollary 1] shows that if $m \geq 6s$ and if X consists of separated clumps that are at least $3\lambda/m$ apart, then

$$\sigma_{\min}(\Phi) \gtrsim \sqrt{\frac{m}{s}} \left(cm\Delta\right)^{\lambda - 1}. \tag{II.1}$$

Compared to the prior result (I.2) derived in [17], inequality (I.2) also achieves the optimal dependence on m and Δ, but with a drastically weaker condition on the clump separation. Note that the clumps must be separated by at least λ/m otherwise we are in a scenario where it is possible that $m < s$.

Let us look at one more example. For a given $\varepsilon \in (0,1)$, consider a *sparse spike train* $X = \{\varepsilon k/m\}_{k=0}^{s-1}$. It was shown in [18], and further improved in [2] with explicit constants, that are $c_1, c_2 > 0$ such that if $s \geq c_1 \log m$, then

$$\sigma_{\min}(\Phi) \lesssim e^{-c_2 \varepsilon s}. \tag{II.2}$$

This has been a troublesome example to deal with because the exponential smallness of inequality (II.2) is not due to any two elements in X being very close together, but rather, due to constructive interference in the columns of Φ.

The sparse spike train is a high density set whereby X does not satisfy the (m,τ) density criterion for any $\tau \leq \varepsilon(s-1)/m$. Invoking Theorem 1 for $\tau = 1/2$ and manipulating the resulting expression (see [14, Section 4.4] for details), whenever $m \geq 6s$, there is an explicit $c_3 > 0$ such that

$$\sigma_{\min}(\Phi) \gtrsim e^{-c_3 \varepsilon s}. \tag{II.3}$$

We discussed three important examples in the literature (well-separated, clumps, and sparse spike train), and Theorem 1 provides optimal scaling for these examples. This illustrates that the density criterion is a natural way of describing the "complexity" of X.

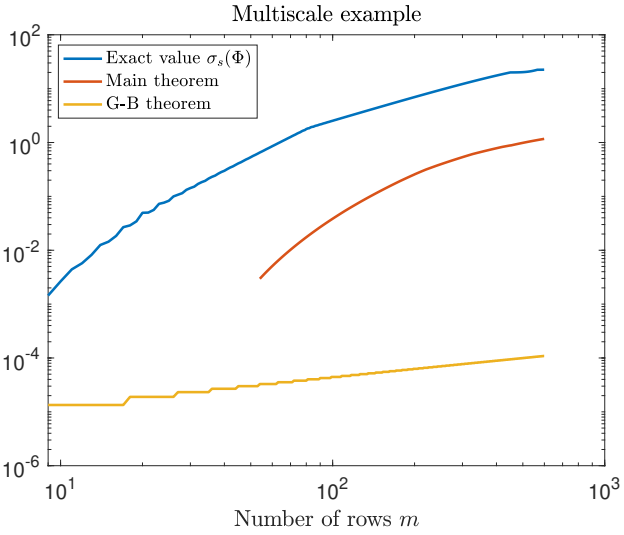


Fig. 1: Plot of $\sigma_{\min}(\Phi)$ and two different lower bounds.

The full version of Theorem 1, namely [14, Theorem 1], also contains accurate numerical constants. For example, let

$$X := X_1 \cup X_2 \cup X_3, \quad X_1 := \{0, \tfrac{1}{90}, \tfrac{2}{90}, \tfrac{3}{90}\},$$
$$X_2 := \tfrac{1}{3} + \{0, \tfrac{1}{200}, \tfrac{2}{200}\}, \quad X_3 := \tfrac{2}{3} + \{0, \tfrac{1}{500}\}.$$

Inequality (I.1) is only applicable when $m > 500$. For $m \leq 500$, the only other result that is applicable is a combination of Gautschi and Bazán [12], [6], which yields the inequality,

$$\sigma_s(\Phi) \geq \sqrt{\frac{1}{s}\left\lfloor \frac{m}{s} \right\rfloor} \min_{1\leq k\leq s} \left\{ \prod_{j\neq k} \frac{|e^{2\pi i x} - e^{2\pi i x_k}|}{2} \right\}. \tag{II.4}$$

[1] To get Theorem 1 from [14, Theorem 1], following the notation in the reference, we observe that $\nu(\tau, \mathcal{G}_k) \leq \nu(\tau, X)$ since $\mathcal{G}_k \subseteq X$, and we have the inequalities $|\mathcal{J}_k| \leq \nu$, $\alpha_k \leq 3\nu/(2m) \leq \tau/2$, and $|\mathcal{I}_k| \leq \nu$ which follow from their definitions and the density criterion.

The true value of $\sigma_{\min}(\Phi)$, lower bound given by [14, Theorem 1], and right side of (II.4) are displayed in Fig. 1. When $m = 400$, [14, Theorem 1] underestimates the true value by a multiplicative factor of ≈ 21 whereas the Gautschi-Bazán bound is off by a factor of $\approx 1.9 \cdot 10^5$.

III. Polynomials, density, and sparsity

In this section, we explain the main ingredients for the proof of Theorem 1, and provide some intuition for why the density criterion is natural.

It was already established in [17] that analyzing $\sigma_{\min}(\Phi)$ can be recast as a minimum norm interpolation problem. Indeed, let P_m denote the space of complex-valued trigonometric polynomials whose Fourier coefficients are supported in $\{0, \dots, m-1\}$, and v is any unit norm right singular vector of Φ corresponding to $\sigma_{\min}(\Phi)$. Then we have the relationship,

$$\sigma_{\min}(\Phi) = \max\left\{\|f\|_{L^2}^{-1} \colon f \in P_m \text{ and } f|_X = v\right\}. \quad \text{(III.1)}$$

Since the right side of (III.1) is a max, to lower bound $\sigma_{\min}(\Phi)$, it suffices to carefully construct an interpolating polynomial.

Through some basic considerations, we can reduce this problem down to finding a Lagrange basis for X whose norms are not too big. Let us pick a reference point $x_k \in X$ and focus on finding a $f_k \in P_m$ such that $f_k(x_j) = \delta_{j,k}$ for all $x_j \in X$. We can use (III.1) for the polynomial $f = \sum_{k=1}^s v_k f_k$, which interpolates the singular vector v on X. We must be careful with the construction of f_k, so that $\|f_k\|_{L^2}$ is not too large, otherwise use of (III.1) would provide a loose estimate.

Throughout this section, fix an arbitrary $\tau \in (0, \frac{1}{2}]$. We decompose X into the bad set $B_k := X \cap I(x_k, \tau)$ and good set $G_k := X \cap I(x_k, \tau)^c$.

To handle the bad set, we pick an appropriate $q \in \mathbb{N}$ to be determined later and use a modified Lagrange interpolant,

$$b_k(x) = \prod_{0<|x_j - x_k| \le \tau} \frac{e^{2\pi i q x} - e^{2\pi i q x_j}}{e^{2\pi i q x_k} - e^{2\pi i q x_j}}. \quad \text{(III.2)}$$

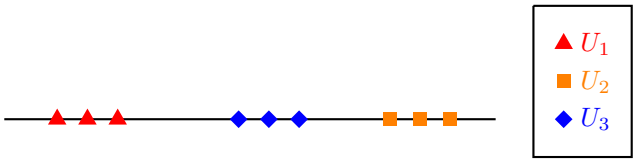
For appropriate q, we have the interpolation property that $b_k(x_j) = \delta_{j,k}$ for all $x_j \in X \cap I(x_k, \tau)$.

The main innovation behind the proof of Theorem 1 is a method for dealing with the good set. A simply greedy argument shows that there exist $r \le \nu(\tau, X)$ many disjoint and non-empty sets $W_{k,1}, \dots, W_{k,r}$ such that $\Delta(W_{k,\ell}) > \tau$ for each ℓ and

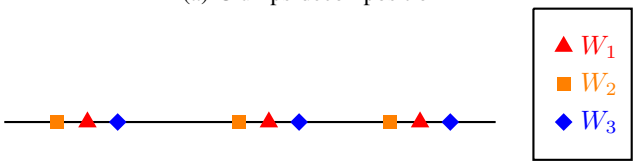
$$G_k = W_{k,1} \cup \cdots \cup W_{k,r}. \quad \text{(III.3)}$$

We call (III.3) the *sparsity decomposition* of G_k. It is perhaps counter-intuitive that this decomposition is helpful. For example, if X consists of clumps separated by at least τ, then G_k consists of all clumps that do not contain x_k, yet (III.3) completely ignores the clump structure. See Figure Fig. 2 for an illustration.

The advantage of decomposition (III.3) is apparent once we translate inequality (I.1) into a statement about interpolating polynomials. The proof of [14, Proposition 7.2] showed that



(a) Clumps decomposition



(b) Sparsity decomposition

Fig. 2: Clumps versus sparsity decomposition of the same set.

for any integer $n > 1/\tau$, there is a $g_{k,\ell} \in P_n$ such that $g_{k,\ell}(x_k) = 1$, $g_{k,\ell}|_{W_{k,\ell}} = 0$, and

$$\|g_{k,\ell}\|_{L^\infty} \le \sqrt{1 + \frac{1}{n\tau - 1}}. \quad \text{(III.4)}$$

This polynomial has remarkably small norm, e.g., uniformly bounded by $\sqrt{2}$ whenever $n \ge 2/\tau$. In contrast, a naive approach of writing down the standard Lagrange interpolant of these $|W_{k,\ell}| + 1$ data points and estimating its sup-norm would yield a loose bound that grows like $(C\tau^{-1})^{|W_{k,\ell}|}$ where $C \ge 1$. Essentially (III.3) decomposes the good set G_k into pieces such that each piece $W_{k,\ell}$ contains significant self-cancellations, as captured by inequality (III.4). This is the key insight of the sparsity decomposition.

With these polynomials at hand, a desired Lagrange basis $\{f_k\}_{k=1}^s$ for X can be defined as

$$f_k = b_k g_{k,1} \cdots g_{k,r}.$$

It satisfies the requisite property that $f_k(x_j) = \delta_{j,k}$ for all $x_j \in X$. We need to ensure that $f_k \in P_m$, which is how the density criterion is used. The degree of f_k is at most $(\nu-1)q + r(n-1)$. If the density criterion holds, then setting $q = \lfloor 1/\tau \rfloor$ and $n = \lfloor 2/\tau \rfloor$ shows that $f \in P_m$. Finally, using Hölder's and (III.4), we have

$$\|f_k\|_{L^2} \le \|b_k\|_{L^2} \prod_{\ell=1}^r \|g_{k,\ell}\|_{L^\infty} \le c^\nu \|b_k\|_{L^2}.$$

Details for upper bounding $\|b_k\|_{L^2}$ can be found in [14]. This is roughly how the localization phenomenon in Theorem 1 appears.

Another way of viewing Theorem 1 is that a low density set can be broken into not too many well-behaved pieces. This argument via polynomials also establishes a more general principle. If X satisfies the (m, τ) density criterion, then

$$\sigma_{\min}(\Phi) \gtrsim c^\nu \min_{1 \le k \le s} \sigma_{\min}\Big(\Phi\big(\lfloor m/\nu \rfloor, B_k\big)\Big).$$

Essentially, the global problem of analyzing a Fourier matrix with node set X can be reduced to the local problem of dealing with just node set B_k, but at the price of introducing a c^ν term

and cutting the number of rows by a factor of ν.

IV. Extension to multivariate matrices

We let $d \geq 2$ be the dimension and $X = \{x_k\}_{k=1}^s \subseteq \mathbb{T} := (\mathbb{R}/\mathbb{Z})^d$ be arbitrary. The sampling set $\{0, \dots, m-1\}$ in one dimension needs to be replaced by a suitable subset of $\mathbb{Z}^d$. There is no canonical choice. For the purposes of this proceeding, suppose the samples are collected inside Q_m, a closed cube in $\mathbb{R}^d$ of side length $2m$ centered at zero.

Then a *non-harmonic multivariate Fourier matrix* is

$$\Phi := \Phi(Q_m, X) = \left[e^{-2\pi i j x_k}\right]_{j \in Q_m \cap \mathbb{Z}^d,\, k=1,2,\dots,s}.$$

Let $|Q_m|_*$ be the number of multi-integers contained in Q_m. When $m \in \mathbb{N}$, this matrix has $|Q_m|_* = (2m+1)^d$ rows. Since our samples are collected in a cube, we let $|\cdot|$ denote the ℓ^∞ norm on $\mathbb{T}^d$. The minimum separation of X is defined as

$$\Delta(X) := \min_{j \neq k} |x_j - x_k|.$$

Much of the theory in [15] extends to more general sampling sets such as multi-integers contained in some ℓ^p ball of radius m centered at zero.

Our first question is how large Δ must be for the condition number of Φ to be uniformly bounded. Inequality (I.1) was proved in [1], [18] by using arguments similar to [21], [19], which all exploited properties of special functions known as the Beurling-Selberg majorant and minorant. There has been significant progress on multivariate analogues, but optimal constructions are not available depending on what function one wants to majorize and minorize. This includes the indicator function of Q_m, the case that is relevant to our problem, see [8] and references therein. Nonetheless, using the construction in [3, Section 2.5], we proved the following theorem, copied from [15, Theorem 2.3]. For convenience, we let $\beta_0 := 1/(2\ln 2) \approx 0.7213$ and $c(\beta) := e^{1/(2\beta)} - 1$.

Theorem 2. *Let $\beta \geq \beta_0$, $X \subseteq \mathbb{T}^d$, and $m \geq 1$. If $\Delta(X) \geq \beta d/m$, then we have*

$$\sigma_{\min}(\Phi) \geq \sqrt{\big(1 - c(\beta)\big)\,|Q_m|_*},$$
$$\sigma_{\max}(\Phi) \leq \sqrt{\big(1 + c(\beta)\big)\,|Q_m|_*}.$$

The $\sqrt{|Q_m|_*}$ term that appears in both bounds is a natural scaling factor since this is precisely the ℓ^2 norm of any column in Φ. Of course, $c(\beta) \in [0,1]$ whenever $\beta \geq \beta_0$ and decreases to zero as β increases. The theorem shows that the condition number of Φ is uniformly bounded by a constant that only depends on β and not on the dimension d. While this conclusion does not suffer from the effects of dimensionality, the separation condition $\Delta(X) \geq \beta d/m$ does, but it is unclear whether linear growth in d is necessary or can be relaxed (or removed altogether).

It is not difficult to show that the matrix $\Phi(Q_m, X)$ is injective whenever $|Q_m|_* \geq |X|^d$. Provided this condition holds, it should be possible to lower bound $\sigma_{\min}(\Phi)$ without any restrictions on X.

To this end, we aim to generalize the results of [14] to higher dimensions. Let $Q(x,\tau) \subseteq \mathbb{T}^d$ be the closed cube of side length 2τ centered at x. The τ *local sparsity* of X is defined as

$$\nu := \nu(\tau, X) := \max_{k=1,\dots,s} \big|X \cap Q(x_k, \tau)\big|.$$

The following theorem is [15, Theorem 3.10].

Theorem 3. *Let $\beta \geq \beta_0$, $X := \{x_k\}_{k=1}^s \subseteq \mathbb{T}^d$, and $m \geq 4s$. If*

$$\tau \leq \frac{1}{4d} \quad \text{and} \quad \frac{2\beta d\,\nu(\tau, X)}{\tau} \leq m, \tag{IV.1}$$

then we have the inequality,

$$\sigma_{\min}(\Phi) \geq C_{m,s,\nu,\beta} \min_{1 \leq k \leq s} \left\{ \prod_{0 < |x_j - x_k|_1 \leq \frac{\nu}{m}} \frac{m}{\nu} |x_j - x_k|_1 \right\},$$

where

$$C_{m,s,\nu,\beta} := \sqrt{\frac{(c(\beta))^\nu}{2^{\nu+1} s} |Q_{m/(2\nu)}|_*}$$

This theorem shows that when (IV.1) holds, the behavior of $\sigma_{\min}(\Phi)$ is predominately determined by the interactions between pairs in X closer than ν/m. This proves a localization effect of the Fourier transform. It may be interesting to note that for the multivariate case, the ℓ^1 distance appears even though the measurements are taken from a cube; this is an effect of duality since Q_m lies in frequency domain, while X is in space, or vice versa. Notice that there is a d factor in condition (IV.1) which appears from the assumptions in Theorem 2.

The primary difference between Theorems 1 and 3 is that $\nu(\tau, X)/(2\tau)$ is legitimately a density in one dimension, but it does not have the same interpretation for $d \geq 2$. Instead, it can be viewed as some projected density. This is of course good, since replacing $\nu(\tau, X)/\tau$ in (IV.1) with $\nu(\tau, X)/(2\tau)^d$ (which is actually a density term) would result in a much stronger assumption than (IV.1).

The polynomial approach is helpful yet again, as it allows us to prove Theorem 3 in an analogous way to Theorem 1. Indeed, the polynomial characterization (III.1) still holds, except P_m is now the space of multivariate trigonometric polynomials whose Fourier coefficients are supported in $Q_m \cap \mathbb{Z}^d$.

To use (III.1), we can construct a Lagrange basis for X. Fix a reference point $x_k \in X$ and we would like to construct $f_k \in P_m$ such that $f_k(x_j) = \delta_{j,k}$ for all $x_j \in X$. We decompose X into a good and bad set again. The bad set is $B_k := X \cap Q(x_k, \tau)$, while the good set is $G_k := X \cap Q(x_k, \tau)^c$. The good set is handled using the sparsity decomposition and a polynomial version of Theorem 2. There are several technical differences for dealing with bad set in the multivariate case, but many general ideas carry over.

Acknowledgments

WL is supported by NSF-DMS Award #2309602 and a Cycle 55 PSC-CUNY grant.

REFERENCES

[1] Céline Aubel and Helmut Bölcskei. Vandermonde matrices with nodes in the unit disk and the large sieve. *Applied and Computational Harmonic Analysis*, 47(1):53–86, 2019.

[2] Alex H Barnett. How exponentially ill-conditioned are contiguous submatrices of the Fourier matrix? *SIAM Review*, 64(1):105–131, 2022.

[3] Jeffrey T. Barton. *Analogs of the Beurling-Selberg functions in N dimensions and their applications*. PhD thesis, The University of Texas at Austin, 1999.

[4] Dmitry Batenkov, Laurent Demanet, Gil Goldman, and Yosef Yomdin. Conditioning of partial nonuniform Fourier matrices with clustered nodes. *SIAM Journal on Matrix Analysis and Applications*, 41(1):199–220, 2020.

[5] Dmitry Batenkov and Gil Goldman. Single-exponential bounds for the smallest singular value of Vandermonde matrices in the sub-Rayleigh regime. *Applied and Computational Harmonic Analysis*, 55:426–439, 2021.

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

[7] Lihu Berman and Arie Feuer. On perfect conditioning of Vandermonde matrices on the unit circle. *The Electronic Journal of Linear Algebra*, 16:157–161, 2007.

[8] Jacob Carruth, Noam Elkies, Felipe Gonçalves, and Michael Kelly. The Beurling-Selberg box minorant problem via linear programming bounds. *arXiv preprint arXiv:1702.04579 (2nd version)*, 2022.

[9] Antonio Córdova, Walter Gautschi, and Stephan Ruscheweyh. Vandermonde matrices on the circle: spectral properties and conditioning. *Numerische Mathematik*, 57(1):577–591, 1990.

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

[11] Alok Dutt and Vladimir Rokhlin. Fast Fourier transforms for nonequispaced data. *SIAM Journal on Scientific computing*, 14(6):1368–1393, 1993.

[12] Walter Gautschi. On inverses of vandermonde and confluent vandermonde matrices. *Numerische Mathematik*, 5:425–430, 1963.

[13] Stefan Kunis and Dominik Nagel. On the smallest singular value of multivariate vandermonde matrices with clustered nodes. *Linear Algebra and its Applications*, 604:1–20, 2020.

[14] Weilin Li. Multiscale estimates for the condition number of nonharmonic Fourier matrices. *Mathematics of Computation*, 2024.

[15] Weilin Li. Nonharmonic multivariate fourier transforms and matrices: condition numbers and hyperplane geometry. *arXiv preprint arXiv:2407.10313*, 2024.

[16] Weilin Li and Wenjing Liao. Conditioning of restricted Fourier matrices and super-resolution of MUSIC. In *2019 13th international conference on sampling theory and applications (sampta)*, pages 1–4. IEEE, 2019.

[17] Weilin Li and Wenjing Liao. Stable super-resolution limit and smallest singular value of restricted fourier matrices. *Applied and Computational Harmonic Analysis*, 51:118–156, 2021.

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

[19] Hugh L Montgomery. The analytic principle of the large sieve. *Bulletin of the American Mathematical Society*, 84(4):547–567, 1978.

[20] Gerlind Plonka, Daniel Potts, Gabriele Steidl, and Manfred Tasche. *Numerical Fourier Analysis*. Springer, 2018.

[21] Atle Selberg. *Collected papers*, volume 1. Springer, 1989.