THE FOURIER RATIO AND COMPLEXITY OF SIGNALS

K. ALDALEH, W. BURSTEIN, G. GARZA, G. HART, A. IOSEVICH, J. IOSEVICH, A. KHALIL, J. KING, N. KULKARNI, T. LE, I. LI, A. MAYELI, B. MCDONALD, K. NGUYEN, AND N. SHAFFER

Abstract. The purpose of this paper is to investigate the degree to which the ratio

$$\mathrm{FR}(f) \equiv \sqrt{N} \cdot \frac{\|\widehat{f}\|_{L^1(\mu)}}{\|\widehat{f}\|_{L^2(\mu)}} = \frac{\sum_{m=0}^{N-1} |\widehat{f}(m)|}{\left(\sum_{m=0}^{N-1} |f(x)|^2\right)^{\frac{1}{2}}}$$

tells us whether the signal $f : \mathbb{Z}_N \to \mathbb{C}$ is learnable, where

$$\widehat{f}(m) = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} e^{-\frac{2\pi i x m}{N}} f(x),$$

and

$$||f||_{L^p(\mu)} = \left(\frac{1}{N} \sum_{x=0}^{N-1} |f(x)|^p \right)^{\frac{1}{p}}.$$

We use the celebrated Bourgain-Talagrand theorem to see that if f is concentrated in a generic set of size $\leqslant c\frac{N}{\log(N)}$, the ratio above is suitably large. We also show that if the ratio is very small, then f can be well-approximated by a low-degree polynomial. One implication of this approximation is that the algorithmic rate distortion (a stable version of Kolmogorov complexity) of signals with a small Fourier ratio is small. We then use these ideas to study the Vapnik-Chervonenkis and statistical query dimensions of the class of functions with a small Fourier ratio. We explain connections between these ideas and the theory of exact signal recovery (e.g., [12], [20], [19]). Finally, we prove a general version of Chang's lemma, which shows that in the presence of the small Fourier Ratio ($\mathrm{FR}(f)$), the set of large Fourier coefficients possesses non-trivial additive structure. In the final section of the paper, we compute this ratio for a variety of real-life data sets, showing that the ratio is a useful guide for the forecastability of time series. In the same section, we run experiments to understand the magnitude of constants that arise in the theoretical results established in this paper.

Contents

Date: November 7, 2025.

A. I. was supported in part by the National Science Foundation under NSF DMS - 2154232. A. M. was supported in part by the AMS-Simons Research Enhancement Grant and by the National Science Foundation under NSF DMS-2453769.

1. Introduction

Let $f : \mathbb{Z}_N \to \mathbb{C}$ be a signal. The purpose of this paper is to explore a simple mechanism to determine whether f contains sufficient structure to learn this signal using a relatively small number of samples. A simple, common-sense idea is that it is challenging to learn a "random" signal, but the task becomes considerably easier in the presence of structure. The question that has been addressed by many authors in a variety of contexts is how to determine the degree to which a given signal behaves like a random one. The central theme of this paper is that the quantity

$$\mathrm{FR}(f) \equiv \sqrt{N} \cdot \frac{\|\widehat{f}\|_{L^1(\mu)}}{\|\widehat{f}\|_{L^2(\mu)}} = \frac{\|\widehat{f}\|_{L^1(\mathbb{Z}_N)}}{\|\widehat{f}\|_{L^2(\mathbb{Z}_N)}}. \tag{1.1}$$

is an effective detector of randomness and structure, where here and throughout, if $g : \mathbb{Z}_N \to \mathbb{C}$,

$$\|g\|_{L^p(\mu)} = \left(\frac{1}{N} \sum_{x \in \mathbb{Z}_N} |g(x)|^p \right)^{\frac{1}{p}}, \tag{1.2}$$

$$\|g\|_p = \left(\sum_{x \in \mathbb{Z}_N} |g(x)|^p \right)^{\frac{1}{p}}, \tag{1.3}$$

and

$$\widehat{f}(m) = \frac{1}{\sqrt{N}} \sum_{x \in \mathbb{Z}_N} \chi(-xm) f(x),$$

where $\chi(t) = e^{\frac{2\pi i t}{N}}$.

The initial draft of this paper was written during the StemForAll2025 ([30]) research program organized by the 5th and 12th listed authors in July/August 2025. The paper was finished under the auspices of the continuous Vertical Integration of Research program, also supervised by the 5th and the 12th listed authors of the paper, designed to bring undergraduates, graduate students, postdocs, and faculty together to work on research.

1.1. Structure of this paper.

We are now going to describe the structure of the paper.

- In Subsection 1.2, we lay down the notation frequently used in this paper.
- In Subsection 1.3, we state the celebrated result due to Talagrand ([31]) on random subsets of sets of orthogonal functions and use it to illustrate the Fourier Ratio $\mathrm{FR}(f)$ (1.1) in a natural context in Subsection 1.4.
- In Subsection 1.5, we study the range of the Fourier Ratio, and show that $1 \leqslant \mathrm{FR}(f) \leqslant \sqrt{N}$.

- in Subsection 1.6, we prove refined lower and upper bounds for the Fourier Ratio and show that it is a natural controlling parameter in the classical Fourier Uncertainty Principle.
- In Subsection 1.7 we study perhaps the key illustrative example, the indicator function of $\mathbb{Z}_p$ embedded in $\mathbb{Z}_{pq}$, where p and q are distinct primes. This example shows that $\mathrm{FR}(f)$ and $FR(\widehat{f})$ should be considered in analyzing the complexity of a signal.
- In Subsection 1.8, we prove that if the Fourier Ratio is small, then the signal can be well-approximated by a low degree trigonometric polynomial, either in L^∞ or L^2 norms using probabilistic tools.
- In Subsection 1.9, we leverage the results from Subsection 1.8 to prove that signals with small Fourier Ratio have small Kolmogorov complexity, or, more precisely, small algorithmic rate distortion (1.17).
- In Subsection 1.10, we use the theory of the Fourier Ratio built up to that point, along with some ideas from compressed sensing to show that a small Fourier Ratio allows one to impute the missing values in time series with high accuracy under some realistic assumptions on the size of the missing set. In the process, we show that the Fourier Ratio of a signal does not change much if we restrict it to a sufficiently large random subset.
- In Subsection 1.11, we show that both the Vapnik-Chervonenkis and statistical query dimension of the class of signals with a small Fourier Ratio are small. Quantitative bounds are provided.
- In Subsection 1.12, we prove a series of results showing that the Fourier Ratio is stable under small perturbations.
- In Subsection 1.13, we study the Fourier Ratio in the context of Chang's lemma from additive number theory. We show that a small Fourier Ratio implies a considerable amount of additive structure.
- Finally, in Section 2, we perform a series of numerical experiments to illustrate some of the results in this paper. First, we compute the Fourier Ratio for several very different real-life data sets and show that the Fourier Ratio is quite small. We compute the Fourier Ratio of a random data set and show that the Fourier Ratio is very large, as the theory suggests. Our final set of numerical experiments explores the value of the constant in Talagrand's inequality, as it has a practical bearing on the applications established in this paper.

1.2. **Normalization and notation.** Throughout this paper, we work on $\mathbb{Z}_N$ with the unitary Fourier transform

$$\widehat{f}(m) = \frac{1}{\sqrt{N}} \sum_{x \in \mathbb{Z}_N} e^{-\frac{2\pi i x m}{N}} f(x),$$

and

$$f(x) = \frac{1}{\sqrt{N}} \sum_{m \in \mathbb{Z}_N} e^{\frac{2\pi i x m}{N}} \widehat{f}(m).$$

The Plancherel identity holds in both ℓ^2 and probability-normalized $L^2(\mu)$:

$$\|\widehat{f}\|_2 = \|f\|_2, \qquad \|\widehat{f}\|_{L^2(\mu)} = \|f\|_{L^2(\mu)}.$$

We frequently convert between ℓ^p and $L^p(\mu)$ using

$$\|g\|_{L^p(\mu)} = N^{-\frac{1}{p}}\|g\|_p \quad (1 \leqslant p < \infty), \qquad \|g\|_{L^\infty(\mu)} = \|g\|_\infty.$$

1.2.1. *Fourier Ratio.* With this normalization,

$$\mathrm{FR}(f) \ := \ \sqrt{N}\,\frac{\|\widehat{f}\|_{L^1(\mu)}}{\|\widehat{f}\|_{L^2(\mu)}} \ = \ \frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_2} \ \in [1, \sqrt{N}],$$

is scale-invariant. More precisely, $FR(\alpha f) = \mathrm{FR}(f)$ for all $\alpha \neq 0$. The extreme cases are $\mathrm{FR}(f) = 1$ iff $\widehat{f}$ is 1-sparse, and $\mathrm{FR}(f) = \sqrt{N}$ when $|\widehat{f}(m)|$ is constant.

1.2.2. *Numerical sparsity and coherence.* The numerical sparsity of a vector x is

$$\mathrm{ns}(x) := \big(\|x\|_1/\|x\|_2\big)^2,$$

hence $\mathrm{FR}(f)^2 = \mathrm{ns}(\widehat{f})$. See, for example, [1]. We also use the standard notion of coherence,

$$\mu(f) \ := \ \frac{N\|f\|_\infty^2}{\|f\|_2^2} \in [1, N],$$

which controls random-restriction sampling rates in our results.

1.2.3. *Bi-Fourier ratio.* Motivated by the example in Subsection 1.7 below, we summarize Fourier complexity using the quantity

$$\mathrm{FR}_{\mathrm{bi}}(f) := \min\{\mathrm{FR}(f),\, FR(\widehat{f})\}.$$

This is invariant (up to constants) under applying the Fourier transform or its inverse.

1.3. **Some key background results.** Our starting point is the following result due to Talagrand ([31])

Theorem 1.1. *Let $(\varphi_j)_{j=1}^n$ be an orthonormal system in $L^2(\mathbb{Z}_N)$ with $\|\varphi_j\|_{L^\infty} \leqslant K$ for $1 \leqslant j \leqslant n$. There exists a constant $\gamma_0 \in (0,1)$ and a subset $I \subset \{1, \ldots, n\}$ with $|I| \geqslant \gamma_0 n$ such that for every $a = (a_i) \in \mathbb{C}^n$,*

$$\left(\sum_{i \in I} |a_i|^2\right)^{1/2} \leqslant C_T\, K\, \big(\log(n)\, \log\log(n)\big)^{1/2} \left\|\sum_{i \in I} a_i \varphi_i\right\|_{L^1},$$

where $C_T > 0$ is a universal constant.

We shall need the following version, stated for signals on $\mathbb{Z}_N$, with the roles of h and $\widehat{h}$ reversed. See [3].

Definition 1.2. Let $0 < p < 1$. We say a random set $S \subset [n] = \{0, 1, \cdots, n-1\}$ is generic if each element of $[n]$ is selected independently with probability p.

The following result can be deduced from Theorem 1.1 (see [3]).

Theorem 1.3. *There exists $\gamma_0 \in (0,1)$ such that if $h : \mathbb{Z}_N \to \mathbb{C}$ supported in a generic set M (in the sense of Definition 1.2) of size $\gamma_0 \frac{N}{\log(N)}$, then with probability $1 - o_N(1)$,*

$$\left(\frac{1}{N} \sum_{m \in \mathbb{Z}_N} |\widehat{h}(m)|^2 \right)^{\frac{1}{2}} \leqslant C_T (\log(N) \log\log(N))^{\frac{1}{2}} \cdot \frac{1}{N} \sum_{m \in \mathbb{Z}_N} |\widehat{h}(m)|, \tag{1.4}$$

where $C_T > 0$ is a constant that depends only on γ_0.

Remark 1.4. It is known (see [31]) that $\sqrt{\log(N)}$ in (1.4) cannot, in general, be removed, and it is not known whether the removing the remaining $\sqrt{\log\log(N)}$ is possible.

In some cases, the term $(\log(N) \log\log(N))^{\frac{1}{2}}$ can be removed. The following result, stated in the setting of this paper, is due to Bourgain ([2]).

Theorem 1.5. *Suppose that M is generic, as above, $|M| = \lceil N^{\frac{2}{q}} \rceil$, $q > 2$. Then for all $h : \mathbb{Z}_N \to \mathbb{C}$, supported in M,*

$$\left(\frac{1}{N} \sum_{m \in \mathbb{Z}_N} |\widehat{h}(m)|^q \right)^{\frac{1}{q}} \leqslant C(q) \cdot \left(\frac{1}{N} \sum_{m \in \mathbb{Z}_N} |\widehat{h}(m)|^2 \right)^{\frac{1}{2}} \tag{1.5}$$

It is not difficult to deduce from the proof that if M is chosen randomly, then the bound holds with probability $1 - \varepsilon$ if $C(q)$ is replaced by $\frac{C(q)}{\varepsilon}$.

We learned the following observation from William Hagerstrom ([15]), which can be proven using Hölder's inequality.

Lemma 1.6. *Suppose that for $h : \mathbb{Z}_N^d \to \mathbb{C}$,*

$$\left(\frac{1}{N} \sum_{x \in \mathbb{Z}_N^d} |\widehat{h}(x)|^q \right)^{\frac{1}{q}} \leqslant C(q) \left(\frac{1}{N} \sum_{x \in \mathbb{Z}_N} |\widehat{h}(x)|^2 \right)^{\frac{1}{2}} \tag{1.6}$$

for some $q > 2$.

Then

$$\left(\frac{1}{N} \sum_{x \in \mathbb{Z}_N} |\widehat{h}(x)|^2 \right)^{\frac{1}{2}} \leqslant (C(q))^{\frac{q}{q-2}} \cdot \frac{1}{N} \sum_{x \in \mathbb{Z}_N} |\widehat{h}(x)|.$$

Remark 1.7. It follows that if the expected size of M in Theorem 1.3 is $O(N^{1-\varepsilon})$ for some $\varepsilon > 0$, then $C_T \sqrt{\log(N) \log\log(N)}$ in (1.4) can be replaced by C'_T, independent of N.

1.4. **The Fourier ratio makes its first appearance.** We are now ready to exploit the setup above. Let $S \subset \mathbb{Z}_N$, chosen randomly. Then in view of Theorem 1.3 and Remark 1.7, with very high probability, we have

$$\|\check{1}_S\|_{L^2(\mu)} \leqslant C_T \|\check{1}_S\|_{L^1(\mu)}. \tag{1.7}$$

Using Plancherel, we see that

$$\|\check{1}_S\|_{L^1(\mu)} \geqslant \frac{1}{C_T}\sqrt{\frac{|S|}{N}},$$

or, in other words,

$$\frac{\|\check{1}_S\|_{L^1(\mu)}}{\|\check{1}_S\|_{L^2(\mu)}} \geqslant \frac{1}{C_T},$$

the largest possible size up to a constant. It is not difficult to see that the same conclusion holds if $\check{1}_S$ is replaced by an arbitrary function with the Fourier transform supported in S.

On the other hand, let's consider the case where S is far from random. Suppose that $S = \mathbb{Z}_N$. Then

$$\check{1}_S(x) = \sqrt{N} \cdot \delta_0(x),$$

where $\delta_0(x) = 1$ if $x = 0$ and 0 otherwise. it is not difficult to check that in this case,

$$\frac{\|\check{1}_S\|_{L^1(\mu)}}{\|\check{1}_S\|_{L^2(\mu)}} = \frac{1}{\sqrt{N}}, \tag{1.8}$$

a much smaller quantity than what we get in (1.7) in the generic case. This suggests that the quantity (1.1) in the case $f = \check{1}_S$ may contain significant information indicating the degree to which a set S is random.

In the remainder of this paper, we explore to what extent the Fourier Ratio captures the degree to which the signal f is learnable, in a variety of different senses. We also apply the resulting ideas to the imputation of time series with missing values.

1.5. **Bounding the Fourier ratio from above and below.** Note that by Fourier inversion and the triangle inequality,

$$|f(x)| \leqslant N^{-\frac{1}{2}} \sum_{m \in \mathbb{Z}_N} |\widehat{f}(m)| = N^{-\frac{1}{2}} \|\widehat{f}\|_1.$$

Squaring both sides, summing over $x \in \mathbb{Z}_N$, and taking square roots yields

$$\|f\|_2 \leqslant \|\widehat{f}\|_1,$$

so by Plancherel, we see that

$$\mathrm{FR}(f) = \frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_2} = \frac{\|\widehat{f}\|_1}{\|f\|_2} \geqslant 1,$$

and we saw in the example described above (1.8) that this lower bound is realized by the constant function 1. Note also that by Cauchy-Schwarz, we have that

$$\mathrm{FR}(f) = \frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_2} \leqslant \frac{\sqrt{N}\|\widehat{f}\|_2}{\|\widehat{f}\|_2} = \sqrt{N}.$$

Thus we have just seen that

$$1 \leqslant \mathrm{FR}(f) \leqslant \sqrt{N}.$$

The same argument used to bound $\mathrm{FR}(f)$ below shows that if f is supported in $E \subset \mathbb{Z}_N$, then

$$\mathrm{FR}(f) \geqslant \frac{\sqrt{N}}{\sqrt{|E|}}. \tag{1.9}$$

To see that (1.9) can be realized, let $E = q\mathbb{Z}_p$, p prime, be a copy of $\mathbb{Z}_p$ sitting inside $\mathbb{Z}_{pq}$ in the obvious way. Then, for the setting $N = pq$, we have

$$\widehat{1}_E(m) = \frac{1}{\sqrt{N}} \sum_{k=0}^{p-1} e^{-\frac{2\pi iqkm}{N}} = \begin{cases} \frac{p}{\sqrt{N}} 1_S(m), & \text{if } p \mid m, \\ 0, & \text{if } p \nmid m. \end{cases}$$

where $S = p\mathbb{Z}_q$, sitting inside $\mathbb{Z}_{pq}$ in the natural way. Setting $f = 1_E$, it follows that

$$\mathrm{FR}(f) = \sqrt{N} \cdot \frac{\|\widehat{f}\|_{L^1(\mu)}}{\|\widehat{f}\|_{L^2(\mu)}} = \frac{\sqrt{N}}{\sqrt{p}} = \frac{\sqrt{N}}{\sqrt{|E|}}. \tag{1.10}$$

1.6. **Refined lower and upper bounds on the Fourier Ratio and the classical Fourier Uncertainty Principle.** We now refine both the lower and the upper bounds for the Fourier Ratio and demonstrate that the Fourier Ratio is a natural controlling parameter in the classical Fourier Uncertainty Principle.

Theorem 1.8. *Let $f : \mathbb{Z}_N \to \mathbb{C}$, L^2-concentrated in $E \subset \mathbb{Z}_N$ at level $a \in (0,1)$, in the sense that*

$$||f||_{L^2(E^c)} \leqslant a||f||_{L^2(\mathbb{Z}_N)},$$

with $\widehat{f}$ L^1- concentrated on $S \subset \mathbb{Z}_N$ at level $b \in (0,1)$, in the sense that

$$||\widehat{f}||_{L^1(S^c)} \leqslant b||\widehat{f}||_{L^1(\mathbb{Z}_N)}.$$

Then

$$\frac{(1-a)^2}{4} \cdot \frac{N}{|E|} \leqslant FR(f)^2 \leqslant \frac{|S|}{(1-b)^2}. \tag{1.11}$$

In particular,

$$|E| \cdot |S| \geqslant (1-a)^2 \cdot (1-b)^2 \cdot \frac{N}{4}, \tag{1.12}$$

a version of the classical uncertainty principle (see e.g. [12]). While (1.12) is well-known, (1.11) shows that the Fourier Ratio is a natural controlling parameter in the Fourier Uncertainty Principle.

1.7. **An important example and its consequences.** The example given above where $N = pq$, p, q distinct primes, and $f = 1_E$, where E denotes $\mathbb{Z}_p$ embedded in $\mathbb{Z}_{pq}$ in a natural way, points to a very interesting phenomenon that we need to address.

As we noted above, $\mathrm{FR}(f)$, in this case, is equal to $\sqrt{\frac{N}{p}}$. If p is much larger than q, then $\mathrm{FR}(f)$ is small, which is consistent with our philosophy that the size of $\mathrm{FR}(f)$ indicates the degree of randomness of the signal, since the indicator of $\mathbb{Z}_p$ inside $\mathbb{Z}_{pq}$ has low complexity in any reasonable sense. However, if p is much smaller than q, then $\mathrm{FR}(f)$ is rather large, and the paradigm appears to break down. The key observation here is that while $\mathrm{FR}(f) = \sqrt{\frac{N}{p}}$, we have $\mathrm{FR}(\widehat{f}) = \sqrt{\frac{N}{q}}$. Why is $\mathrm{FR}(\widehat{f})$ relevant? Simply because we can go from f to $\widehat{f}$ by applying the Fourier matrix, and go back by applying the inverse Fourier matrix. Since the $\mathbb{Z}_N$ Fourier transform (or inverse Fourier transform) runtime is $\approx N\log(N)$ (see e.g., [4, 10, 13, 27]), low complexity of f implies the low complexity for $\widehat{f}$ and vice versa.

In summary, the quantity

$$\mathrm{FR}_{\mathrm{bi}}(f) = \min\left\{\mathrm{FR}(f), \mathrm{FR}(\widehat{f})\right\} \tag{1.13}$$

is better at capturing the complexity of a given signal in view of the discussion above. We now turn our attention towards justifying this paradigm.

1.8. **Approximation by low degree polynomials.** Our first main result amplifies the idea described in Subsection 1.4, which indicates that in the presence of randomness, the Fourier Ratio is large. Indeed, we will see that if the signal is even concentrated in a random set, then the ratio $\mathrm{FR}(f)$ is suitably large.

Theorem 1.9. *Let $f : \mathbb{Z}_N \to \mathbb{C}$. Suppose that there exists a generic set M such that*

$$\|f\|_{L^2(M^c)} \leqslant r\|f\|_2$$

for some $r \in (0,1)$, with $|M| \leqslant \gamma_0 \frac{N}{\log(N)}$, where γ_0 is as in Theorem 1.3. Suppose that

$$r < \frac{1-r}{C_T\sqrt{\log(N)\log\log(N)}}. \tag{1.14}$$

Then

$$FR(f) \geqslant \sqrt{N} \cdot \frac{1 - r\frac{C_T\sqrt{\log(N)\log\log(N)}}{1-r}}{\frac{C_T\sqrt{\log(N)\log\log(N)}}{1-r}} \tag{1.15}$$

with probability $1 - o_N(1)$.

Remark 1.10. Using the observation in Remark 1.7, if $|M| \leqslant N^{1-\varepsilon}$ for some $\varepsilon > 0$, we can replace (1.14) with

$$r < \frac{1-r}{C_T},$$

and we can replace (1.15) with

$$\mathrm{FR}(f) \geqslant \sqrt{N} \cdot \frac{\left(1 - r\frac{C_T}{1-r}\right)}{\frac{C_T}{1-r}}.$$

Our next result concerns L^∞ approximations and it shows that if the corresponding ratio $\frac{\|\widehat{f}\|_{L^1(\mu)}}{\|f\|_\infty}$ is small, then we can approximate f in L^∞ norm by a trigonometric polynomial with degree logarithmic in N.

Theorem 1.11. *Let $f : \mathbb{Z}_N \to \mathbb{C}$ and let $\eta > 0$. Then for any integer k such that*

$$k > 8 \left(\frac{\|\widehat{f}\|_{L^1(\mu)}}{\|f\|_\infty} \right)^2 \frac{N \log(4N)}{\eta^2},$$

there exists a trigonometric polynomial

$$P(x) = \sum_{i=1}^{k} c_i \chi(m_i x)$$

such that

$$\|f - P\|_\infty < \eta \|f\|_\infty.$$

Remark 1.12. Note that the triangle inequality shows $\frac{\|\widehat{f}\|_{L^1(\mu)}}{\|f\|_\infty} \geqslant N^{-\frac{1}{2}}$ (with the equality for a dirac Delta $f = 1_{\{x_0\}}$), and so in the best case, Theorem 1.11 indeed gives a polynomial of degree $O(\log(N))$.

Remark 1.13. It is interesting to note that a slight variant of the quantity arising above, namely $\mu(f) = \frac{N\|f\|_\infty^2}{\|f\|_2^2} \in [1, N]$, is known as coherence, and it was introduced in a related context by Candes and Recht ([8]). This quantity is going to come up several times in this paper.

The proof of Theorem 1.11 is probabilistic, and similar techniques allow for good L^2 polynomial approximations in terms of the ratio $\mathrm{FR}(f)$, as in the following theorem.

Theorem 1.14. *Let $f : \mathbb{Z}_N \to \mathbb{C}$, and let $\eta > 0$. Then for any k such that*

$$k > \frac{FR(f)^2 - 1}{\eta^2},$$

there exists a trigonometric polynomial

$$P(x) = \sum_{i=1}^{k} c_i \chi(m_i x)$$

such that

$$\|f - P\|_2 < \eta \|f\|_2.$$

Remark 1.15. Theorem 1.14 can be interpreted as follows. Define

$$\mathcal{C}(r) := \{f : \mathrm{FR}(f) \leqslant r\}\,. \tag{1.16}$$

Fix $\varepsilon > 0$ and suppose $f \in \mathcal{C}(r)$. Then there exists a degree $\frac{r^2}{\varepsilon^2}$ trigonometric polynomial P such that

$$\|f - P\|_2 < \varepsilon \cdot \|f\|_2.$$

In other words, the class of signals with a small Fourier ratio is a subset of a set of signals of low complexity, as measured by the degree of the approximating polynomial. We shall build on this theme in the next section, where we use the notion of Kolmogorov complexity.

We can continue to use probabilistic techniques to get good L^1 approximations, now in terms of the corresponding ratio $\frac{\|\widehat{f}\|_1}{\|f\|_1}$.

Theorem 1.16. *Let $f : \mathbb{Z}_N \to \mathbb{C}$ and let $\eta > 0$. If*

$$k \geqslant 32\pi \left(\frac{\|\widehat{f}\|_1}{\|f\|_1}\right)^2 \cdot \frac{N}{\eta^2},$$

then there exists a trigonometric polynomial P of degree k such that

$$\|f - P\|_1 \leqslant \eta \|f\|_1.$$

Remark 1.17. We can again use the triangle inequality to see that $\frac{\|\widehat{f}\|_1}{\|f\|_1} \geqslant \frac{1}{\sqrt{N}}$, and consequently, in the best case, Theorem 1.16 gives a polynomial of constant degree. Notice, in the probability-normalized notation, the same argument yields, i.e., $\|\widehat{f}\|_{L^1(\mu)} \geqslant \frac{1}{\sqrt{N}}\|f\|_{L^1(\mu)}$.

1.9. **Connections with the theory of Kolmogorov complexity.** We shall need the following definition. See e.g. [25], Chapter 2 and Chapter 7.

Definition 1.18. Let U be a fixed universal Turing machine. For a function $f : \mathbb{Z}_N \to \mathbb{C}$, its Kolmogorov complexity $K(f)$ is defined as

$$K(f) \;=\; \min\{\, |p| \;:\; U(p,x) = f(x) \text{ for all } x \in \mathbb{Z}_N \,\},$$

where $|p|$ denotes the length of the binary program p.

Since we aim to understand to what extent the Fourier ratio allows us to determine how well we can expect to learn a given time series, the Kolmogorov complexity is too restrictive because it is easily led astray by small perturbations. To be precise, let p, q denote two odd primes, and let $E = q\mathbb{Z}_p$, embedded in $\mathbb{Z}_{pq}$ in the natural way. As we saw in (1.10), $FR(1_E) = \frac{\sqrt{N}}{\sqrt{|E|}} = \frac{\sqrt{N}}{\sqrt{p}} = \sqrt{q}$. In the case when p is much larger than q, this ratio is rather small, and correctly suggests that 1_E is easy to learn. However, suppose that we modify 1_E by changing one of its values from 1 to $1 + \delta$, where δ is a very small positive number with k digits in its base 2 expansion. It follows that

$$K(f) = K(1_E) + O(k)$$

since it is not difficult to see that it takes $O(k)$ lines of a fixed universal Turing code to describe δ. It is not difficult to see that up to a small error, $K(1_E) = \log(pq)$, so if k is much larger than $\log(pq)$, the Kolmogorov complexity becomes arbitrarily large. And yet, f is very well-approximated by a function with low Kolmogorov complexity. To deal with this issue, we introduce a refined version of Kolmogorov complexity (see e.g [33]).

Given $f : \mathbb{Z}_N \to \mathbb{C}$, the algorithmic rate–distortion function of f at distortion level $\varepsilon > 0$ is defined by

$$r_f(\varepsilon) \;=\; \min\{\, \mathrm{K}(g) : g : \mathbb{Z}_N \to \mathbb{C},\ \|f - g\|_2 \leqslant \varepsilon \|f\|_2 \,\}, \tag{1.17}$$

where $\mathrm{K}(g)$ is the Kolmogorov complexity of g with respect to a fixed universal Turing machine. Thus $r_f(\varepsilon)$ quantifies the minimum description length of an approximation to f within distortion ε.

We can use Theorem 1.14 to show that functions with small $\mathrm{FR}(f)$ must have small algorithmic rate-distortion, by showing that low-degree trigonometric polynomials have small Kolmogorov complexity. Our next result is the following.

Theorem 1.19. *Suppose $f : \mathbb{Z}_N \to \mathbb{C}$, and let $\varepsilon > 0$. If $k = \frac{FR(f)^2}{\varepsilon^2}$, then we have that*

$$r_f(2\varepsilon) \leqslant C_U k \log\left(\frac{(1+\varepsilon)N\sqrt{k}}{\varepsilon}\right) + C_U',$$

where C_U and C_U' are constants depending only on the fixed universal Turing machine U.

We note that for $f : \mathbb{Z}_N \to \mathbb{C}$, $\mathrm{K}(f) \geqslant C \log N$ for some constant C depending on the Turing machine U. Thus, we see that if $\mathrm{FR}(f)$ is close to 1, the algorithmic rate-distortion of f is close to $\log N$, which is best possible.

1.10. **Applications to the recovery of missing values in time series.** We are now going to further exploit Theorem 1.14. Suppose that $f : \mathbb{Z}_N \to \mathbb{C}$ is a signal, and that only the values $\{f(x) : x \in X\}$ are known for some subset $X \subset \mathbb{Z}_N$. The following result shows that if $f \in \mathcal{C}(r)$, the missing values can be recovered efficiently.

Theorem 1.20. *Sample q indices of $[N]$ uniformly and independently where*

$$q = Cr^2/\varepsilon^2 \log(r/\varepsilon)^2 \log(N).$$

We name the sample set X. If C is a sufficiently large universal constant, then with high probability, for all $f \in \mathcal{C}(r)$, x^*, the solution to the linear program

$$\min_x \|\widehat{x}\|_1 \text{ subject to } \|f - x\|_{L^2(X)} \leqslant \|f\|_2 \cdot \varepsilon$$

satisfies

$$\|x^* - f\|_2 \leqslant 11.47\|f\|_2 \cdot \varepsilon.$$

Here, the empirical $L^2(X)$-norm is defined by $\|g\|_{L^2(X)} := \left(\frac{1}{|X|}\sum_{x\in X} |g(x)|^2\right)^{1/2}$.

Remark 1.21. It should be noted that the optimization problem of Theorem 1.20 is a convex program, so it is algorithmically feasible.

A slight deficiency of Theorem 1.20 is the a priori assumption that $f \in \mathcal{C}(r)$. Since we are only given the values of f restricted to X, it would be useful for us to have a result that says that the Fourier Ratio is not likely to be substantially different if it is computed over X rather than over the whole $\mathbb{Z}_N$. This is what our next result is about.

Theorem 1.22 (Random restriction preserves $\mathrm{FR}(f)$)**.** *Let $X \subset \mathbb{Z}_N$ be created by keeping each x independently with probability $p \in (0,1]$, and define the restriction*

$$f_X(x) = \begin{cases} f(x), & x \in X, \\ 0, & x \notin X. \end{cases}$$

Consider

$$\mu(f) = \frac{N\|f\|_\infty^2}{\|f\|_2^2} \in [1, N],$$

and fix parameters $\varepsilon \in (0, 1/2)$ and $u \geqslant 1$. There is a universal constant $C > 0$ such that if

$$p \geqslant C\,\frac{\mu(f)}{\varepsilon^2}\,\frac{\log N + u}{N},$$

then with probability at least $1 - 2e^{-u}$ the following estimates hold:

$$\big|\|\widehat{f_X}\|_2 - \sqrt{p}\,\|\widehat{f}\|_2\big| \leqslant \varepsilon\sqrt{p}\,\|\widehat{f}\|_2, \tag{1.18}$$

$$\big|\|\widehat{f_X}\|_1 - p\,\|\widehat{f}\|_1\big| \leqslant \varepsilon p\,\|\widehat{f}\|_1. \tag{1.19}$$

It follows that

$$\frac{FR(f_X)}{FR(f)} \in \left[\frac{1-\varepsilon}{\sqrt{1+\varepsilon}},\ \frac{1+\varepsilon}{\sqrt{1-\varepsilon}}\right] \subseteq [1 - 3\varepsilon,\, 1 + 3\varepsilon].$$

In particular,

$$|FR(f_X) - FR(f)| \leqslant 3\varepsilon\, FR(f).$$

Remark 1.23. Observe that $\mu(f) \leqslant \mathrm{FR}(f)^2$, so Theorem 1.22 allows for restriction to random sets of size $\gtrsim \frac{r^2 \log(N)}{\varepsilon^2}$.

We now address the fact that we only know $||f_X||_2$, not $||f||_2$.

Theorem 1.24. *Let $N \geqslant 2$ and let $f : \mathbb{Z}_N \to \mathbb{C}$ be arbitrary. For $p \in (0,1]$, let $X \subset \mathbb{Z}_N$ be formed by keeping each $x \in \mathbb{Z}_N$ independently with probability p, and define $f_X(x) = f(x)$ if $x \in X$ and $f_X(x) = 0$ otherwise. Let $\mu(f)$ be as above.*

Fix $\varepsilon \in (0,1)$ and $u \geqslant 1$. There is a universal constant $C > 0$ such that if

$$p \geqslant C\frac{\mu(f)}{\varepsilon^2}\frac{u}{N},$$

then with probability at least $1 - 2e^{-u}$ we have

$$|\|f_X\|_2 - \sqrt{p}\,\|f\|_2| \leqslant \varepsilon\sqrt{p}\,\|f\|_2.$$

Equivalently,

$$(1-\varepsilon)\sqrt{p}\,\|f\|_2 \leqslant \|f_X\|_2 \leqslant (1+\varepsilon)\sqrt{p}\,\|f\|_2.$$

Remark 1.25. In view of Theorem 1.24 and Theorem 1.22, we can replaced $\mathrm{FR}(f)$ and $||f||_2$ in Theorem 1.20 with suitably scaled $FR(f_X)$ and $||f_X||_2$, which resolves the data leakage problem in the sense that we do not know f away from X.

1.11. **VC Dimension and Statistical Query Dimension.** We are now going to conceptualize these results further by bringing in the concepts of the Vapnik-Chervonenkis and statistical query dimension.

Definition 1.26. For a concept class, $\mathcal{C} \subset \{-1,1\}^X$, and probability distribution, $\mathcal{D}$, on X, the statistical query dimension of $\mathcal{C}$ with respect to $\mathcal{D}$ is the largest number d such that $\mathcal{C}$ contains d functions $f_1, f_2, ..., f_d$ such that for all $i \neq j$,

$$|\mathbb{E}_{x\sim\mathcal{D}}\left[f_i(x)\cdot f_j(x)\right]| \leqslant \frac{1}{d}.$$

The Statistical Query Dimension of $\mathcal{C}$ is the maximum of the statistical query dimension of $\mathcal{C}$ with respect to $\mathcal{D}$ over all $\mathcal{D}$.

Definition 1.27. Given a concept class $\mathcal{C} \subset \{-1,1\}^X$, we say that $\mathcal{C}$ shatters the points $c_1, c_2, \ldots, c_n$ from X if the restriction of $\mathcal{C}$ to $\{c_1, c_2, \ldots, c_n\}$ is the set of all functions from $\{c_1, c_2, \ldots, c_n\}$ to $\{-1,1\}$. We say that the VC dimension of $\mathcal{C}$ is equal to d if $\mathcal{C}$ shatters some set of d points, but it does not shatter any set with $d+1$ points.

The following result from [28] says that the statistical query dimension bounds the VC dimension.

Theorem 1.28. *Let $\mathcal{C} \subset \{-1,1\}^X$. If the VC-dimension of $\mathcal{C}$ is d, then the Statistical Query dimension of $\mathcal{C}$ is $\Omega(d)$, where here, and throughout, $A = \Omega(B)$ if there exists a finite $c > 0$ such that $A \geqslant cB$.*

Next, we consider the statistical query dimension and VC-dimension of the class of functions with bounded Fourier Ratio. In order to match the above setting, define

$$\mathcal{B}(r) = \{f : \mathbb{Z}_N \to \{-1,1\} \; : \; \mathrm{FR}(f) \leqslant r\} \subset \{-1,1\}^{\mathbb{Z}_N}.$$

The next result bounds the statistical query dimension and VC-dimension of $\mathcal{B}(r)$.

Theorem 1.29. *The statistical query dimension of $\mathcal{B}(r)$ is*

$$\leqslant \frac{1}{\varepsilon^{\frac{r^2}{\varepsilon^2}}} \cdot \left(\frac{e}{\frac{r^2}{\varepsilon^2}}\right)^{\frac{r^2}{\varepsilon^2}} N^{\frac{r^2}{\varepsilon^2}}, \tag{1.20}$$

where $r = FR(f)$ and $\varepsilon = \frac{1}{4(1+2r)}$. In particular, the dimension is

$$\leqslant \left(\frac{eN}{4r^2(1+2r)}\right)^{\frac{r^2}{\varepsilon^2}} = \left(\frac{eN}{4r^2(1+2r)}\right)^{16r^2(1+2r)^2}, \qquad \textit{since } \frac{r^2}{\varepsilon^2} = r^2 \cdot 16(1+2r)^2.$$

Remark 1.30. The VC-dimension of $\mathcal{B}(r)$ can be estimated from (1.20) using Theorem 1.28.

1.12. **Stability of the Fourier ratio.** We are now going to investigate how the Fourier Ratio FR(f) behaves under small perturbations.

Theorem 1.31. *Let $f : \mathbb{Z}_N \to \mathbb{C}$. For any perturbation $n : \mathbb{Z}_N \to \mathbb{C}$, set $A = \|\widehat{f}\|_1$, $B = \|\widehat{n}\|_1$, $s = \|\widehat{f}\|_2$, $t = \|\widehat{n}\|_2$. If $t < s$, then*

$$\left|FR(f+n) - FR(f)\right| \leqslant \frac{\|\widehat{n}\|_1 + FR(f)\|\widehat{n}\|_2}{\|\widehat{f}\|_2 - \|\widehat{n}\|_2}. \tag{1.21}$$

Moreover, using $\|\widehat{n}\|_1 \leqslant \sqrt{N}\|\widehat{n}\|_2$ and $FR(f) \leqslant \sqrt{N}$, we have

$$\left|FR(f+n) - FR(f)\right| \leqslant \frac{\left(\sqrt{N} + FR(f)\right)\|\widehat{n}\|_2}{\|\widehat{f}\|_2 - \|\widehat{n}\|_2} \leqslant \frac{2\sqrt{N}\|\widehat{n}\|_2}{\|\widehat{f}\|_2 - \|\widehat{n}\|_2}.$$

In particular, if $\|\widehat{n}\|_2 \leqslant \frac{1}{2}\|\widehat{f}\|_2$, then

$$\left|FR(f+n) - FR(f)\right| \leqslant 2\left(\frac{\|\widehat{n}\|_2}{\|\widehat{f}\|_2}\right)\left(\sqrt{N} + FR(f)\right).$$

One way of interpreting Theorem 1.31 is the following. Let $\delta = \frac{||\widehat{n}||_2}{||\widehat{f}||_2}$. Then (1.21) tells us that

$$|\mathrm{FR}(f+n) - \mathrm{FR}(n)| \leqslant \frac{\delta}{1-\delta} \cdot (\mathrm{FR}(f) + \mathrm{FR}(n)).$$

If $\frac{\delta}{1-\delta} \leqslant \frac{1}{\sqrt{N}}$, we obtain

$$|\mathrm{FR}(f+n) - \mathrm{FR}(n)| \leqslant \frac{1}{\sqrt{N}}\mathrm{FR}(f) + 1.$$

Our next result tells us what happens if n is the Gaussian noise with a given variance.

Theorem 1.32. *Let $f, n : \mathbb{Z}_N \to \mathbb{C}$. Assume $n(x)$ are independent circular complex Gaussian variables with variance σ^2, and fix $\gamma \in (0,1)$. With probability at least $1-\gamma$,*

$$|FR(f+n) - FR(f)| \leqslant \frac{(FR(n) + FR(f))\ t_\gamma}{\|\widehat{f}\|_2 - t_\gamma},$$

where

$$t_\gamma = \sigma\left(\sqrt{N} + \sqrt{\log(1/\gamma)}\right).$$

We will now examine the situation where $f_i : \mathbb{Z}_N \to \mathbb{C}$, $1 \leqslant i \leqslant n$ is a collection of time series such that $f_i(x) = s(x) + n_i(x)$, where $s(x)$ is the "true" (unknown) signal, and n_i is Gaussian noise. The following well-known result shows that the average of f_i is, with high probability, much closer to the true signal $s(x)$ than any of the individual f_is.

Theorem 1.33 (Smoothing by averaging independent Gaussian noise)**.** *Let $f_1, \dots, f_n$: $\mathbb{Z}_N \to \mathbb{C}$ be given by $f_i(x) = s(x) + n_i(x)$, where $s : \mathbb{Z}_N \to \mathbb{C}$ is an unknown signal and,* *for each i and x, the noise $n_i(x)$ is circular complex Gaussian with mean 0 and variance σ^2, independent across i and x. Define the average*

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x) \qquad (x \in \mathbb{Z}_N).$$

Then:

(a) Unbiasedness and pointwise variance reduction:

$$\mathbb{E}\, f(x) = s(x), \qquad f(x) - s(x) \sim \mathcal{CN}\Big(0, \frac{\sigma^2}{n}\Big) \text{ for each } x.$$

(b) Mean squared error in ℓ^2:

$$\mathbb{E}\, \|f - s\|_2^2 = \frac{N\sigma^2}{n}, \quad \text{where } \|h\|_2^2 = \sum_{x\in\mathbb{Z}_N} |h(x)|^2.$$

(c) High-probability bound (norm concentration): for any $\gamma \in (0,1)$, with probability at least $1-\gamma$,

$$\|f - s\|_2 \leqslant \frac{\sigma}{\sqrt{n}}\Big(\sqrt{N} + \sqrt{\log(1/\gamma)}\Big).$$

A slightly sharper but longer bound is

$$\|f - s\|_2 \leqslant \frac{\sigma}{\sqrt{n}}\Big(\sqrt{N} + \sqrt{\log(1/\gamma)} + \frac{\log(1/\gamma)}{2\sqrt{N}}\Big) \quad \text{with probability at least } 1-\gamma.$$

Remark 1.34. Parts (a)-(b) of Theorem 1.33 are proved in [21, 24, 36] (and discussed in [26, 16]); the concentration ingredient for part (c) is proved in [34, 5] (Lipschitz Gaussian concentration) or [22, 35] (chi–square tails).

Our next result analyzes the relationship between the Fourier Ratio of the average f and the true signal s. We are going to see that averaging brings the Fourier ratio toward the signal at an $\frac{1}{\sqrt{n}}$ rate.

Theorem 1.35. *Let $f_i : \mathbb{Z}_N \to \mathbb{C}$ be given by $f_i = s + n_i$ for $i = 1, \dots, n$, where $s : \mathbb{Z}_N \to \mathbb{C}$ is fixed and the noises $n_i(x)$ are independent across i and x, each circular complex Gaussian with mean 0 and variance σ^2. Define the average*

$$f = \frac{1}{n}\sum_{i=1}^{n} f_i = s + \overline{n}, \qquad \overline{n} := \frac{1}{n}\sum_{i=1}^{n} n_i.$$

Fix $\gamma \in (0,1)$ and set $r_\gamma = \sqrt{N} + \sqrt{\log(1/\gamma)}$. Then the following hold.

(i) With probability at least $1-\gamma$,

$$\|\overline{n}\|_2 \leqslant \frac{\sigma}{\sqrt{n}} r_\gamma, \qquad \|n_i\|_2 \leqslant \sigma r_\gamma \quad \text{for each fixed } i.$$

Consequently, if $\|s\|_2 \geqslant 2\sigma r_\gamma$*, then with probability at least* $1-\gamma$*,*

$$\big|FR(f) - FR(s)\big| \leqslant \frac{2\sigma r_\gamma}{\sqrt{n}\,\|s\|_2}\big(FR(\overline{n}) + FR(s)\big),$$

and for each fixed i*,*

$$\big|FR(f_i) - FR(s)\big| \leqslant \frac{2\sigma r_\gamma}{\|s\|_2}\big(FR(n_i) + FR(s)\big).$$

(ii) In particular, on the event in (i), the averaged estimate satisfies

$$\big|FR(f) - FR(s)\big| \leqslant \frac{1}{\sqrt{n}} C_\gamma(s, n_i, \overline{n}),$$

where

$$C_\gamma(s, n_i, \overline{n}) = \frac{2\sigma r_\gamma}{\|s\|_2}\big(FR(\overline{n}) + FR(s)\big),$$

so the deviation of $FR(f)$ *from* $FR(s)$ *is smaller by a factor* $1/\sqrt{n}$ *relative to the corresponding single-shot bound (up to the benign replacement of* $FR(n_i)$ *by* $FR(\overline{n})$*, which has the same distribution because* $FR(\alpha e) = FR(e)$ *for all* $\alpha > 0$*).*

1.13. **Chang's Lemma and additive structure.** Next, we consider the extent to which the ratio $\mathrm{FR}(f)$ indicates an additive structure of the function f. For indicator functions, we have the following result, which shows that large spectrum sets have some additive structure.

Theorem 1.36 (Sparse-spectrum approximation from small $\mathrm{FR}(f)$)**.** *Let* $f : \mathbb{Z}_N \to \mathbb{C}$*, and fix* $\eta > 0$*. Define the large spectrum*

$$\Gamma := \Big\{ m \in \mathbb{Z}_N \ : \ |\widehat{f}(m)| \geqslant \eta \|f\|_{L^2(\mu)} \Big\}.$$

Then

$$|\Gamma| \leqslant \frac{FR(f)}{\eta}\sqrt{N},$$

and moreover if

$$P(x) := \frac{1}{N^{\frac{1}{2}}} \sum_{m \in \Gamma} \widehat{f}(m)\chi(xm),$$

then

$$\|f - P\|_2 \leqslant \eta \|f\|_2.$$

In other words, f *can be approximated by a polynomial of degree*

$$\frac{FR(f)}{\eta} \cdot \sqrt{N}$$

up to an error $\leqslant \eta$*.*

Remark 1.37. Note that our bounds on $\mathrm{FR}(f)$ show that in the best case, this approximating polynomial could be of degree $\frac{1}{\eta}\sqrt{N}$. While this is of worse degree than in Theorem 1.14, in this case the approximating polynomial is deterministic. Moreover, we will see that we can ensure the Fourier support of the polynomial $P(x)$ has an additive structure by using a suitable generalization of a result due to Chang.

A result due to Chang ([11]) shows that large spectrum sets, as in Theorem 1.36, have additive structure in the case that f is the indicator function of a set.

Lemma 1.38 (Chang's lemma). *Let $A \subset \mathbb{Z}_N$ have density $\alpha = \frac{|A|}{N}$, and for $\eta > 0$ define the large spectrum set*

$$\Gamma = \left\{ m \in \mathbb{Z}_N \,:\, |\widehat{1_A}(m)| \geqslant \eta \alpha N^{\frac{1}{2}} \right\}.$$

Then there exists $\Lambda \subset \Gamma$ with

$$|\Lambda| \leqslant C\eta^{-2} \log\left(\frac{1}{\alpha}\right)$$

such that every $m \in \Gamma$ is a $\{-1, 0, 1\}$-linear combination of elements of Λ.

Lemma 1.38 was used by Chang originally to improve on bounds in Freiman's theorem ([11]), as well as by Green in [14] to improve on a result due to Bourgain on arithmetic progressions in sumsets.

Chang's lemma suggests that large spectrum sets have some arithmetic structure, and by generalizing Chang's lemma from indicator functions to general functions $f : \mathbb{Z}_N \to \mathbb{C}$, we can show that a small $\mathrm{FR}(f)$ implies additive structure in the large valued sets of a function. To this end, we have the following result.

Theorem 1.39 (Generalized Chang's Lemma). *Let $f : \mathbb{Z}_N \to \mathbb{C}$, and for $\eta > 0$ define the large spectrum set*

$$\Gamma = \left\{ m \in \mathbb{Z}_N \,:\, |\widehat{f}(m)| \geqslant \eta \|f\|_{L^2(\mu)} \right\}.$$

Then there exists a constant C and some $\Lambda \subset \Gamma$ with

$$|\Lambda| \leqslant C\eta^{-2} \left(\frac{\|f\|_{\frac{\log N}{\log N - 1}}}{\|f\|_2} \right)^2 \log N \tag{1.22}$$

and

$$|\Lambda| \leqslant C\eta^{-2} \frac{\|f\|_1}{\|f\|_2} \log\left(\left(\frac{\|f\|_2}{\|f\|_1} \right)^2 N \right) \tag{1.23}$$

such that every $m \in \Gamma$ is a $\{-1, 0, 1\}$-linear combination of elements of Λ.

Remark 1.40. Applying Theorem 1.39 to $\widehat{f}$, we obtain

$$|\Lambda| \leqslant C\eta^{-2} \mathrm{FR}(f) \log\left(\mathrm{FR}(f)^{-2} N \right),$$

from (1.23), where Λ is a set such that every

$$x \in \Gamma := \left\{ x \in \mathbb{Z}_N : |f(x)| \geqslant \eta \|f\|_{L^2(\mu)} \right\}$$

is a $\{-1, 0, 1\}$-linear combination of elements of Λ. Since the $\mathrm{FR}(f)$ term outside the log dominates, this implies that a small $\mathrm{FR}(f)$ means a small Λ, pointing towards additive structure in the large valued set of f. Alternatively, for large N, (1.22) approaches

$$|\Lambda| \leqslant C\eta^{-2} \mathrm{FR}(f) \log N,$$

giving the same result.

2. Numerical experiments

In this section, we are going to describe some numerical experiments to estimate the Fourier ratio for reasonable data sets and the Talagrand constant.

2.1. **Fourier ratio for real-life data sets.** We will illustrate the utility of the ratio $\mathrm{FR}(f) \equiv \sqrt{N} \cdot \frac{||\widehat{f}||_{L^1(\mu)}}{||\widehat{f}||_{L^2(\mu)}}$ on real-life data sets. The plots of the data sets are in Figure 1 below. The results are as follows:

- Peyton Manning Wikipedia Visits: $\mathrm{FR}(f) = 1.917$
- Electric Production: $\mathrm{FR}(f) = 2.133$
- Delhi Daily Climate: $\mathrm{FR}(f) = 2.715$
- Australia Monthly Beer Production: $\mathrm{FR}(f) = 2.884$

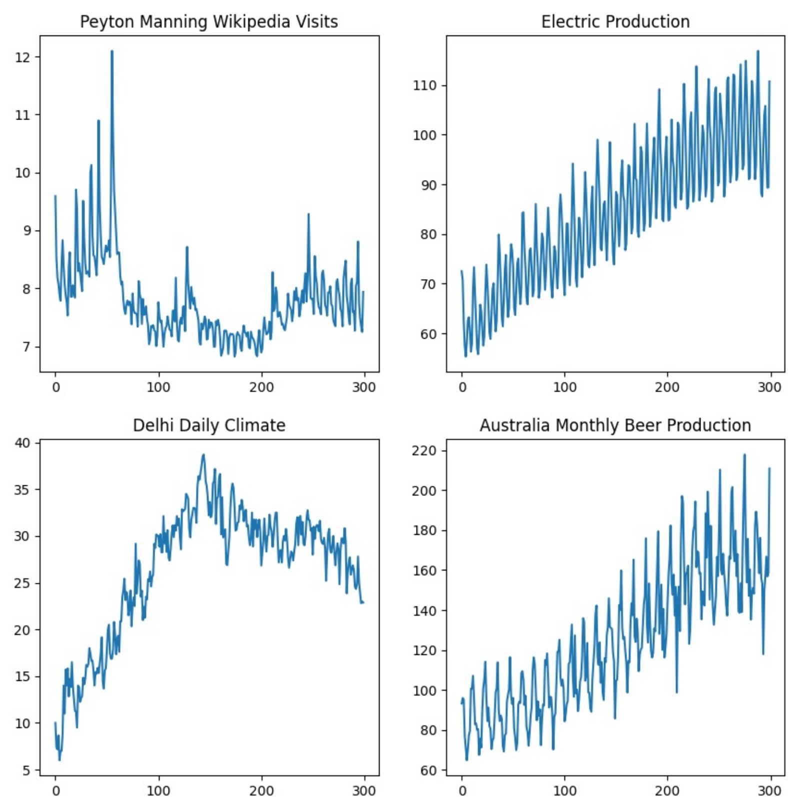

Figure 1. Four real-world datasets: plots show raw values (not Fourier ratios), so the y-axis is the observed series value.

These data sets are unrelated to one another as they are dominated by event-driven fluctuations, like football game times, global weather patterns, etc. These examples, with Fourier ratios close to the minimal value of 1, give heuristic credibility to the definition of randomness detection studied in this paper.

2.2. **Talagrand's Constant.** In this section, we are exploring the constants C_T (Talagrand's Constant) and $C(q)$. (Bourgain's Constant) In lieu of Theorem 1.3, there exists $\gamma_0 \in (0,1)$ such that if $h : \mathbb{Z}_N \to \mathbb{C}$ supported in a generic set M (in the sense of Definition 1.2) of size $\gamma_0 \frac{N}{\log(N)}$, then with probability $1 - o_N(1)$,

$$\left(\frac{1}{N}\sum_{m\in\mathbb{Z}_N} \left|\widehat{h}(m)\right|^2\right)^{\frac{1}{2}} \leqslant C_T(\log(N)\log\log(N))^{\frac{1}{2}} \cdot \frac{1}{N}\sum_{m\in\mathbb{Z}_N} |\widehat{h}(m)|, \tag{2.1}$$

where $C_T > 0$ is a constant that depends only on γ_0.

In lieu of Remark 1.7, when $|M| = O(N^{1-\varepsilon})$ for some $\varepsilon > 0$, then we may replace $C_T(\log(N)\log\log(N))^{\frac{1}{2}}$ with just C_T'. So in this case, we will take Talagrand's constant to be C_T'. We will write this as C_T for the remainder of this section, as we will let $|M| = O(N^{\frac{2}{q}})$ so $\varepsilon = \frac{q-2}{q}$ for some $q > 2$. We will now write 2.1 as

$$\left(\frac{1}{N}\sum_{m\in\mathbb{Z}_N} \left|\widehat{h}(m)\right|^2\right)^{\frac{1}{2}} \leqslant C_T\frac{1}{N}\sum_{m\in\mathbb{Z}_N} |\widehat{h}(m)| \tag{2.2}$$

Now we look at Theorem 1.5. Suppose that M is generic, as above, $|M| = \lceil N^{\frac{2}{q}} \rceil$, $q > 2$. Then for all $f : \mathbb{Z}_N \to \mathbb{C}$, supported in M,

$$\left(\frac{1}{N}\sum_{m\in\mathbb{Z}_N} \left|\widehat{f}(m)\right|^q\right)^{\frac{1}{q}} \leqslant C(q) \cdot \left(\frac{1}{N}\sum_{m\in\mathbb{Z}_N} \left|\widehat{f}(m)\right|^2\right)^{\frac{1}{2}} \tag{2.3}$$

Looking at Lemma 1.6 we apply Hölder's inequality to 2.3 to see that

$$\left(\frac{1}{N}\sum_{x\in\mathbb{Z}_N} \left|\widehat{f}(x)\right|^2\right)^{\frac{1}{2}} \leqslant (C(q))^{\frac{q}{q-2}} \cdot \frac{1}{N}\sum_{x\in\mathbb{Z}_N} |\widehat{f}(x)|.$$

Combining the above, we see that $C_T \leqslant C(q)^{\frac{q}{q-2}}$. Recalling the notation

$$\|g\|_{L^p(\mu)} = \left(\frac{1}{N}\sum_{x\in\mathbb{Z}_N} |g(x)|^p\right)^{\frac{1}{p}}, \tag{2.4}$$

we get that 2.2 becomes

$$\|\widehat{h}\|_{L^2(\mu)} \leqslant C_T\|\widehat{h}\|_{L^1(\mu)} \tag{2.5}$$

and 2.3 turns into

$$\|\widehat{f}\|_{L^q(\mu)} \leqslant C(q)\|\widehat{f}\|_{L^2(\mu)}. \tag{2.6}$$

Now, we want to perform numerical experiments to compute $\frac{\|\widehat{f}\|_{L^q(\mu)}}{\|\widehat{f}\|_{L^2(\mu)}}$ and $\frac{\|\widehat{f}\|_{L^2(\mu)}}{\|\widehat{f}\|_{L^1(\mu)}}$ so we may investigate if $C_T \approx C(q)^{\frac{q}{q-2}}$ or is C_T much smaller.

All of the following algorithms were implemented and tested in Python. We will also be computing the ratio of the norms using indicator functions on the random sets M. The values obtained are not for general functions f. Instead we will compute it when $f = \check{1}_M$. As these are random signals, there is expected to be some variance between the data for varying trials. In lieu of this, exact results won't be possible for evaluating these constants, but to counter this, we used large values of N and a million trials. Although computationally expensive, we were able to get a good estimate. In all these algorithms, we computed the ratios with random sizes M. We took the 90th percentile of all these trials as the estimate for the constants to avoid the far outliers. As for the Bourgain constant, the algorithm will compute the ratio $\frac{\|\widehat{f}\|_{L^q(\mu)}}{\|\widehat{f}\|_{L^2(\mu)}}$ and then raise this value to the $\frac{q}{q-2}$ power so we may compare C_T and $C(q)^{\frac{q}{q-2}}$.

2.3. **Computing Talagrand/Bourgain Constants.** Here, we present 7 numerical experiments which pin down estimates for $C(q)$ and C_T.

In the first algorithm, we first sought to see a relationship between $C(q)^{\frac{q}{q-2}}$ and q. We fixed $N = 100000$ and let $q \in [3, 4]$. Letting $q \to 2$ would make the exponent blow up, so we avoided computing the constant close to $q = 2$.

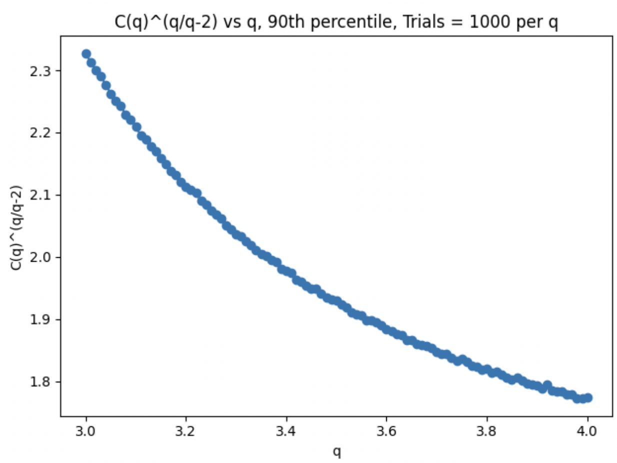

FIGURE 2. $C(q)^{\frac{q}{q-2}}$ vs q

In this second algorithm, we sought to perform a sanity check to see that Bourgain's constant does not depend on N. We fixed $q = 4$ and let $N \in [10, 100000]$.

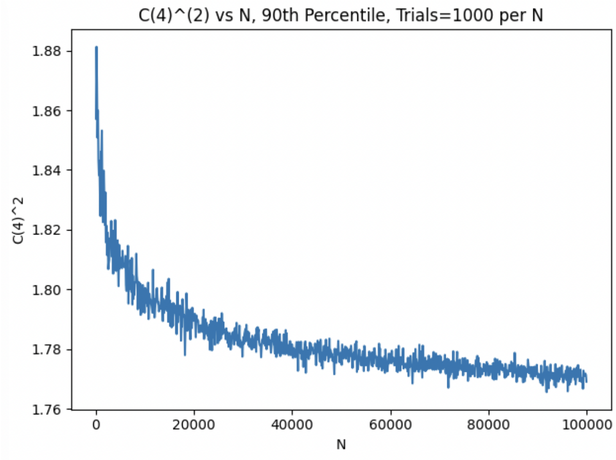

FIGURE 3. $C(q)^{\frac{q}{q-2}}$ vs N

In this third algorithm, we created a heatmap for varying q and N and see how the two variables affect $C(q)^{\frac{q}{q-2}}$.

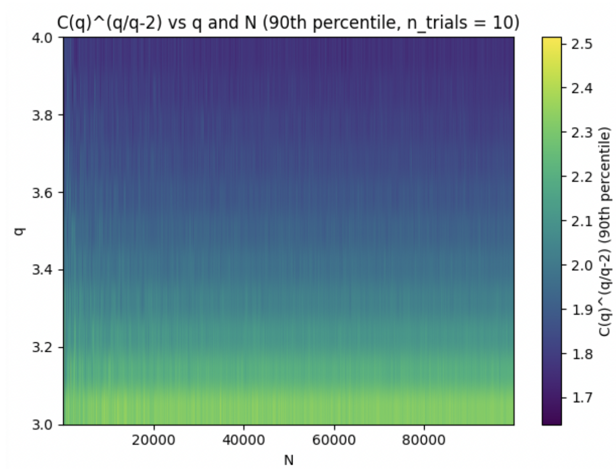

FIGURE 4. $C(q)^{\frac{q}{q-2}}$ vs q and N

In the fourth algorithm, we first sought to see a relationship between C_T and q. We fixed $N = 100000$ and let $q \in [3, 4]$ to match up with Bourgain's constant.

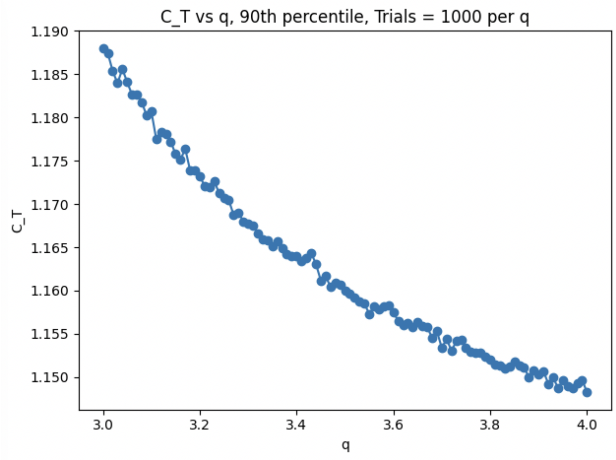

FIGURE 5. C_T vs q

In this fifth algorithm, we sought to see a relationship between C_T and N. We fixed $q = 4$ and let $N \in [10, 100000]$.

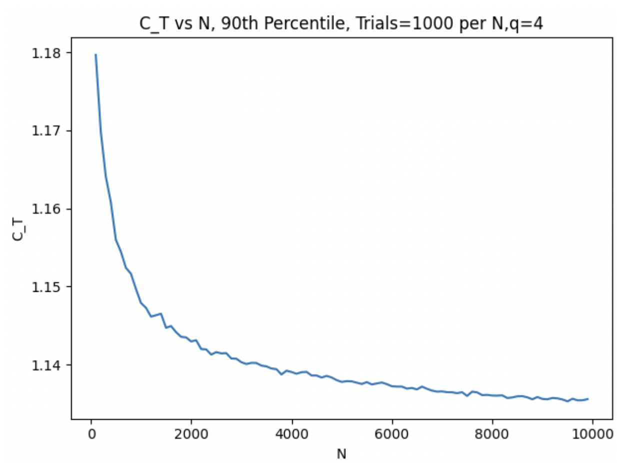

FIGURE 6. C_T vs N

In this sixth algorithm, we created a heatmap for varying q and N and see how the two variables affect C_T.

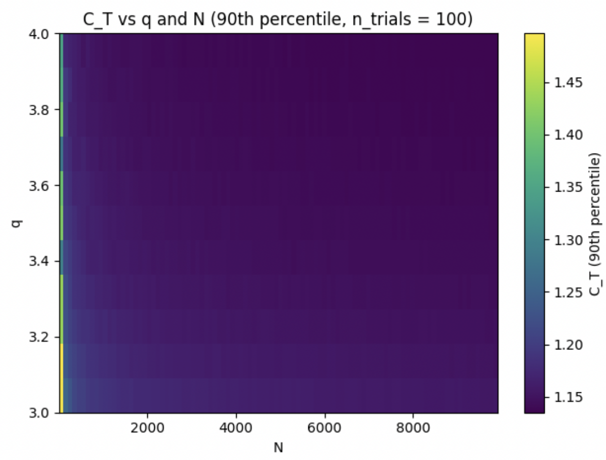

FIGURE 7. C_T vs q and N

In this final algorithm, we have a graph comparing the values of $C(q)^{\frac{q}{q-2}}$ and C_T varying both q and N.

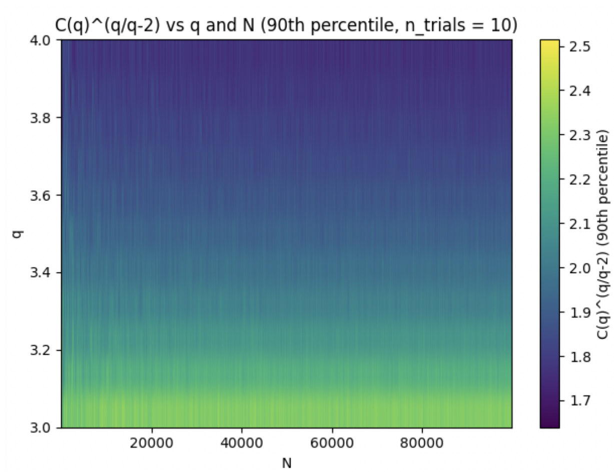

FIGURE 8. $(C(q))^{\frac{q}{q-2}}$ vs q and N

From the numerical experiments with Talagrand's constant, and Bourgain's constant, we observe that $C_T \leqslant (C(q))^{\frac{q}{q-2}}$. We see that C_T is approximated by $1 - 1.2$ whereas $(C(q))^{\frac{q}{q-2}}$ is approximated by $1.5 - 2.5$. So this can infer that Talagrand's Constant is a universal constant, and the N and q dependence is negligible, and it can be somewhat approximated by $(C(q))^{\frac{q}{q-2}}$.

3. Proofs of the main results

3.1. Proof of Theorem 1.8. Since f is L^2-concentrated on $E \subset \mathbb{Z}_N$ at level a ($a \in (0,1)$), we have

$$\|f\|_{L^2(E^c)} \leqslant \frac{a}{\sqrt{1-a^2}}\,\|f\|_{L^2(E)} \leqslant \frac{a}{1-a}\,\|f\|_{L^2(E)}, \qquad a \in (0,1),$$

$$\begin{aligned}\|f\|_2 &\leqslant \frac{\sqrt{1+a^2}}{(1-a)}\,\|f\|_{L^2(E)} \leqslant \frac{2}{1-a}\,\|f\|_{L^2(E)} \leqslant \frac{2}{1-a}\,|E|^{1/2}\,\|f\|_\infty \\ &\leqslant \frac{2}{1-a}\,|E|^{1/2}\,N^{-1/2}\,\|\widehat{f}\|_1.\end{aligned}$$

It follows that

$$\frac{(1-a)}{2}\cdot\sqrt{\frac{N}{|E|}} \leqslant \frac{\|\widehat{f}\|_1}{\|f\|_2} = FR(f). \tag{3.1}$$

Since $\widehat{f}$ is L^1-concentrated on S at level b, the same argument as above yields

$$||\widehat{f}||_1 \leqslant \frac{1}{1-b}\cdot||\widehat{f}||_{L^1(S)} \leqslant \frac{1}{1-b}\cdot|S|^{\frac{1}{2}}\cdot||\widehat{f}||_2.$$

Dividing both sides by $||\widehat{f}||_2$ yields

$$FR(f) \leqslant \frac{\sqrt{|S|}}{1-b}. \tag{3.2}$$

Combining (3.1) and (3.2) yields the conclusion of the theorem, i.e.,

$$|S|\,|E| \geqslant \frac{N}{4}(1-b)^2(1-a)^2.$$

3.2. Proof of Theorem 1.9. We have

$$\|\widehat{f}\|_2 \leqslant \|\widehat{1_M f}\|_2 + \|\widehat{1_{M^c} f}\|_2.$$

By Talagrand,

$$\begin{aligned}\|\widehat{1_M f}\|_2 = N^{\frac{1}{2}}\|\widehat{1_M f}\|_{L^2(\mu)} &\leqslant C_T\sqrt{N}\sqrt{\log(N)\log\log(N)}\|\widehat{1_M f}\|_{L^1(\mu)} \\ &= C_T N^{-\frac{1}{2}}\sqrt{\log(N)\log\log(N)}\|\widehat{1_M f}\|_1.\end{aligned}$$

By the concentration assumption,

$$\|\widehat{1_{M^c} f}\|_2 \leqslant r\|f\|_2.$$

It follows that

$$\|\widehat{f}\|_2(1-r) \leqslant C_T N^{-\frac{1}{2}}\sqrt{\log(N)\log\log(N)}\|\widehat{1_M f}\|_1,$$

and we conclude that

$$\|\widehat{f}\|_2 \leqslant \frac{C_T N^{-\frac{1}{2}}\sqrt{\log(N)\log\log(N)}}{1-r}\|\widehat{1_M f}\|_1.$$

We must now unravel $\|\widehat{1_M f}\|_1$. This expression equals

$$\begin{aligned}\|\widehat{f} - \widehat{1_{M^c} f}\|_1 &\leqslant \|\widehat{f}\|_1 + \|\widehat{1_{M^c} f}\|_1 \\ &\leqslant \|\widehat{f}\|_1 + rN^{\frac{1}{2}}\|\widehat{f}\|_2.\end{aligned}$$

It follows that

$$\|\widehat{f}\|_2 \cdot \left(1 - \frac{C_T r\sqrt{\log(N)\log\log(N)}}{1-r}\right) \leqslant \frac{C_T N^{-\frac{1}{2}}\sqrt{\log(N)\log\log(N)}}{1-r}\|\widehat{f}\|_1.$$

We conclude that

$$\frac{\|\widehat{f}\|_{L^1(\mu)}}{\|\widehat{f}\|_{L^2(\mu)}} \geqslant \frac{\left(1 - r\frac{C_T\sqrt{\log(N)\log\log(N)}}{1-r}\right)}{\frac{C_T\sqrt{\log(N)\log\log(N)}}{1-r}},$$

as claimed. $\square$

3.3. **Proof of Theorem 1.11.** The proof is probabilistic, so suppose $f : \mathbb{Z}_N \to \mathbb{C}$ and define the random function $Z : \mathbb{Z}_N \to \mathbb{C}$, where for each m,

$$Z(x) = \|\widehat{f}\|_1 \operatorname{sgn}(\widehat{f}(m)) N^{-\frac{1}{2}}\chi(mx)$$

with probability $\frac{|\widehat{f}(m)|}{\|\widehat{f}\|_1}$, where $\operatorname{sgn}(z) = \frac{z}{|z|}$. Observe that the expected value of Z is just our function f.

Let $Z_1, \ldots, Z_k$ be i.i.d. random functions with the same distribution as Z. Note that for each i, we have

$$|Z_i(x)| = N^{-\frac{1}{2}}\|\widehat{f}\|_1,$$

and thus

$$-N^{-\frac{1}{2}}\|\widehat{f}\|_1 \leqslant \operatorname{Re}(Z_i(x)) \leqslant N^{-\frac{1}{2}}\|\widehat{f}\|_1, \tag{3.3}$$

$$-N^{-\frac{1}{2}}\|\widehat{f}\|_1 \leqslant \operatorname{Im}(Z_i(x)) \leqslant N^{-\frac{1}{2}}\|\widehat{f}\|_1. \tag{3.4}$$

Next, by the union bound note that for each x we have

$$\begin{aligned}\mathbb{P}\left(\left|\mathbb{E}\left(Z(x)\right) - \frac{1}{k}\sum_{i=1}^k Z_i(x)\right| \geqslant \eta\right) &\leqslant \mathbb{P}\left(\left|\mathbb{E}\left(\operatorname{Re}(Z(x))\right) - \frac{1}{k}\sum_{i=1}^k \operatorname{Re}(Z_i(x))\right| \geqslant \frac{\eta}{2}\right) \\ &\quad + \mathbb{P}\left(\left|\mathbb{E}\left(\operatorname{Im}(Z(x))\right) - \frac{1}{k}\sum_{i=1}^k \operatorname{Im}(Z_i(x))\right| \geqslant \frac{\eta}{2}\right).\end{aligned}$$

By Hoeffding's inequality ([18]), as well as (3.3) and (3.4), we get that the right-hand side is bounded by

$$4\exp\left(-\frac{2\left(\frac{\eta k}{2}\right)^2}{k\left(2N^{-\frac{1}{2}}\|\widehat{f}\|_1\right)^2}\right) = 4\exp\left(-\frac{\eta^2 kN}{8\|\widehat{f}\|_1^2}\right). \tag{3.5}$$

Now, suppose the right-hand side in (3.5) is less than $\frac{1}{N}$, and set $P(x) = \frac{1}{k}\sum_{i=1}^k Z_i(x)$. Then by the union bound as well as the earlier observation that $\mathbb{E}[Z(x)] = f(x)$, we have

$$\mathbb{P}\left(\|f-P\|_\infty \geqslant \eta\right) \leqslant \sum_{x\in\mathbb{Z}_N} \mathbb{P}\left(|f(x)-P(x)| \geqslant \eta\right) < 1,$$

so there is a deterministic choice of $Z_1, \ldots, Z_k$ with $\|f-P\|_\infty < \eta$. The assumption that (3.5) is less than $\frac{1}{N}$ amounts to

$$4N\exp\left(-\frac{\eta^2 kN}{8\|\widehat{f}\|_1^2}\right) < 1,$$

i.e.,

$$\log(4N) - \frac{\eta^2 kN}{8\|\widehat{f}\|_1^2} < 0.$$

Then by setting η to equal $\varepsilon\|f\|_\infty$ and rearranging, we get that for any k such that

$$k > 8\left(\frac{\|\widehat{f}\|_{L^1(\mu)}}{\|f\|_{L^\infty}}\right)^2 \frac{N\log(4N)}{\varepsilon^2},$$

there is a P such that

$$\|f-P\|_\infty < \varepsilon\|f\|_\infty,$$

and we are done. $\square$

3.4. **Proof of Theorem 1.14.** We proceed as in the proof of Theorem 1.11, so suppose $f : \mathbb{Z}_N \to \mathbb{C}$, and again define the random function $Z : \mathbb{Z}_N \to \mathbb{C}$, where for each m,

$$Z(x) = \|\widehat{f}\|_1 \operatorname{sgn}(\widehat{f}(m)) N^{-\frac{1}{2}}\chi(mx)$$

with probability $\frac{|\widehat{f}(m)|}{\|\widehat{f}\|_1}$. We again have that $\mathbb{E}[Z(x)] = f(x)$. Moreover, note that for each x,

$$\begin{aligned}
\mathbb{E}|Z(x)|^2 &= \sum_{m\in\mathbb{Z}_N}\left|\|\widehat{f}\|_1 \frac{\widehat{f}(m)}{|\widehat{f}(m)|} N^{-\frac{1}{2}}\chi(mx)\right|^2 \cdot \frac{|\widehat{f}(m)|}{\|\widehat{f}\|_1} \\
&= \frac{1}{N}\|\widehat{f}\|_1 \sum_{m\in\mathbb{Z}_N} |\widehat{f}(m)| \\
&= \frac{1}{N}\|\widehat{f}\|_1^2,
\end{aligned}$$

and thus the variance of $Z(x)$ is

$$\operatorname{Var}(Z(x)) = \mathbb{E}|Z(x)|^2 - |\mathbb{E}[Z(x)]|^2 = \frac{1}{N}\|\widehat{f}\|_1^2 - |f(x)|^2.$$

Now, let $Z_1, \dots, Z_k$ be random i.i.d. functions with distribution Z, and define the random trigonometric polynomial P by

$$P(x) = \frac{1}{k}\sum_{i=1}^{k} Z_i(x).$$

Note that $\mathbb{E}[P(x)] = f(x)$ and that by independence,

$$\mathrm{Var}(P(x)) = \frac{1}{k}\mathrm{Var}(Z(x)),$$

and thus

$$\begin{aligned}
\mathbb{E}\|f - P\|_2^2 &= \sum_{x\in\mathbb{Z}_N} \mathbb{E}|f(x) - P(x)|^2 \\
&= \sum_{x\in\mathbb{Z}_N} \mathrm{Var}(P(x)) \\
&= \frac{1}{k}\sum_{x\in\mathbb{Z}_N} \mathrm{Var}(Z(x)) \\
&= \frac{1}{k}\sum_{x\in\mathbb{Z}_N} \frac{1}{N}\|\widehat{f}\|_1^2 - |f(x)|^2 \\
&= \frac{1}{k}\left(\|\widehat{f}\|_1^2 - \|f\|_2^2\right).
\end{aligned}$$

Now, if we assume this final value is less than $\eta^2\|f\|_2^2$, then there exists a deterministic choice of P such that $\|f - P\|_2 < \eta\|f\|_2$.
This assumption on k amounts to

$$\begin{aligned}
k &> \frac{1}{\eta^2}\cdot\frac{\|\widehat{f}\|_1^2 - \|f\|_2^2}{\|f\|_2^2} \\
&= \frac{1}{\eta^2}\left(\frac{\|\widehat{f}\|_1^2}{\|f\|_2^2} - 1\right) \\
&= \frac{\mathrm{FR}(f)^2 - 1}{\eta^2},
\end{aligned}$$

and thus for any such k, there is a trigonometric polynomial P with $\|f - P\|_2 < \eta\|f\|_2$, and we are done. □

3.5. **Proof of Theorem 1.16.** Again define the random function $Z : \mathbb{Z}_N \to \mathbb{C}$ by choosing $m \in \mathbb{Z}_N$ with probability $\frac{|\widehat{f}(m)|}{\|\widehat{f}\|_1}$ and setting

$$Z(x) = \|\widehat{f}\|_1 \,\mathrm{sgn}(\widehat{f}(m)) N^{-1/2}\chi(mx).$$

If $Z_1, \dots, Z_k$ are independent copies of Z, and if

$$P(x) = \frac{1}{k}\sum_{i=1}^{k} Z_i(x)$$

Then $\mathbb{E}[P(x)] = f(x)$, and moreover by the bound (3.5) from the proof of Theorem 1.11, we get that for every $x \in \mathbb{Z}_N$,

$$\mathbb{P}(|f(x) - P(x)| \geqslant t) \leqslant 4\exp\left(-\frac{t^2 kN}{8\|\widehat{f}\|_1^2}\right).$$

Consequently, we have that

$$\begin{aligned}
\mathbb{E}\|f - P\|_1 &= \sum_{x\in\mathbb{Z}_N} \mathbb{E}|f(x) - P(x)| \\
&= \sum_{x\in\mathbb{Z}_N} \int_0^\infty \mathbb{P}(|f(x) - P(x)| \geqslant t)dt \\
&\leqslant \sum_{x\in\mathbb{Z}_N} \int_0^\infty 4\exp\left(-\frac{t^2 kN}{8\|\widehat{f}\|_1^2}\right) dt \\
&= 4N \int_0^\infty \exp\left(-t^2 \cdot \frac{kN}{8\|\widehat{f}\|_1^2}\right) dt \\
&= 4N \left(\frac{\sqrt{8}\|\widehat{f}\|_1}{\sqrt{kN}}\right) \int_0^\infty \exp\left(-t^2\right) dt \\
&= \frac{2\sqrt{8\pi N}}{\sqrt{k}} \|\widehat{f}\|_1.
\end{aligned}$$

Now, if we assume this final value is less than $\eta\|f\|_1$, then there is a deterministic choice of P such that $\|f - P\|_1 < \eta\|f\|_1$.
This assumption amounts to

$$\frac{2\sqrt{8\pi N}}{\sqrt{k}} \|\widehat{f}\|_1 < \eta\|f\|_1,$$

or in other words

$$k > 32\pi \left(\frac{\|\widehat{f}\|_1}{\|f\|_1}\right)^2 \frac{N}{\eta^2}.$$

Thus for any such k, there is a trigonometric polynomial P with $\|f - P\|_1 < \eta\|f\|_1$, which completes the proof.

3.6. **Proof of Theorem 1.19.** Let $f : \mathbb{Z} \to \mathbb{C}$ and let $\varepsilon > 0$. If $k = \frac{\mathrm{FR}(f)^2}{\varepsilon^2}$, then by Theorem 1.14, there exists a trigonometric polynomial P of the form

$$P(x) = \sum_{i=1}^{k} c_i \chi(m_i x)$$

such that $\|f - P\|_2 < \varepsilon\|f\|_2$. We show that there is another trigonometric polynomial P' with low Kolmogorov complexity such that $\|P - P'\|_2 < \varepsilon\|f\|_2$, since this will imply $r_f(2\varepsilon) \leqslant \mathrm{K}(P')$.

First, note that

$$\begin{aligned}\widehat{P}(m_j) &= \frac{1}{\sqrt{N}} \sum_{x \in \mathbb{Z}_N} \chi(-xm_j)P(x)\\ &= \frac{1}{\sqrt{N}} \sum_{x \in \mathbb{Z}_N} \sum_{i=1}^{k} c_i \chi(x(m_i - m_j))\\ &= \sqrt{N} c_j.\end{aligned}$$

Now, let M be a constant to be chosen later, and define P' to be

$$P'(x) = \sum_{i=1}^{k} c_i' \chi(m_i x),$$

where c_i' is c_i, truncated to M digits after the decimal point. Note that since

$$\begin{aligned}|c_j'| &\leqslant |c_j|\\ &= \frac{1}{\sqrt{N}} \left|\widehat{P}(m_j)\right|\\ &\leqslant \frac{1}{\sqrt{N}} \|\widehat{P}\|_2\\ &= \frac{1}{\sqrt{N}} \|P\|_2\\ &\leqslant \frac{(1+\varepsilon)\|f\|_2}{\sqrt{N}},\end{aligned}$$

and since each c_j' contains at most M digits after the decimal point, we can encode each c_j' in length

$$C_U \left(M + \log\left(\frac{(1+\varepsilon)\|f\|_2}{\sqrt{N}} \right) \right),$$

where C_U is some constant depending on the fixed universal Turing machine.

Since encoding each m_i takes length $C_U \log N$, we see that we can encode P' using a Turing machine of length

$$C_U k \left(M + \log\left(\frac{(1+\varepsilon)\|f\|_2}{\sqrt{N}} \right) + \log N \right) + C_U',$$

where C_U' is another constant depending on the universal Turing machine, giving the length required to obtain the trigonometric polynomial P' from the values of c_i' and m_i. This thus gives a bound on $\mathrm{K}(P')$.

Next, note that since c_i' is obtained by truncating c_i, we have that

$$|c_i - c_i'| \leqslant 2^{-M},$$

so that

$$\begin{aligned}\|P-P'\|_2^2 &= \|\widehat{P}-\widehat{P'}\|_2^2\\ &= N\sum_{i=1}^{k}|c_i-c_i'|^2\\ &\leqslant Nk2^{-2M}.\end{aligned}$$

Thus, if we want $\|P-P'\|_2^2 \leqslant \varepsilon^2\|f\|_2^2$, we need that

$$k2^{-2M}N \leqslant \varepsilon^2\|f\|_2^2,$$

so in other words that

$$M \geqslant \frac{1}{2}\log\left(\frac{kN}{\varepsilon^2\|f\|_2^2}\right) = \log\left(\frac{\sqrt{k}\sqrt{N}}{\varepsilon\|f\|_2}\right).$$

Taking M such that equality holds above, we have that

$$\begin{aligned}\mathrm{K}(P') &\leqslant C_U k\left(M+\log\left(\frac{(1+\varepsilon)\|f\|_2}{\sqrt{N}}\right)+\log N\right)+C_U'\\ &= C_U k\left(\log\left(\frac{\sqrt{k}\sqrt{N}}{\varepsilon\|f\|_2}\right)+\log\left(\frac{(1+\varepsilon)\|f\|_2}{\sqrt{N}}\right)+\log N\right)+C_U'\\ &= C_U k\log\left(\frac{(1+\varepsilon)N\sqrt{k}}{\varepsilon}\right)+C_U'.\end{aligned}$$

and thus since

$$\|f-P'\|_2 \leqslant \|f-P\|_2+\|P-P'\|_2 < 2\varepsilon\|f\|_2,$$

we get that

$$\begin{aligned}r_f(2\varepsilon) &\leqslant C_U k\log\left(\frac{(1+\varepsilon)N\sqrt{k}}{\varepsilon}\right)+C_U'\\ &\leqslant O\left(k\log\left(\frac{(1+\varepsilon)N\sqrt{k}}{\varepsilon}\right)\right).\end{aligned}$$

□

3.7. **Proof of Theorem 1.20.** We shall need the following classical definition. See e.g. [6], [7], and [29].

Definition 3.1 (Restricted Isometry Condition)**.** A matrix $A \in \mathbb{R}^{m\times n}$ is said to satisfy the *restricted isometry condition of order* k with constant $\delta_k \in (0,1)$ if for all vectors $x \in \mathbb{R}^n$ with at most k nonzero entries and for sufficient large C, the following holds:

$$C(1-\delta_k)\|x\|_2^2 \leqslant \|Ax\|_2^2 \leqslant C(1+\delta_k)\|x\|_2^2.$$

Define $\Phi := \left(\frac{\chi(s\cdot r)}{\sqrt{N}}\right)_{s,r} \in \mathbb{C}^{N\times N}$ to be the $N\times N$ inverse Fourier matrix. Set $S = \frac{r^2}{\varepsilon^2}$. Consider $\delta_{4S} = 1/4$, a restricted isometry constant corresponding to vectors with at most

$4S$ non-zero entries. Noting that $\delta_{3S} \leqslant \delta_{4S}$. Sample $q' = CS\log^2(S)\log(N)$ rows of Φ uniformly and independently, denoting $q \leqslant q'$ to be the number of distinct rows. Denote the resulting matrix of distinct rows as $A \in \mathbb{C}^{q\times N}$. Then, with probability $1-2^{-\Omega((\log N)\cdot(\log S))}$, A satisfies the restricted isometry condition with $\delta_{4S} = 1/4$, by Theorem 4.5 of [17], and with high-probability the condition

$$\delta_{3S} + 3\delta_{4S} < 2 \tag{3.6}$$

is satisfied.

Let $f \in \mathcal{C}(r)$. By Theorem 1.14, there exists an S-sparse polynomial P, such that,

$$\|f - P\|_2 \leqslant \|f\|_2 \cdot \varepsilon.$$

Let f_X be the restriction of f on X, $y = f_X, x_0 = \widehat{P}$. We get $\|y - Ax_0\| = \|f_X - A\widehat{P}\|_2 = \|f - P\|_{L^2(X)} \leqslant \|f - P\|_2 \leqslant \|f\|_2 \cdot \varepsilon$. Consider the following optimization problem:

$$(*):\ \min \|\widehat{x}\|_1 \text{ subject to } \|f - x\|_{L^2(X)} \leqslant \eta.$$

The following result (Theorem 1 of [9]) gives us guarantees for (*).

Theorem 3.2. *Let S be such that $\delta_{3S} + 3\delta_{4S} < 2$. Then for any signal x_0 supported on T_0 with $|T_0| \leqslant S$ and any perturbation e with $\|e\|_2 \leqslant \eta$, the solution x^* to (*) obeys*

$$\|\widehat{x^*} - x_0\|_2 \leqslant C_S \cdot \eta$$

where the constant C_S may only depend on δ_{4S}. For reasonable values of δ_{4S}, C_S is well behaved; e.g. $C_S \approx 8.82$ for $\delta_{4S} = 1/5$ and $C_S \approx 10.47$ for $\delta_{4S} = 1/4$.

Remark 3.3. Theorem 3.2 is stated for real-valued matrices. However, the authors note in [9], the real assumption is for proof simplicity, and the Theorem also applies to complex-valued orthogonal matrices, such as the Fourier matrix.

Putting everything together, setting $\eta = \|f\|_2 \cdot \varepsilon$, and using equation (3.6) and Theorem 3.2, we obtain

$$\|\widehat{x^*} - \widehat{P}\|_2 \leqslant 10.47 \cdot \|f\|_2 \cdot \varepsilon.$$

By Parseval's Theorem, $\|\widehat{x^*} - \widehat{P}\|_2 = \|x - P\|_2$, so by using triangle's inequality, we get

$$\|x^* - f\|_2 \leqslant \|x^* - P\|_2 + \|f - P\|_2 \leqslant 11.47 \cdot \|f\|_2 \cdot \varepsilon.$$

3.8. **Proof of Theorem 1.22.** Let $\sigma_x = \mathbf{1}_{\{x\in X\}}$ so that $f_X = f \cdot \mathbf{1}_X$ and $\mathbb{E}\sigma_x = p$.

3.8.1. *Control of $\|\widehat{f_X}\|_2$.* By Plancherel, $\|\widehat{f_X}\|_2 = \|f_X\|_2$ and

$$\|f_X\|_2^2 = \sum_{x=0}^{N-1} \sigma_x |f(x)|^2, \qquad \mathbb{E}\|f_X\|_2^2 = p\|f\|_2^2.$$

The summands are independent, bounded by $\|f\|_\infty^2$, and have variance at most $p\|f\|_\infty^2|f(x)|^2$. A standard Bernstein bound yields, for some absolute $c_1 > 0$,

$$\mathbb{P}\Big(\big|\|f_X\|_2^2 - p\|f\|_2^2\big| > t\Big) \leqslant 2\exp\left(-c_1 \min\left\{\frac{t^2}{p\|f\|_\infty^2\|f\|_2^2}, \frac{t}{\|f\|_\infty^2}\right\}\right).$$

Choose $t = c_2\,\varepsilon^2 p\,\|f\|_2^2$ and use $\|f\|_\infty^2 \leqslant \mu(f)\|f\|_2^2/N$ to obtain

$$\big|\|f_X\|_2^2 - p\|f\|_2^2\big| \leqslant c_2\,\varepsilon^2 p\,\|f\|_2^2$$

with probability at least $1 - 2e^{-u}$ provided $p \geqslant C_0 \frac{\mu(f)}{\varepsilon^2}\frac{u}{N}$. Taking square roots and using $|a-b| \leqslant \varepsilon b$ implies

$$\big|\|f_X\|_2 - \sqrt{p}\,\|f\|_2\big| \leqslant \varepsilon\sqrt{p}\,\|f\|_2.$$

By Plancherel again, this is the claimed bound for $\|\widehat{f_X}\|_2$.

3.8.2. *A decomposition of* $\widehat{f_X}$. With our normalization,

$$\widehat{f_X} = \widehat{f \cdot \mathbf{1}_X} = N^{-1/2}\,\widehat{f} * \widehat{\mathbf{1}_X}.$$

Decompose $\widehat{\mathbf{1}_X} = p\sqrt{N}\,\delta_0 + W$, where

$$W(m) = N^{-1/2}\sum_{x=0}^{N-1}(\sigma_x - p)e^{-2\pi i x m/N}, \quad \mathbb{E}W(m) = 0, \quad \mathbb{E}|W(m)|^2 = p(1-p).$$

Hence

$$\widehat{f_X} = p\widehat{f} + \Delta, \qquad \Delta = N^{-1/2}(\widehat{f} * W).$$

Equivalently, in the time domain

$$\Delta(m) = N^{-1}\sum_{x=0}^{N-1}(\sigma_x - p)f(x)e^{-2\pi i x m/N}.$$

Thus, as a vector in $\mathbb{C}^{\mathbb{Z}_N}$,

$$\Delta = \sum_{x=0}^{N-1}(\sigma_x - p)\,v_x, \qquad v_x(m) = N^{-1}f(x)e^{-2\pi i x m/N}.$$

Note the norms $\|v_x\|_1 = |f(x)|$ and $\|v_x\|_2 = N^{-1/2}|f(x)|$.

3.8.3. *Some tools from probability.* Symmetrization inequality for norms. Let $Z_1, \dots, Z_n$ be independent mean-zero random vectors in a normed space, and let $\varepsilon_1, \dots, \varepsilon_n$ be i.i.d. Rademacher signs ($\mathbb{P}(\varepsilon_i = \pm 1) = 1/2$), independent of the Z_i. Then

$$\mathbb{E}\Big\|\sum_{i=1}^n Z_i\Big\| \leqslant 2\,\mathbb{E}\Big\|\sum_{i=1}^n \varepsilon_i Z_i\Big\|.$$

To see this, let $Z_1', \dots, Z_n'$ be an independent copy of $Z_1, \dots, Z_n$, and write $\mathbb{E}'$ for expectation over the copy. By Jensen and convexity of the norm,

$$\mathbb{E}\Big\|\sum_i Z_i\Big\| = \mathbb{E}\Big\|\mathbb{E}'\sum_i (Z_i - Z_i')\Big\| \leqslant \mathbb{E}\mathbb{E}'\Big\|\sum_i (Z_i - Z_i')\Big\|.$$

Conditional on (Z_i, Z_i'), the distribution of $\sum_i (Z_i - Z_i')$ is the same as that of $\sum_i \varepsilon_i (Z_i - Z_i')$. Hence

$$\mathbb{E}\Big\|\sum_i Z_i\Big\| \leqslant \mathbb{E}\Big\|\sum_i \varepsilon_i (Z_i - Z_i')\Big\| \leqslant \mathbb{E}\Big\|\sum_i \varepsilon_i Z_i\Big\| + \mathbb{E}\Big\|\sum_i \varepsilon_i Z_i'\Big\| = 2\,\mathbb{E}\Big\|\sum_i \varepsilon_i Z_i\Big\|.$$

This proves the symmetrization inequality.

Khintchine inequality for $p = 1$ (scalar case). Let $\varepsilon_1, \dots, \varepsilon_n$ be i.i.d. Rademacher random variables, and $a_1, \dots, a_n \in \mathbb{C}$. Then

$$\mathbb{E}\Big|\sum_{i=1}^{n} \varepsilon_i a_i\Big| \leqslant \Big(\sum_{i=1}^{n} |a_i|^2\Big)^{1/2}.$$

This is the classical Khintchine inequality with optimal constant 1 at $p = 1$. We will apply it coordinatewise to vector-valued sums.

3.8.4. *Bounding* $\mathbb{E}\|\Delta\|_1$ *by symmetrization and Khintchine.* Apply the symmetrization inequality to $Z_x = (\sigma_x - p)v_x$:

$$\mathbb{E}\|\Delta\|_1 = \mathbb{E}\Big\|\sum_x (\sigma_x - p)v_x\Big\|_1 \leqslant 2\,\mathbb{E}\Big\|\sum_x \varepsilon_x(\sigma_x - p)v_x\Big\|_1.$$

Expand the ℓ_1 norm as a sum over coordinates $m \in \mathbb{Z}_N$ and use Fubini to move expectations inside:

$$\mathbb{E}\Big\|\sum_x \varepsilon_x(\sigma_x - p)v_x\Big\|_1 = \sum_{m\in\mathbb{Z}_N} \mathbb{E}\Big|\sum_x \varepsilon_x(\sigma_x - p)v_x(m)\Big|.$$

For each fixed m, condition on the variables (σ_x) and apply the scalar Khintchine inequality to the Rademacher series in x:

$$\mathbb{E}_\varepsilon\Big|\sum_x \varepsilon_x(\sigma_x - p)v_x(m)\Big| \leqslant \Big(\sum_x |(\sigma_x - p)v_x(m)|^2\Big)^{1/2}.$$

Now take expectation over (σ_x) and use $\mathbb{E}(\sigma_x - p)^2 = p(1-p)$ and $|v_x(m)| = N^{-1}|f(x)|$ (independent of m):

$$\mathbb{E}\Big(\sum_x |(\sigma_x - p)v_x(m)|^2\Big)^{1/2} \leqslant \Big(\sum_x \mathbb{E}(\sigma_x - p)^2|v_x(m)|^2\Big)^{1/2} = \sqrt{p(1-p)}\Big(\sum_x |v_x(m)|^2\Big)^{1/2}.$$

Since $\sum_x |v_x(m)|^2 = N^{-2}\sum_x |f(x)|^2 = N^{-2}\|f\|_2^2$, we obtain, for each m,

$$\mathbb{E}\Big|\sum_x \varepsilon_x(\sigma_x - p)v_x(m)\Big| \leqslant \sqrt{p(1-p)} \cdot \frac{\|f\|_2}{N}.$$

Summing over all $m \in \mathbb{Z}_N$ gives

$$\mathbb{E}\Big\|\sum_x \varepsilon_x(\sigma_x - p)v_x\Big\|_1 \leqslant N \cdot \sqrt{p(1-p)} \cdot \frac{\|f\|_2}{N} = \sqrt{p(1-p)}\,\|f\|_2.$$

Finally, restore the factor 2 from symmetrization:

$$\mathbb{E}\|\Delta\|_1 \leqslant 2\sqrt{p(1-p)}\,\|f\|_2 \leqslant 2\sqrt{p}\,\|f\|_2.$$

3.8.5. *High-probability control of* $\|\Delta\|_1$. Consider the function of the N independent Bernoulli variables $\sigma = (\sigma_x)_x$,

$$\Phi(\sigma) = \Bigg\|\sum_x (\sigma_x - p)v_x\Bigg\|_1 = \|\Delta\|_1.$$

If we change only the coordinate σ_x to another value in $\{0,1\}$ while keeping all other σ_y fixed, the vector inside $\|\cdot\|_1$ changes by at most $|\,\sigma_x - \sigma_x'\,|\,\|v_x\|_1 \leqslant \|v_x\|_1 = |f(x)|$. Hence Φ

has bounded differences with constants $c_x = |f(x)|$. McDiarmid's inequality yields, for all $t > 0$,

$$\mathbb{P}\big(\Phi - \mathbb{E}\Phi \geqslant t\big) \leqslant \exp\left(-\frac{2t^2}{\sum_x c_x^2}\right) = \exp\left(-\frac{2t^2}{\|f\|_2^2}\right).$$

Equivalently, with probability at least $1 - e^{-u}$,

$$\|\Delta\|_1 \leqslant \mathbb{E}\|\Delta\|_1 + \frac{1}{\sqrt{2}}\sqrt{u}\,\|f\|_2 \leqslant 2\sqrt{p}\,\|f\|_2 + \frac{1}{\sqrt{2}}\sqrt{u}\,\|f\|_2.$$

This inequality is correct but not yet in the most useful scaling for small p. To express the deviation in terms that shrink when pN grows, we further bound $\|f\|_2 \leqslant \|f\|_\infty\sqrt{N}$ to obtain the coarser but handier form

$$\|\Delta\|_1 \leqslant C_1\left(\sqrt{p}\,\|f\|_2 + \|f\|_\infty\sqrt{Nu}\right) \quad \text{with probability at least } 1 - e^{-u},$$

for some absolute C_1.

3.8.6. *Control of $\|\widehat{f_X}\|_1$ and comparison.* From above,

$$\big|\|\widehat{f_X}\|_1 - p\|\widehat{f}\|_1\big| \leqslant \|\Delta\|_1.$$

Combine this with the bound from Step 5 and use $\|\widehat{f}\|_1 \geqslant \|\widehat{f}\|_2 = \|f\|_2$ to get

$$\frac{\big|\|\widehat{f_X}\|_1 - p\|\widehat{f}\|_1\big|}{p\|\widehat{f}\|_1} \leqslant \frac{C_1}{p}\left(\frac{\sqrt{p}\,\|f\|_2}{\|f\|_2} + \frac{\|f\|_\infty\sqrt{Nu}}{\|f\|_2}\right) \leqslant C_1\left(\frac{1}{\sqrt{p}} + \sqrt{\frac{\mu(f)}{pN}}\sqrt{u}\right).$$

If $p \geqslant C\frac{\mu(f)}{\varepsilon^2}\frac{\log N + u}{N}$ with C sufficiently large, then the right-hand side is at most ε, and therefore

$$\big|\|\widehat{f_X}\|_1 - p\|\widehat{f}\|_1\big| \leqslant \varepsilon p\|\widehat{f}\|_1$$

with probability at least $1 - e^{-u}$. Intersect with the high-probability event from Step 1 (probability at least $1 - 2e^{-u}$) to get both norm relations simultaneously.

3.8.7. *Bounds on $FR(f)$.* Let $A = \|\widehat{f_X}\|_1$, $B = \|\widehat{f_X}\|_2$, $a = p\|\widehat{f}\|_1$, $b = \sqrt{p}\,\|\widehat{f}\|_2$. In the case of the high probability event, we have $|A - a| \leqslant \varepsilon a$ and $|B - b| \leqslant \varepsilon b$. Then

$$\frac{A}{B} \leqslant \frac{a(1+\varepsilon)}{b(1-\varepsilon)} = \frac{1+\varepsilon}{\sqrt{1-\varepsilon}}\,\frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_2}, \qquad \frac{A}{B} \geqslant \frac{a(1-\varepsilon)}{b(1+\varepsilon)} = \frac{1-\varepsilon}{\sqrt{1+\varepsilon}}\,\frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_2}.$$

For $\varepsilon \in (0, 1/2)$ a straightforward calculation shows that

$$1 - 3\varepsilon \leqslant (1 \pm \varepsilon)/(1 \pm \varepsilon)^{1/2} \leqslant 1 + 3\varepsilon,$$

which gives the final assertion. This completes the proof.

3.9. **Proof of Theorem 1.24.** We write $\|f\|_2^2 = \sum_{x\in\mathbb{Z}_N} |f(x)|^2$ and note that

$$\|f_X\|_2^2 = \sum_{x\in\mathbb{Z}_N} |f(x)|^2 1_X(x),$$

and

$$\mathbb{E}\|f_X\|_2^2 = p\|f\|_2^2.$$

For each $x \in \mathbb{Z}_N$, let

$$Y_x = \big(1_X(x) - p\big)\,|f(x)|^2,$$

so that the Y_x are independent, mean zero, and

$$\|f_X\|_2^2 - p\|f\|_2^2 = \sum_x Y_x.$$

We will apply Bernstein's inequality to the sum $S = \sum_x Y_x$. First, for each x,

$$|Y_x| \leqslant |f(x)|^2 \leqslant \|f\|_\infty^2 =: M,$$

and

$$\mathrm{Var}(Y_x) = \mathrm{Var}(1_X(x)\,|f(x)|^4) = p(1-p)|f(x)|^4 \leqslant p\,|f(x)|^4.$$

Hence

$$\sigma^2 := \sum_x \mathrm{Var}(Y_x) \leqslant p\sum_x |f(x)|^4 \leqslant p\,\|f\|_\infty^2 \sum_x |f(x)|^2 = p\,\|f\|_\infty^2\|f\|_2^2.$$

Bernstein's inequality yields, for all $t > 0$,

$$\mathbb{P}\big(|S| \geqslant t\big) \leqslant 2\exp\Big(-\frac{t^2}{2(\sigma^2 + Mt/3)}\Big).$$

Choose

$$t = \varepsilon\, p\,\|f\|_2^2.$$

Using $\varepsilon \in (0,1)$, we bound the denominator by

$$\sigma^2 + Mt \leqslant p\,\|f\|_\infty^2\|f\|_2^2 + \varepsilon p\,\|f\|_\infty^2\|f\|_2^2 \leqslant 2p\,\|f\|_\infty^2\|f\|_2^2.$$

Therefore

$$\mathbb{P}\big(|\|f_X\|_2^2 - p\|f\|_2^2| \geqslant \varepsilon p\|f\|_2^2\big) \leqslant 2\exp\Big(-\frac{\varepsilon^2 p^2\|f\|_2^4}{4p\,\|f\|_\infty^2\|f\|_2^2}\Big) = 2\exp\Big(-c\,\varepsilon^2 p\,\frac{\|f\|_2^2}{\|f\|_\infty^2}\Big),$$

for a universal constant $c > 0$ (for instance $c = 1/4$). Since

$$\mu(f) = N\|f\|_\infty^2/\|f\|_2^2,$$

we have

$$\frac{\|f\|_2^2}{\|f\|_\infty^2} = \frac{N}{\mu(f)}.$$

Hence

$$\mathbb{P}\big(|\|f_X\|_2^2 - p\|f\|_2^2| \geqslant \varepsilon p\|f\|_2^2\big) \leqslant 2\exp\Big(-c\,\varepsilon^2 p\,\frac{N}{\mu(f)}\Big).$$

If

$$p \geqslant C\,\frac{\mu(f)}{\varepsilon^2}\,\frac{u}{N}$$

with $C \geqslant c^{-1}$, then the right side is at most $2e^{-u}$. Thus, with probability at least $1 - 2e^{-u}$,

$$|\|f_X\|_2^2 - p\|f\|_2^2| \leqslant \varepsilon p\|f\|_2^2.$$

To finish the proof, observe that

$$\left|\sqrt{a}-\sqrt{b}\right| = \frac{|a-b|}{\sqrt{a}+\sqrt{b}} \leqslant \frac{|a-b|}{\sqrt{b}}$$

valid for $a, b \geqslant 0$. Taking $a = \|f_X\|_2^2$ and $b = p\|f\|_2^2$ gives

$$\left|\|f_X\|_2 - \sqrt{p}\,\|f\|_2\right| \leqslant \frac{|\|f_X\|_2^2 - p\|f\|_2^2|}{\sqrt{p}\,\|f\|_2} \leqslant \varepsilon\sqrt{p}\,\|f\|_2,$$

and the proof is complete.

3.10. **Proof of Theorem 1.29.** If $\mathcal{B}(r)$'s maximum statistical query dimension is d, then $\mathcal{B}(r)$'s VC-dimension is $O(d)$ by Theorem 1.28. Hence, it suffices to calculate the statistical query dimension of $\mathcal{C}$ over each distribution $\mathcal{D}$.

Note first that for $f \in \mathcal{B}(r)$, since $f(x) \in \{-1, 1\}$ for each x, we have by Plancherel that

$$\|\widehat{f}\|_2 = \|f\|_2 = \sqrt{N},$$

and thus

$$\mathrm{FR}(f) = \frac{\|\widehat{f}\|_1}{\sqrt{N}}.$$

Fix $\varepsilon > 0$ and $\mathcal{D}$, a probability distribution on $X := \mathbb{Z}_N$. We will approximate $\mathcal{B}(r)$ with an ε-net using the probabilistic method. Fix $f \in \mathcal{B}(r)$ and define the random function, Z, where

$$Z(x) = \frac{\|\widehat{f}\|_1}{\sqrt{N}} \cdot \operatorname{sign}(\widehat{f}(m)) \cdot \chi(mx),$$

with probability $\frac{|\widehat{f}(m)|}{\|\widehat{f}\|_1}$ for each $m \in X$. Thus,

$$\mathbb{E}(Z) = f.$$

Draw $Z_1, \cdots, Z_k$ i.i.d. from the distribution of Z. We have that

$$\begin{aligned}
\mathbb{E}_Z\left(\mathbb{E}_{x\sim\mathcal{D}}\left[f(x) - \frac{1}{k}\sum_{i=1}^k Z_k(x)\right]^2\right) &= \sum_{x\in X} p(x)\cdot\mathbb{E}_Z\left[f(x) - \frac{1}{k}\sum_{i=1}^k Z_k(x)\right]^2 \\
&= \sum_{x\in X} p(x)\cdot\mathrm{Var}\left[\frac{1}{k}\sum_{i=1}^k Z_k(x)\right] = \sum_{x\in X}\frac{p(x)}{k^2}\cdot\sum_{i=1}^k \mathrm{Var}Z(x) \\
&\leqslant \sum_{x\in X}\frac{p(x)}{k}\cdot\mathbb{E}_Z\left[Z(x)\right]^2 \\
&= \sum_{x\in X}\frac{p(x)}{k}\cdot\frac{\|\widehat{f}\|_1^2}{N}\sum_{m\in X}\frac{|\widehat{f}(m)|}{\|\widehat{f}\|_1} \\
&= \frac{\mathrm{FR}(f)^2}{k} < \varepsilon^2
\end{aligned}$$

if $k > \frac{\mathrm{FR}(f)^2}{\varepsilon^2}$. Here, $p(x)$ is the probability mass of x under the distribution $\mathcal{D}$. Hence, by the probabilistic method, there exists a degree $k = \frac{\mathrm{FR}(f)^2}{\varepsilon^2}$ polynomial, P, such that

$$\|f - P\|_{L^2(\mathcal{D})} < \varepsilon.$$

We can divide the unit circle on $\mathbb{C}$ into $M = 1/\varepsilon$ pieces, $v_1, ..., v_M$. Note that the v_i are just the M-th roots of unity. Hence, P can be approximated by a degree k polynomial of the form,

$$\tilde{P}(x) = \frac{\|\widehat{f}\|_1}{k\sqrt{N}} \cdot \sum_{i=1}^{k} v_{s_i} \cdot \chi(m_i \cdot x),$$

where v_{s_i} is the closest element to $\mathrm{sign}(\widehat{f}(m_i))$. Indeed,

$$\|\tilde{P} - P\|_{L^2(\mathcal{D})} \leqslant \|\tilde{P} - P\|_\infty$$

$$= \left| \frac{\|\widehat{f}\|_1}{k\sqrt{N}} \cdot \sum_{i=1}^{k} \left(v_{s_i} - \widehat{f}(m) \right) \cdot \chi(m_i \cdot z) \right| \leqslant \frac{\|\widehat{f}\|_1}{k\sqrt{N}} \sum_{i=1}^{k} |v_{s_i} - \widehat{f}(m)|$$

$$\leqslant \frac{\|\widehat{f}\|_1}{k\sqrt{N}} \sum_{i=1}^{k} 2\varepsilon = 2\varepsilon \cdot \mathrm{FR}(f).$$

By the triangle inequality,

$$\|\tilde{P} - f\|_{L^2(\mathcal{D})} \leqslant \|\tilde{P} - P\|_{L^2(\mathcal{D})} + \|f - P\|_{L^2(\mathcal{D})} \leqslant \varepsilon + 2\varepsilon \cdot \mathrm{FR}(f). \tag{3.7}$$

Putting it all together, each $f \in \mathcal{B}(r)$ can be approximated by a polynomial $\tilde{P}$ up to an error of $\varepsilon(1 + 2\mathrm{FR}(f))$ (as in 3.7). There are at most

$$\binom{N}{k} \cdot M^k \leqslant \left(\frac{Ne}{k}\right)^k M^k = \frac{1}{\varepsilon^{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}}} \left(\frac{Ne}{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}} \right)^{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}}.$$

different choices of $\tilde{P}$. Hence, the covering number of $\mathcal{B}(r)$ is estimated by

$$\mathcal{N}(\mathcal{C}, \|\cdot\|_{L^2(\mathcal{D})}, \varepsilon(1 + 2\mathrm{FR}(f))) \leqslant \frac{1}{\varepsilon^{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}}} \left(\frac{Ne}{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}} \right)^{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}}.$$

Hence, the packing number of $\mathcal{B}(r)$ is estimated by,

$$\mathcal{P}(\mathcal{C}, \|\cdot\|_{L^2(\mathcal{D})}, 4\varepsilon(1 + 2\mathrm{FR}(f))) \leqslant \frac{1}{\varepsilon^{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}}} \left(\frac{Ne}{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}} \right)^{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}}.$$

Suppose that $f_1, .., f_d \in \mathcal{B}(r)$, where

$$d = \frac{1}{\varepsilon^{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}}} \left(\frac{Ne}{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}} \right)^{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}}. \tag{3.8}$$

By definition of packing, since $d > \mathcal{P}$, there exists $i \neq j$ such that

$$\|f_i - f_j\|_{L^2(\mathcal{D})} \leqslant 4\varepsilon(1 + 2\mathrm{FR}(f)).$$

Hence,

$$\mathbb{E}_{x\sim\mathcal{D}}\left(f_i(x)\cdot f_j(x)\right) \geqslant \frac{2 - (4\varepsilon(1+2\mathrm{FR}(f))^2}{2} > \frac{1}{d}.$$

This shows that we may set $\varepsilon = \frac{1}{4(1+2r)}$, where $r = \mathrm{FR}(f)$.

Hence, the statistical query dimension is bounded above by CN^α, where

$$C = \frac{1}{\varepsilon^{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}}} \cdot \left(\frac{e}{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}}\right)^{\frac{\mathrm{FR}(f)^2}{\varepsilon^2}}$$

and

$$\alpha = \left(\frac{\mathrm{FR}(f)}{\varepsilon}\right)^2,$$

with

$$\varepsilon = \frac{1}{4(1+2r)}.$$

This completes the proof.

3.11. **Proof of Theorem 1.31.** By the triangle and reverse triangle inequalities,

$$\|\widehat{f}\|_1 - \|\widehat{n}\|_1 \leqslant \|\widehat{f+n}\|_1 \leqslant \|\widehat{f}\|_1 + \|\widehat{n}\|_1,$$
$$\|\widehat{f}\|_2 - \|\widehat{n}\|_2 \leqslant \|\widehat{f+n}\|_2 \leqslant \|\widehat{f}\|_2 + \|\widehat{n}\|_2.$$

In terms of A, B, s, t this gives, for $t < s$,

$$\frac{\max\{A-B, 0\}}{s+t} \leqslant \mathrm{FR}(f+n) \leqslant \frac{A+B}{s-t}.$$

For the upper deviation,

$$\mathrm{FR}(f+n) - \frac{A}{s} \leqslant \frac{A+B}{s-t} - \frac{A}{s} = \frac{Bs+At}{s(s-t)} = \frac{B+(A/s)t}{s-t} = \frac{B+\mathrm{FR}(f)t}{s-t}.$$

For the lower deviation, using $\max\{A-B, 0\} \geqslant A - B$,

$$\frac{A}{s} - \mathrm{FR}(f+n) \leqslant \frac{A}{s} - \frac{A-B}{s+t} = \frac{Bs+At}{s(s+t)} \leqslant \frac{Bs+At}{s(s-t)} = \frac{B+\mathrm{FR}(f)t}{s-t}.$$

Combining the two one-sided bounds yields

$$\left|\mathrm{FR}(f+n) - \mathrm{FR}(f)\right| \leqslant \frac{B+\mathrm{FR}(f)t}{s-t} = \frac{\|\widehat{n}\|_1 + \mathrm{FR}(f)\|\widehat{n}\|_2}{\|\widehat{f}\|_2 - \|\widehat{n}\|_2}.$$

For the stated corollary, apply Cauchy–Schwarz to get $\|\widehat{n}\|_1 \leqslant \sqrt{N}\|\widehat{n}\|_2$ and $\|\widehat{f}\|_1 \leqslant \sqrt{N}\|\widehat{f}\|_2$, hence $\mathrm{FR}(f) \leqslant \sqrt{N}$, which gives the first displayed bound. The second bound follows from $\mathrm{FR}(f) \leqslant \sqrt{N}$. If $\|\widehat{n}\|_2 \leqslant \frac{1}{2}\|\widehat{f}\|_2$, then $\|\widehat{f}\|_2 - \|\widehat{n}\|_2 \geqslant \frac{1}{2}\|\widehat{f}\|_2$, yielding the final inequality.

3.12. **Proof of Theorem 1.32.** Set $A = \|\widehat{f}\|_1$, $B = \|\widehat{n}\|_1$, $s = \|\widehat{f}\|_2$, $t = \|\widehat{n}\|_2$. By the triangle and reverse triangle inequalities,

$$\|\widehat{f}\|_1 - \|\widehat{n}\|_1 \leqslant \|\widehat{f+n}\|_1 \leqslant \|\widehat{f}\|_1 + \|\widehat{n}\|_1, \qquad \|\widehat{f}\|_2 - \|\widehat{n}\|_2 \leqslant \|\widehat{f+n}\|_2 \leqslant \|\widehat{f}\|_2 + \|\widehat{n}\|_2.$$

For $t < s$ this implies

$$\frac{\max\{A-B,0\}}{s+t} \leqslant \mathrm{FR}(f+n) \leqslant \frac{A+B}{s-t}.$$

Hence

$$\mathrm{FR}(f+n) - \frac{A}{s} \leqslant \frac{A+B}{s-t} - \frac{A}{s} = \frac{Bs+At}{s(s-t)}, \qquad \frac{A}{s} - \mathrm{FR}(f+n) \leqslant \frac{Bs+At}{s(s+t)} \leqslant \frac{Bs+At}{s(s-t)}.$$

Therefore

$$|\mathrm{FR}(f+n) - \mathrm{FR}(f)| \leqslant \frac{Bs+At}{s(s-t)} = \frac{B+(A/s)t}{s-t}.$$

Since $A/s = \mathrm{FR}(f)$ and $B = t\,\mathrm{FR}(n)$, we obtain the deterministic bound

$$|\mathrm{FR}(f+n) - \mathrm{FR}(f)| \leqslant \frac{(\mathrm{FR}(n)+\mathrm{FR}(f))\,t}{s-t} \quad \text{whenever } t < s. \tag{3.9}$$

Because the DFT is unitary and n has i.i.d. circular complex Gaussian entries with variance σ^2, the Fourier coefficients $\widehat{n}(m)$ are i.i.d. circular complex Gaussian with the same variance. Equivalently,

$$\|\widehat{n}\|_2^2 = \sum_{m=0}^{N-1} |\widehat{n}(m)|^2 \sim \frac{\sigma^2}{2}\,\chi^2_{2N}.$$

Let $g \in \mathbb{R}^{2N}$ be standard normal and write $\widehat{n} = (\sigma/\sqrt{2})\,g$. We will use the following standard tail bounds at this point.

Theorem 3.4 (Gaussian concentration for Lipschitz functions; [34] (Prop. 3.1.8), [5] (Thm. 5.6), and [23] (Thm. 5.6 with the corollary to the Euclidean norm).)**.** *Let $g \sim \mathcal{N}(0, I_m)$ and let $f : \mathbb{R}^m \to \mathbb{R}$ be L-Lipschitz with respect to the Euclidean norm. Then for all $t \geqslant 0$,*

$$\Pr\big(f(g) \geqslant \mathbb{E}f(g) + t\big) \leqslant \exp\left(-\frac{t^2}{2L^2}\right), \qquad \Pr\big(|f(g) - \mathbb{E}f(g)| \geqslant t\big) \leqslant 2\exp\left(-\frac{t^2}{2L^2}\right).$$

In particular, taking $f(u) = \|u\|_2$ (which is 1-Lipschitz) and any $\gamma \in (0,1)$,

$$\Pr\left(\|g\|_2 \leqslant \sqrt{m} + \sqrt{2\log(1/\gamma)}\right) \geqslant 1-\gamma.$$

Applying this with $m = 2N$ gives

$$\|g\|_2 \leqslant \sqrt{2N} + \sqrt{2\log\left(\frac{1}{\gamma}\right)} \quad \text{with probability at least } 1-\gamma,$$

and therefore

$$\|\widehat{n}\|_2 \leqslant \frac{\sigma}{\sqrt{2}}\left(\sqrt{2N} + \sqrt{2\log\left(\frac{1}{\gamma}\right)}\right) = \sigma\left(\sqrt{N} + \sqrt{\log\left(\frac{1}{\gamma}\right)}\right) =: t_\gamma$$

with probability at least $1-\gamma$.

We also record a slightly sharper alternative that can replace the previous one when desired.

Theorem 3.5 (Laurent–Massart χ^2-inequality; [22] (Lemma 1, original source) and [35] (Prop. 2.1, a self-contained proof).)**.** *Let $X \sim \chi^2_m$. Then for all $x \geqslant 0$,*

$$\Pr\left(X - m \geqslant 2\sqrt{mx} + 2x\right) \leqslant e^{-x}, \qquad \Pr\left(m - X \geqslant 2\sqrt{mx}\right) \leqslant e^{-x}.$$

Equivalently, with probability at least $1 - e^{-x}$,

$$\sqrt{X} \leqslant \sqrt{m} + \sqrt{2x} + \frac{x}{\sqrt{m}}.$$

Taking $X = \|g\|_2^2$, $m = 2N$, and $x = \log(1/\gamma)$ yields

$$\|g\|_2 \leqslant \sqrt{2N} + \sqrt{2\log\left(\frac{1}{\gamma}\right)} + \frac{\log\left(\frac{1}{\gamma}\right)}{\sqrt{2N}}$$

with probability at least $1 - \gamma$, which translates to

$$\|\widehat{n}\|_2 \leqslant \sigma\left(\sqrt{N} + \sqrt{\log\left(\frac{1}{\gamma}\right)} + \frac{\log\left(\frac{1}{\gamma}\right)}{2\sqrt{N}}\right).$$

Returning to (3.10), on the event $\{\|\widehat{n}\|_2 \leqslant t_\gamma\}$ and provided $t_\gamma < s = \|\widehat{f}\|_2$, we obtain

$$|\mathrm{FR}(f+n) - \mathrm{FR}(f)| \leqslant \frac{(\mathrm{FR}(n) + \mathrm{FR}(f))\, t_\gamma}{s - t_\gamma}.$$

Finally, using $\|\widehat{f}\|_2 = s = \|f\|_2$ gives the equivalent form in terms of $\|f\|_2$.

3.13. **Proof of Theorem 1.33.** Write $\overline{n}(x) = \frac{1}{n}\sum_{i=1}^n n_i(x)$, so that $f = s + \overline{n}$. Since each $n_i(x)$ is mean zero and independent, $\mathbb{E}\,\overline{n}(x) = 0$, and by Gaussian stability $\overline{n}(x) \sim \mathcal{CN}(0, \sigma^2/n)$. This proves (a).

For (b), independence across x gives

$$\mathbb{E}\,\|f - s\|_2^2 = \sum_{x \in \mathbb{Z}_N} \mathbb{E}\,|\overline{n}(x)|^2 = \sum_{x \in \mathbb{Z}_N} \frac{\sigma^2}{n} = \frac{N\sigma^2}{n}.$$

For (c), stack real and imaginary parts of $\overline{n}$ into $g \in \mathbb{R}^{2N}$ with i.i.d. $\mathcal{N}(0,1)$ entries, so that

$$\overline{n} = \frac{\sigma}{\sqrt{2n}}\, g \quad \text{and} \quad \|\overline{n}\|_2 = \frac{\sigma}{\sqrt{2n}}\|g\|_2.$$

By Gaussian concentration for Lipschitz functions (take $f(u) = \|u\|_2$, which is 1-Lipschitz; see for example [34, 5, 23]), for all $t \geqslant 0$,

$$\Pr\left(\|g\|_2 \geqslant \sqrt{2N} + t\right) \leqslant e^{-t^2/2}.$$

Taking $t = \sqrt{2\log(1/\gamma)}$ yields

$$\|g\|_2 \leqslant \sqrt{2N} + \sqrt{2\log(1/\gamma)} \quad \text{with probability at least } 1 - \gamma,$$

hence

$$\|\overline{n}\|_2 \leqslant \frac{\sigma}{\sqrt{2n}}\left(\sqrt{2N} + \sqrt{2\log(1/\gamma)}\right) = \frac{\sigma}{\sqrt{n}}\left(\sqrt{N} + \sqrt{\log(1/\gamma)}\right)$$

with probability at least $1-\gamma$, proving the displayed inequality. Using the Laurent–Massart chi-square inequality for χ^2_{2N} (see [22], and also [35] for a modern proof) gives

$$\|g\|_2 \leqslant \sqrt{2N} + \sqrt{2\log(1/\gamma)} + \frac{\log(1/\gamma)}{\sqrt{2N}}$$

with probability at least $1-\gamma$, which translates to the refined bound after multiplying by $\sigma/\sqrt{2n}$.

3.14. **Proof of Theorem 1.35.** Let $x = \widehat{s}$ and $e = \widehat{u}$ for some perturbation u. By the triangle and reverse triangle inequalities,

$$\|x\|_1 - \|e\|_1 \leqslant \|x+e\|_1 \leqslant \|x\|_1 + \|e\|_1, \qquad \|x\|_2 - \|e\|_2 \leqslant \|x+e\|_2 \leqslant \|x\|_2 + \|e\|_2.$$

Writing $A = \|x\|_1$, $B = \|e\|_1$, $s_2 = \|x\|_2$, $t = \|e\|_2$, for $t < s_2$ one obtains

$$\frac{\max\{A-B,0\}}{s_2+t} \leqslant \frac{\|x+e\|_1}{\|x+e\|_2} \leqslant \frac{A+B}{s_2-t}.$$

A short rearrangement gives

$$\left|\frac{\|x+e\|_1}{\|x+e\|_2} - \frac{A}{s_2}\right| \leqslant \frac{B+(A/s_2)t}{s_2-t}.$$

Since $A/s_2 = \mathrm{FR}(s)$ and $B = \|\widehat{u}\|_1 = \mathrm{FR}(u)\|u\|_2$, and $\|x+e\|_1/\|x+e\|_2 = \mathrm{FR}(s+u)$, we have the deterministic bound

$$|\mathrm{FR}(s+u) - \mathrm{FR}(s)| \leqslant \frac{\big(\mathrm{FR}(u)+\mathrm{FR}(s)\big)\|u\|_2}{\|s\|_2 - \|u\|_2} \qquad \text{whenever } \|u\|_2 < \|s\|_2. \tag{3.10}$$

In particular, if $\|u\|_2 \leqslant \frac{1}{2}\|s\|_2$, then

$$|\mathrm{FR}(s+u) - \mathrm{FR}(s)| \leqslant \frac{2\|u\|_2}{\|s\|_2}\big(\mathrm{FR}(u)+\mathrm{FR}(s)\big). \tag{3.11}$$

We now establish the high-probability radii for the noises. Stack the real and imaginary parts of a complex vector into $\mathbb{R}^{2N}$. Since $n_i(x) \sim \mathcal{CN}(0,\sigma^2)$ are independent and the map $u \mapsto \|u\|_2$ is 1-Lipschitz, the same standard Gaussian norm concentration bound as the one we used in the proof of Theorem 1.32 implies

$$\Pr\big(\|n_i\|_2 \leqslant \sigma r_\gamma\big) \geqslant 1-\gamma, \qquad \Pr\big(\|\overline{n}\|_2 \leqslant \sigma r_\gamma/\sqrt{n}\big) \geqslant 1-\gamma,$$

where $r_\gamma = \sqrt{N} + \sqrt{\log(1/\gamma)}$.

We now apply the deterministic bound. Assume $\|s\|_2 \geqslant 2\sigma r_\gamma$. Then on the events above we have $\|\overline{n}\|_2 \leqslant \frac{1}{2}\|s\|_2$ and $\|n_i\|_2 \leqslant \frac{1}{2}\|s\|_2$. Apply (3.11) with $u = \overline{n}$ to get

$$|\mathrm{FR}(f) - \mathrm{FR}(s)| = |\mathrm{FR}(s+\overline{n}) - \mathrm{FR}(s)| \leqslant \frac{2\|\overline{n}\|_2}{\|s\|_2}\big(\mathrm{FR}(\overline{n})+\mathrm{FR}(s)\big) \leqslant \frac{2\sigma r_\gamma}{\sqrt{n}\|s\|_2}\big(\mathrm{FR}(\overline{n})+\mathrm{FR}(s)\big),$$

with probability at least $1-\gamma$. Similarly, with $u = n_i$,

$$|\mathrm{FR}(f_i) - \mathrm{FR}(s)| \leqslant \frac{2\|n_i\|_2}{\|s\|_2}\big(\mathrm{FR}(n_i)+\mathrm{FR}(s)\big) \leqslant \frac{2\sigma r_\gamma}{\|s\|_2}\big(\mathrm{FR}(n_i)+\mathrm{FR}(s)\big),$$

with probability at least $1-\gamma$ for each fixed i.

Finally, note that $\mathrm{FR}(\alpha e) = \mathrm{FR}(e)$ for any $\alpha > 0$, so $\mathrm{FR}(\overline{n})$ and $\mathrm{FR}(n_i)$ have the same distribution (both are the FR of a circular complex Gaussian vector up to a positive scale). This justifies the comparison stated in part (ii).

3.15. **Proof of Theorem 1.36.** Let Γ be defined as above. By our assumption on f as well as Markov's inequality, we have that

$$\mathrm{FR}(f)\|f\|_2 \geqslant \sum_{m\in\mathbb{Z}_N} |\widehat{f}(m)| \geqslant \eta N^{-\frac{1}{2}}\|f\|_2|\Gamma|,$$

and thus if f is nonzero,

$$|\Gamma| \leqslant \frac{\mathrm{FR}(f)}{\eta}\sqrt{N}.$$

Next, defining P as above, note that P is the inverse Fourier transform of $\widehat{f}\cdot 1_\Gamma$. Additionally, for $m \notin \Gamma$, we have that

$$|\widehat{f}(m)| < \eta\frac{\|f\|_2}{N^{\frac{1}{2}}}.$$

Thus, by the Plancherel, we have that

$$\begin{aligned}\|f-P\|_2 &= \|\widehat{f}-\widehat{P}\|_2\\ &= \|\widehat{f}-\widehat{f}1_\Gamma\|_2\\ &= \left(\sum_{m\notin\Gamma}|\widehat{f}(m)|^2\right)^{\frac{1}{2}}\\ &\leqslant \eta\|f\|_2,\end{aligned}$$

and we are done. □

3.16. **Proof of Theorem 1.39.** Our proof mostly follows Chang's original proof found in [11]. Recall

$$\Gamma = \left\{m\in\mathbb{Z}_N \;:\; |\widehat{f}(m)| \geqslant \eta\frac{\|f\|_2}{\sqrt{N}}\right\}.$$

Let Λ be a maximal dissociated subset of Γ, that is, a maximal subset with the property that all $\{-1,0,1\}$-linear combinations of elements of Λ are distinct, and note Γ must be contained in the $\{-1,0,1\}$-span of Λ.

Let

$$g(x) = \frac{1}{\|1_\Lambda\widehat{f}\|_2}\sum_{n\in\Lambda}\widehat{f}(n)\chi(xn),$$

and $p' = \frac{p}{p-1}$. Then

$$\begin{aligned}\|f\|_{L^{p'}(\mu)}\|g\|_{L^p(\mu)} &\geqslant \|fg\|_{L^1(\mu)} \\ &\geqslant \frac{1}{N\|1_\Lambda \widehat{f}\|_2}\left|\sum_{x\in\mathbb{Z}_N} f(x)\sum_{n\in\Lambda}\overline{\widehat{f}(n)}\chi(-xn)\right| \\ &= \frac{1}{N\|1_\Lambda \widehat{f}\|_2}\left|\sum_{n\in\Lambda}\overline{\widehat{f}(n)}\sum_{x\in\mathbb{Z}_N} f(x)\chi(-xn)\right| \\ &= \frac{\sqrt{N}}{N\|1_\Lambda \widehat{f}\|_2}\sum_{n\in\Lambda}|\widehat{f}(n)|^2 \\ &= \frac{1}{\sqrt{N}}\|1_\Lambda \widehat{f}\|_2 \\ &\geqslant \eta\|f\|_2\frac{\sqrt{|\Lambda|}}{N}.\end{aligned}$$

By definition,

$$\|f\|_{L^{p'}(\mu)} = N^{\frac{1}{p}-1}\|f\|_{p'}.$$

We also have that there is some absolute constant C such that

$$\|g\|_{L^p(\mu)} \leqslant C\sqrt{p}$$

by Rudin's inequality (see Lemma 4.33 in [32]). Combining the above yields

$$\eta\|f\|_2\frac{\sqrt{|\Lambda|}}{N} \leqslant C\sqrt{p}N^{\frac{1}{p}-1}\|f\|_{p'},$$

i.e.,

$$\sqrt{|\Lambda|} \leqslant C\eta^{-1}\sqrt{p}N^{\frac{1}{p}}\frac{\|f\|_{p'}}{\|f\|_2}.$$

Now, taking $p = \log N$, we obtain

$$\sqrt{|\Lambda|} \leqslant C\eta^{-1}\sqrt{\log N}\frac{\|f\|_{p'}}{\|f\|_2},$$

i.e.,

$$|\Lambda| \leqslant C'\eta^{-2}\log N\frac{\|f\|_{p'}^2}{\|f\|_2^2},$$

where $C' = (Ce)^2$, thus proving (1.22).

To prove (1.23), observe that

$$\begin{aligned}
\|f\|_{L^{p'}(\mu)} &= N^{\frac{1}{p}-1}\|f\|_{p'} \\
&= N^{\frac{1}{p}-1}\left(\sum_{x\in\mathbb{Z}_N}|f(x)|^{\frac{p}{p-1}}\right)^{\frac{p-1}{p}} \\
&= N^{\frac{1}{p}-1}\left(\sum_{x\in\mathbb{Z}_N}|f(x)||f(x)|^{\frac{1}{p-1}}\right)^{\frac{p-1}{p}} \\
&= N^{\frac{1}{p}-1}\left(\|f\|_1\sum_{x\in\mathbb{Z}_N}\frac{|f(x)|}{\|f\|_1}|f(x)|^{\frac{1}{p-1}}\right)^{\frac{p-1}{p}} \\
&\leqslant N^{\frac{1}{p}-1}\left(\|f\|_1\left(\sum_{x\in\mathbb{Z}_N}\frac{|f(x)|}{\|f\|_1}|f(x)|\right)^{\frac{1}{p-1}}\right)^{\frac{p-1}{p}}
\end{aligned}$$

by Jensen's inequality. This implies

$$\|f\|_{L^{p'}(\mu)} \leqslant N^{\frac{1}{p}-1}\|f\|_1^{\frac{p-1}{p}}\left(\frac{\|f\|_2^2}{\|f\|_1}\right)^{\frac{1}{p}} = N^{\frac{1}{p}-1}\|f\|_1\left(\frac{\|f\|_2}{\|f\|_1}\right)^{\frac{2}{p}}.$$

Again, combining with our previous bounds,

$$\eta\|f\|_2\frac{\sqrt{|\Lambda|}}{N} \leqslant C\sqrt{p}N^{\frac{1}{p}-1}\|f\|_1\left(\frac{\|f\|_2}{\|f\|_1}\right)^{\frac{2}{p}},$$

i.e.,

$$\sqrt{|\Lambda|} \leqslant C\eta^{-1}\sqrt{p}\left(\frac{\|f\|_2^2}{\|f\|_1^2}N\right)^{\frac{1}{p}}\frac{\|f\|_1}{\|f\|_2}.$$

Now, taking

$$p = \log\left(\frac{\|f\|_2^2}{\|f\|_1^2}N\right),$$

we obtain

$$\sqrt{|\Lambda|} \leqslant C\eta^{-1}\sqrt{\log\left(\frac{\|f\|_2^2}{\|f\|_1^2}N\right)}\frac{\|f\|_1}{\|f\|_2}.$$

Thus,

$$|\Lambda| \ll \eta^{-2}\log\left(\frac{\|f\|_2^2}{\|f\|_1^2}N\right)\frac{\|f\|_1}{\|f\|_2},$$

and this proves (1.23). $\square$

References

[1] A.S. Bandeira, M.E.Lewis, and D.G.Mixon, *Discrete uncertainty principles and sparse signal processing*, arXiv:1504.01014, (2015). 5

[2] J. Bourgain, *Bounded orthogonal systems and the $\Lambda(p)$-set problem*, Acta Math. **162** (1989), no. 3-4, 227–245. 6

[3] W. Burstein, A. Iosevich, A. Mayeli, and H. Nathan, *Fourier minimization and time series imputation*, (arXiv:2506.19226), (2025). 5, 6

[4] L.I. Bluestein, *A linear filtering approach to the computation of the discrete Fourier transform*, IEEE Transactions on Audio and Electroacoustics, vol. AU-18, no.4, pp.451-455, (1970). 9

[5] S. Boucheron, G. Lugosi, P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, 2013. Theorem 5.6 (Gaussian concentration for Lipschitz functions) provides Version A; see also Chapter 2 for χ^2 concentration results. 16, 40, 41

[6] E. Candès and T. Tao, *Decoding by linear programming. IEEE Transactions on Information Theory*, 51(12), (2005), 4203–4215. 31

[7] E. Candès, *The Restricted Isometry Property and its implications for compressed sensing*, Comptes Rendus Mathematique, 346(9–10), (2008), 589–592. 31

[8] E.J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics, vol.9, no.6, pp.717-772, (2009). doi:10.1007/s10208-009-9045-5, (arXiv:0805.4471), (2008). 10

[9] E. Candès, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Communications on Pure and Applied Mathematics 59 (2005). 32

[10] J. W. Cooley and J. W. Tukey, *An Algorithm for the Machine Calculation of Complex Fourier Series*, Mathematics of Computation, vol.19, no.90, pp. 297-301, (1965). 9

[11] M. Chang, (2002). *A polynomial bound in Freiman's theorem*, Duke Mathematical Journal. 113. 10.1215/S0012-7094-02-11331-3. 18, 43

[12] D. Donoho and P. Stark, *Uncertainty principle and signal processing*, SIAM Journal of Applied Math., (1989), Society for Industrial and Applied Mathematics, volume 49, No. 3, pp. 906-931. 1, 9

[13] I.J. Good, *The interaction algorithm and practical Fourier analysis*, Journal of the Royal Statistical Society, Series B, vol.20, no.2, pp.361-372, (1958). 9

[14] B. Green, *Arithmetic progressions in sumsets*. GAFA, Geom. funct. anal. 12, 584–597 (2002). https://doi.org/10.1007/s00039-002-8258-4 18

[15] W. Hagerstrom, *A number of perspectives on signal recovery*, University of Rochester Honors Thesis (2025). 6

[16] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. Wiley, 1996. (Section on averaging/ensemble averaging and noise power reduction.) 16

[17] I. Haviv and O. Regev, *The Restricted Isometry Property of Subsampled Fourier Matrices*, (arXiv:1507.01768), (2015). 32

[18] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association **58** (301): 13–30. 27

[19] A. Iosevich, B. Kashin, I. Limonova, and A. Mayeli, *Subsystems of orthogonal systems and the recovery of sparse signals in the presence of random losses*, Russian Mathematical Surveys, (2024), Volume 79, Issue 6, Pages 1095-1097. 1

[20] A. Iosevich and A. Mayeli, *Uncertainty Principles, Restriction, Bourgain's Λ_q theorem, and Signal Recovery*, Applied and Computational Harmonic Analysis 76 (2025): 101734. 1

[21] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993, (Chapter 2–3: unbiasedness and variance of sample means; averaging i.i.d. noise reduces variance by $1/n$.) 16

[22] B. Laurent and P. Massart. *Adaptive estimation of a quadratic functional by model selection*, Annals of Statistics, 28(5):1302–1338, 2000. Lemma 1 states the χ^2 tail bounds used as Version B. 16, 41, 42

[23] M. Ledoux, *The Concentration of Measure Phenomenon*, American Mathematical Society, 2001. 40, 41

[24] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. Springer, (1998). (Sections on unbiased estimators and variances; sample mean properties.) 16

[25] M. Li and P. Vitanyi, *An Introduction to Kolmogorov Complexity and Its Applications*, 4th Edition, Springer, (2019). 11

[26] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. Prentice Hall, 2009. (Discussions of ensemble/time averaging as noise reduction; reduction of noise power with averaging.) 16

[27] C.M. Rader, *Discrete Fourier transforms when the number of data samples is prime*, Proceedings of the IEEE, vol.56, no.6, pp.1107-1108, (1968). 9

[28] R. Reyzin, *Statistical Queries and Statistical Algorithms: Foundations and Applications*, (arXiv:2004.00557), (2020). 14

[29] Rudelson, M. & Vershynin, R. Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements. (2006), https://arxiv.org/abs/math/0602559 31

[30] stemforall2025, *Undergraduate research program at the University of Rochester*, https://alexiosevich.com/stemforall2025.html, (2025). 3

[31] M. Talagrand, *Selecting a proportion of characters*, Israel J. Math. **108** (1998), 173-191. 3, 5, 6

[32] T. C. Tao and V. H. Vu, *Additive combinatorics*, paperback edition, Cambridge Studies in Advanced Mathematics, 105, Cambridge Univ. Press, Cambridge, 2010; MR2573797. 44

[33] N. Vereshchagin and P. Vitányi. *Rate–Distortion Theory for Kolmogorov Complexity*, IEEE Transactions on Information Theory, 56(7):3438–3454, (2010). 12

[34] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge University Press, 2018. 16, 40, 41

[35] M. J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019. Proposition 2.1 reproduces the Laurent–Massart χ^2 deviation inequality used for Version B. 16, 41, 42

[36] L. Wasserman, *All of Statistics*, Springer, 2004. (Chapters 1–3: properties of the sample mean; variance σ^2/n.) 16

Department of Mathematics, University of Rochester, Rochester, NY, USA

Email address: iosevich@gmail.com

Department of Mathematics, CUNY Graduate Center, New York, NY, USA

Email address: amayeli@gc.cuny.edu