Sparsity, Signal Recovery, and Rényi Entropy

Nathaniel Shaffer

May 2026

Abstract

In the field of compressed sensing, a central idea is that of *sparsity*. In this paper, we analyze notions of sparsity for $f : \mathbb{Z}_N^d \to \mathbb{C}$ by examining the Rényi entropy

$$H_\alpha(f) = \frac{1}{1-\alpha} \log \left(\sum_x \frac{|f(x)|^{2\alpha}}{\|f\|_2^{2\alpha}} \right).$$

We prove upper and lower bounds on established measures of approximate sparsity in terms of $H_\alpha(f)$, and give a qualitative description of which signals maximize $H_\alpha(f)$. Lastly, we describe some connections between H_α and the analysis of Boolean functions.

Contents

1 Introduction

The primary focus of the field of compressed sensing is understanding the reconstruction of a signal from a limited number of linear measurements. This problem is under-determined in general, as when the number of measurements is fewer than the length of the signal, there may exist multiple solutions matching those measurements. The key idea of compressed sensing, however, is that in the presence of *sparsity*, one can often achieve exact reconstruction.

We can formalize this problem by considering a signal to be a map $f : \mathbb{Z}_N^d \to \mathbb{C}$, and taking the linear measurement to be the Fourier transform of f, denoted $\widehat{f}$. Alternatively, we can view this as losing some values of $\widehat{f}$, so that one measures $\{\widehat{f}(m)\}_{m \notin S}$ for some subset $S \subseteq \mathbb{Z}_N^d$, and asks when it is possible to recover f entirely from this limited measurement. In this setting, we may view f as sparse provided it has few non-zero values, and quantitatively look at the set

$$E = \operatorname{supp}(f) = \{x \in \mathbb{Z}_N^d \;:\; f(x) \neq 0\}.$$

In their seminal paper [2], Donoho and Stark used the classical Fourier uncertainty principle to show that in the case that

$$|E| \cdot |S| < \frac{N^d}{2},$$

it is in fact possible to recover f exactly from the measurements $\{\widehat{f}(m)\}_{m \notin S}$. In particular, recovery is possible whenever the set of unobserved frequencies satisfies

$$|S| < \frac{N^d}{2|E|}.$$

Thus if f is sparse in the sense that it has few non-zero values, we have $|E| \ll N^d$, and hence we may recover f even when many frequencies are unobserved.

Additionally, we may note that this recovery problem is symmetric in $f, \widehat{f}$, so that we just as well could have taken the measurement $\{f(x)\}_{x \notin M}$ and asked to recover $\widehat{f}$. In this way, we have a good recovery result if either f or $\widehat{f}$ is sparse in the sense of having small support.

In many settings however, defining sparsity to mean having small support is too strong of an assumption, as a signal may be non-zero everywhere and yet be concentrated on some smaller set of values. One way to quantify this is by supposing that for some $g : \mathbb{Z}_N^d \to \mathbb{C}$ with small support, we have

$$\|f - g\|_2 \leq \epsilon \|f\|_2$$

for some small ϵ. In other words, we can obtain a good relative error by approximating f by some sparse function g. Many reconstruction results in compressed sensing which are exact for a sparse function may thus go through at the cost of the error ϵ.

To quantify this notion of sparsity even further, in [1], Aldahleh et al. examine the Fourier Ratio

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

They show that we have $1 \leq \mathrm{FR}(f)^2 \leq |\operatorname{supp}(\widehat{f})|$, and moreover given $\epsilon > 0$ we can approximate f in the above sense by some g with

$$|\operatorname{supp}(\widehat{g})| \leq \frac{\mathrm{FR}(f)^2}{\epsilon^2}.$$

In this way, $\mathrm{FR}(f)$ measures the approximate sparsity of $\widehat{f}$.

The main goal of this work is to generalize the Fourier Ratio using information theory and general Rényi entropy. We thus argue that the quantity

$$H_\alpha(f) = \frac{1}{1-\alpha} \log \left(\sum_x \frac{|f(x)|^{2\alpha}}{\|f\|_2^{2\alpha}} \right)$$

is a good measure of approximate sparsity of $f : \mathbb{Z}_N^d \to \mathbb{C}$.

In Section 2, we introduce the basic Fourier analytic and information theoretic background to define the Rényi entropy of both $f, \widehat{f}$. The main focus of this work is Section 3, in which we relate $H_\alpha(f)$ to established notions of sparsity. Next, in Section 4 we show that H_α obeys an uncertainty principle between f and $\widehat{f}$. In Section 5, we utilize Pinsker's inequality to understand functions which are maximal with respect to H_α. Finally, in Section 6, we describe the connections between Fourier entropy and the analysis of Boolean functions, and summarize some recent results on the Fourier-Entropy-Influence conjecture.

2 Definitions

For functions $f : \mathbb{Z}_N^d \to \mathbb{C}$, we have the ℓ^p norms

$$\|f\|_p = \left(\sum_{x \in \mathbb{Z}_N^d} |f(x)|^p \right)^{\frac{1}{p}}.$$

For the case $p = \infty$, we set

$$\|f\|_\infty = \max_{x \in \mathbb{Z}_N^d} |f(x)|,$$

noting this corresponds to the limit $p \to \infty$. Additionally, as another special case we define

$$\|f\|_0 = |\{x \in \mathbb{Z}_N^d \,:\, f(x) \neq 0\}|,$$

in other words if we define

$$\operatorname{supp}(f) = \{x \in \mathbb{Z}_N \,:\, f(x) \neq 0\},$$

then $\|f\|_0$ is the size of the support of f. As in the interpretation from Section 1, the quantity $\|f\|_0$ is commonly referred to as the "sparsity" of f.

Additionally, for a subset $S \subseteq \mathbb{Z}_N^d$, we define

$$\|f\|_{\ell^p(S)} = \left(\sum_{x \in S} |f(x)|^p \right)^{\frac{1}{p}},$$

in other words $\|f\|_{\ell^p(S)}$ is the ℓ^p norm of the restriction of f to S.

2.1 Fourier Analysis Background

We define the primitive character $\chi : \mathbb{Z}_N \to \mathbb{C}$ by

$$\chi(t) = \exp\left(\frac{2\pi i t}{N}\right),$$

and we have the collection of character functions $\chi_m : \mathbb{Z}_N^d \to \mathbb{C}$ given by

$$\chi_m(x) = \chi(x \cdot m),$$

where $x \cdot m = x_1 m_1 + \cdots + x_d m_d$.

Note that $\{\chi_m\}_{m \in \mathbb{Z}_N^d}$ forms an orthogonal basis for the set of functions $f : \mathbb{Z}_N^d \to \mathbb{C}$, so that since $\|\chi_m\|_2^2 = N^d$, the collection $\{N^{-d/2}\chi_m\}_{m \in \mathbb{Z}_N^d}$ is an orthonormal basis. Thus, we have the expansion

$$f = N^{-d/2} \sum_{m \in \mathbb{Z}_N} \chi_m \left\langle f, N^{-d/2}\chi_m \right\rangle.$$

We then define the Fourier transform of f to be the function $\widehat{f} : \mathbb{Z}_N^d \to \mathbb{C}$ given by

$$\begin{aligned}\widehat{f}(m) &= \left\langle f, N^{-d/2}\chi_m \right\rangle \\ &= N^{-d/2} \sum_{x \in \mathbb{Z}_N^d} f(x)\overline{\chi_m(x)} \\ &= N^{-d/2} \sum_{x \in \mathbb{Z}_N^d} f(x)\chi(-m \cdot x).\end{aligned}$$

From the expansion of f in the characters χ_m, we get the inversion identity

$$f(x) = N^{-d/2} \sum_{m \in \mathbb{Z}_N^d} \widehat{f}(m)\chi(m \cdot x).$$

Note that orthonormality of the character functions immediately gives the Parseval identity $\langle f, g \rangle = \langle \widehat{f}, \widehat{g} \rangle$, so that we have $\|f\|_2 = \|\widehat{f}\|_2$.

2.2 Information Theory Background

For a probability distribution p and $\alpha > 0$ with $\alpha \neq 1$, we first define the α Rényi entropy of p by

$$H_\alpha(p) = \frac{1}{1-\alpha} \log\left(\sum_x p(x)^\alpha\right).$$

For the case $\alpha = 0$ or $\alpha = 1$, we define $H_\alpha(p)$ in the limit, so that $H_0(p) = \log|\operatorname{supp}(p)|$, and by a quick application of L'Hopital's rule we have

$$H_1(p) = \sum_x p(x) \log\left(\frac{1}{p(x)}\right).$$

$H_1(p)$ is also called the Shannon entropy of p, and in the literature is often simply written $H(p)$.

The quantities $H_\alpha(p)$ are nonincreasing in α, in other words for $\alpha \leq \beta$ we have $H_\alpha(p) \geq H_\beta(p)$. As a special case of this, we get that for any α,

$$H_\alpha(p) \leq H_0(p) = \log|\operatorname{supp}(p)|,$$

so that if p is supported in a set of size k, $H_\alpha(p)$ does not exceed $\log k$.

Additionally, for probability distributions p, q and $\alpha > 0$, the α Rényi divergence of p and q is given by

$$D_\alpha(p\|q) = \frac{1}{\alpha - 1} \log\left(\sum_m \frac{p(m)^\alpha}{q(m)^{\alpha-1}}\right).$$

Similarly to before, for $\alpha = 1$, $D_1(p\|q)$ is defined by taking the limit $\alpha \to 1$, to get that

$$D_1(p\|q) = \sum_m p(m) \log \frac{p(m)}{q(m)}.$$

$D_1(p\|q)$ is sometimes called the Kullback-Leibner divergence, often written $D_{KL}(p\|q)$ in the literature, and is related to the Shannon entropy.

2.3 Fourier Entropy

Motivated by the Plancherel identity that $\|f\|_2 = \|\widehat{f}\|_2$, we can set $g(x) = f(x)/\|f\|_2$ to get that $\widehat{g}(m) = \widehat{f}(m)/\|\widehat{f}\|_2$. Thus both $|g(x)|^2$ and $|\widehat{g}(m)|^2$ define probability distributions on $\mathbb{Z}_N$.

Hence, we can consider the α Rényi entropy of both of these distributions, and accordingly we make the following definition.

Definition 2.1. For $f : \mathbb{Z}_N^d \to \mathbb{C}$, we define the probability distribution $p_f(x) = \frac{|f(x)|^2}{\|f\|_2^2}$, and the α Rényi entropy of f by

$$H_\alpha(f) = H_\alpha(p_f).$$

Note that in the case $\alpha = 0$, we see that $H_0(f), H_0(\widehat{f})$ are simply the logarithm of the size of the supports of $f, \widehat{f}$ respectively, while these quantities decrease for increasing α.

Additionally, observe that we can write $H_\alpha(f)$ in terms of the $\ell^{2\alpha}$ norm of f by

$$H_\alpha(f) = \frac{1}{1-\alpha} \log\left(\sum_x \left(\frac{|f(x)|^2}{\|f\|_2^2}\right)^\alpha\right) = \frac{1}{1-\alpha} \log\left(\frac{\|f\|_{2\alpha}^{2\alpha}}{\|f\|_2^{2\alpha}}\right) = \frac{2\alpha}{1-\alpha} \log \frac{\|f\|_{2\alpha}}{\|f\|_2}. \tag{1}$$

In this form we see that $H_\alpha(f)$ is scale invariant, in other words $H_\alpha(cf) = H_\alpha(f)$, and thus in the normalized case $\|f\|_2 = 1$, we note that

$$H_\alpha(f) = \frac{2\alpha}{1-\alpha} \log \|f\|_{2\alpha},$$

in other words

$$e^{(1-\alpha)H_\alpha(f)} = \|f\|_{2\alpha}^{2\alpha}.$$

We interpret $H_\alpha(f)$ as measuring the "sparsity" or "compressibility" of the function f, and the main objective of this work is to quantify ways in which this interpretation holds.

3 Approximate Sparsity and Rényi Entropy

The central theme of compressed sensing is that signal recovery or compressing problems can be simplified or made tractable in the presence of sparsity. However, most signals are not exactly sparse in the sense that they have many entries which are 0, but often signals are "approximately" sparse in the sense that most of their entries are small.

This notion of "approximate" sparsity is often quantified by the ℓ^q error resulting from a sparse approximation. This is quantified in the following definition.

Definition 3.1. For $f : \mathbb{Z}_N^d \to \mathbb{C}$, the ℓ^q error of the best s-term approximation is

$$\sigma_s(f)_q = \inf\{\|f-P\|_q \,:\, \|P\|_0 \le s\}.$$

We will focus on the case $q = 2$, and moreover we are typically concerned with quantifying relative error, and thus will usually examine the quantity $\sigma_s(f/\|f\|_2)_2$. Note that this is equivalent to normalizing f so that $\|f\|_2 = 1$.

Conversely to looking at the best error from a fixed sparse approximation, we can examine what degree of sparsity is required to obtain a specified error. Thus we make the following definition.

Definition 3.2. For $f : \mathbb{Z}_N^d \to \mathbb{C}$, the concentration function of f, $k_\epsilon(f)$, is defined by

$$k_\epsilon(f) = \inf\{\|P\|_0 \,:\, \|f-P\|_2 \le \epsilon\|f\|_2\}.$$

Remark 3.3. Observe that the infimum defining both $\sigma_s(f)_2$ and $k_\epsilon(f)$ is attained by $P = f|_S$ for some suitable S. To see this, for any P supported in S, note that if we set $P' = f|_S$, then $\|P'\|_0 = \|P\|_0 = |S|$, and moreover

$$\|f-P'\|_2^2 = \sum_{x\notin S} |f(x)|^2 \le \sum_{x\notin S}|f(x)|^2 + \sum_{x\in S}|f(x)-P(x)|^2 = \|f-P\|_2^2.$$

Thus the best error is achieved when $P = f|_S$ for some subset S, and this shows that we can write $k_\epsilon(f)$ in the following way:

$$k_\epsilon(f) = \inf\{|S| \,:\, \|f\|_{\ell^2(S^c)} \le \epsilon\|f\|_2\}.$$

In fact, if we assume that $x_1, \ldots, x_{N^d}$ is an ordering of $\mathbb{Z}_N^d$ satisfying

$$|f(x_1)| \ge |f(x_2)| \ge \cdots \ge |f(x_{N^d})|,$$

then the infimum defining $k_\epsilon(f)$ is attained by $P = f|_S$ for $S = \{x_1, \ldots, x_s\}$, since for any S' of size s, we have

$$\|f - f|_{S'}\|_2^2 = \sum_{x \notin S'} |f(x)|^2 \geq \sum_{j>s} |f(x_j)|^2 = \|f - f|_S\|_2^2.$$

Additionally, this fact implies that more generally,

$$\sigma_s(f)_q^q = \sum_{j>s} |f(x_j)|^q.$$

Remark 3.4. If we assume that $\|f\|_2 = 1$, then we can relate $k_\epsilon(f)$ and $\sigma_s(f)$ in the following way.

$$k_\epsilon(f) \leq s \quad \text{if and only if} \quad \sigma_s(f)_2 \leq \epsilon.$$

In this way, $k_\epsilon(f)$ and $\sigma_s(f)$ are inverse to one another, and bounds on one can be converted to bounds on the other. This is a fact we will use throughout.

3.1 Approximate Sparsity Upper Bounds

Our first goal is to relate the quantities $H_\alpha(f)$ to well-established notions of sparsity, namely $\sigma_s(f)$ above.

In [1], Aldahleh et al. examine the Fourier Ratio $\mathrm{FR}(f) = \frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_2}$, and demonstrate that $\widehat{f}$ is concentrated on some set S with $|S|$ bounded in terms of $\mathrm{FR}(f)$. In the context of Rényi entropy, by (1) the Fourier Ratio can be written

$$\mathrm{FR}(f)^2 = \exp(H_{1/2}(\widehat{f})),$$

so that the Fourier α-Rényi entropy generalizes the Fourier Ratio as α varies.

The statement that $\widehat{f}$ is concentrated on a set S of size s amounts to the fact that $k_\epsilon(\widehat{f}) \leq s$. Thus, we can rephrase Fourier concentration results from [1] in terms of bounds on $k_\epsilon(\widehat{f})$.

Proposition 3.5 (Theorem 1.15, [1]). *If $f : \mathbb{Z}_N^d \to \mathbb{C}$, then for any $\epsilon > 0$ we have*

$$k_\epsilon(\widehat{f}) \leq \frac{\exp(H_{1/2}(\widehat{f}))}{\epsilon^2}.$$

Remark 3.6. Up to Fourier inversion, we can write the bound in Proposition 3.5 in terms of $\sigma_s(f)$ to see that in the case $\|f\|_2 = 1$,

$$\sigma_s(f)_2 \leq \frac{\exp(\frac{1}{2} H_{1/2}(f))}{\sqrt{s}}.$$

Note moreover that under the assumption $\|f\|_2 = 1$, the above bound takes the form

$$\sigma_s(f)_2 \leq \frac{\|f\|_1}{\sqrt{s}}.$$

More generally, $\sigma_s(f)$ can be bounded in terms of the ℓ^p norm of f, as in the following proposition in [3], originally an observation due to Stechkin.

Proposition 3.7 (Proposition 2.3, [3]). *For any* $0 < p < q$ *and* $f : \mathbb{Z}_N^d \to \mathbb{C}$,

$$\sigma_s(f)_q \leq \frac{1}{s^{1/p-1/q}} \|f\|_p.$$

Proof. Let $x_1, \ldots, x_{N^d}$ be ordered so that

$$|f(x_1)| \geq |f(x_2)| \geq \cdots \geq |f(x_{N^d})|.$$

First observe that for any s, we have

$$|f(x_s)| \leq \frac{1}{s} \left(\sum_{i=1}^{s} |f(x_i)|^p \right)^{1/p}$$

since $|f(x_s)| \leq |f(x_i)|$ for all $1 \leq i \leq s$.

Thus, note that by Remark 3.3 and the observation above, we have that

$$\begin{aligned} \sigma_s(f)_q^q &= \sum_{j>s} |f(x_j)|^q \\ &\leq |f(x_s)|^{q-p} \sum_{j>s} |f(x_j)|^p \\ &\leq \left(\frac{1}{s} \sum_{i=1}^{s} |f(x_i)|^p \right)^{\frac{q-p}{p}} \sum_{j>s} |f(x_j)|^p \\ &\leq \left(\frac{1}{s} \|f\|_p^p \right)^{\frac{q-p}{p}} \|f\|_p^p \\ &= \frac{1}{s^{\frac{q}{p}-1}} \|f\|_p^q, \end{aligned}$$

so that taking qth roots,

$$\sigma_s(f)_q \leq \frac{1}{s^{1/p-1/q}} \|f\|_p.$$

∎

Thus, the bound in Proposition 3.5 amounts to the case $\|f\|_2 = 1$, $q = 2$, and $p = 1$ in Proposition 3.7.

In [3], Foucart and Rauhut remark that Proposition 3.7 can be strengthened to obtain a better constant using convex optimization. This gives the following theorem.
Theorem 3.8 (Theorem 2.5, [3]). *For any* $0 < p < q$ *and* $f : \mathbb{Z}_N^d \to \mathbb{C}$,

$$\sigma_s(f)_q \leq \frac{c_{p,q}}{s^{1/p-1/q}} \|f\|_p$$

with

$$c_{p,q} = \left[\left(\frac{p}{q} \right)^{p/q} \left(1 - \frac{p}{q} \right)^{1-p/q} \right]^{1/p} \leq 1.$$

The constant $c_{p,q}$ *is best possible.*

By (1), we can write $H_\alpha(f)$ in terms of the $\ell^{2\alpha}$ norm of f, and thus the above bounds immediately give a relation between $\sigma_s(f)$ and $H_\alpha(f)$. For simplicity we use the bound with constant 1, but note that the following result can be improved by leveraging Theorem 3.8.

Theorem 3.9. *For any* $0 < \alpha < 1$ *and* $f : \mathbb{Z}_N^d \to \mathbb{C}$, *if* $\|f\|_2 = 1$ *then*

$$\sigma_s(f)_2 \le \left(\frac{e^{H_\alpha(f)}}{s}\right)^{\frac{1-\alpha}{2\alpha}}.$$

In other words,

$$k_\epsilon(f) \le \frac{e^{H_\alpha(f)}}{\epsilon^{\frac{2\alpha}{1-\alpha}}}.$$

While the bound of Proposition 3.5 is a specific case of Theorem 3.9, the proof technique used is distinct and can be generalized to give another upper bound on $\sigma_s(f)$ and $k_\epsilon(f)$ in terms of a convex combination of $H_\alpha(f)$ and $H_{1-\alpha}(f)$.

Theorem 3.10. *For any* $\alpha \in (0,1)$ *and* $\epsilon > 0$,

$$k_\epsilon(f) \le \frac{\exp\left((1-\alpha)H_\alpha(f) + \alpha H_{1-\alpha}(f)\right)}{\epsilon^2}.$$

In other words, if $\|f\|_2 = 1$ *then*

$$\sigma_s(f)_2 \le \left(\frac{\exp\left((1-\alpha)H_\alpha(f) + \alpha H_{1-\alpha}(f)\right)}{s}\right)^{1/2}$$

Proof. We prove the equivalent bound

$$k_\epsilon(\widehat{f}) \le \frac{\exp\left((1-\alpha)H_\alpha(\widehat{f}) + \alpha H_{1-\alpha}(\widehat{f})\right)}{\epsilon^2}.$$

The proof proceeds similarly to the methods of [1], so define the random function Z by letting

$$Z(x) = N^{-d/2}\chi(x\cdot m)\widehat{f}(m)\cdot\frac{\|\widehat{f}\|_{2\alpha}^{2\alpha}}{|\widehat{f}(m)|^{2\alpha}} \quad \text{with probability } \frac{|\widehat{f}(m)|^{2\alpha}}{\|\widehat{f}\|_{2\alpha}^{2\alpha}},$$

where m ranges over the Fourier support of f. Observe that $Z(x)$ is a scalar multiple of some character χ_m.

We have by construction that $\mathbb{E}[Z(x)] = f(x)$, and moreover we can compute

$$\begin{aligned}\mathbb{E}[|Z(x)|^2] &= \sum_m \left|N^{-d/2}\chi(x\cdot m)\widehat{f}(m)\cdot\frac{\|\widehat{f}\|_{2\alpha}^{2\alpha}}{|\widehat{f}(m)|^{2\alpha}}\right|^2\cdot\frac{|\widehat{f}(m)|^{2\alpha}}{\|\widehat{f}\|_{2\alpha}^{2\alpha}}\\ &= N^{-d}\|\widehat{f}\|_{2\alpha}^{2\alpha}\sum_m |\widehat{f}(m)|^{2-2\alpha}\\ &= N^{-d}\|\widehat{f}\|_{2\alpha}^{2\alpha}\|\widehat{f}\|_{2(1-\alpha)}^{2(1-\alpha)},\end{aligned}$$

and thus

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

Thus, letting $Z_1, \ldots, Z_k$ be i.i.d. copies of Z and then setting

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

by a similar computation as in [1] we get that

$$\begin{aligned}
\mathbb{E}[\|f - P\|_2^2] &= \sum_x \mathrm{Var}(P(x)) \\
&= \frac{1}{k}\sum_x \mathrm{Var}(Z(x)) \\
&= \frac{1}{k}\left(\|\widehat{f}\|_{2\alpha}^{2\alpha}\|\widehat{f}\|_{2(1-\alpha)}^{2(1-\alpha)} - \|f\|_2^2\right).
\end{aligned}$$

When this quantity is less than $\epsilon^2\|f\|_2^2$, we get the wanted result, and by Plancherel this amounts to the assumption that

$$\begin{aligned}
k &> \frac{1}{\epsilon^2}\left(\frac{\|\widehat{f}\|_{2\alpha}^{2\alpha}\|\widehat{f}\|_{2(1-\alpha)}^{2(1-\alpha)}}{\|\widehat{f}\|_2^2} - 1\right) \\
&= \frac{1}{\epsilon^2}\left(\frac{\|\widehat{f}\|_{2\alpha}^{2\alpha}}{\|\widehat{f}\|_2^{2\alpha}} \cdot \frac{\|\widehat{f}\|_{2(1-\alpha)}^{2(1-\alpha)}}{\|\widehat{f}\|_2^{2(1-\alpha)}} - 1\right) \\
&= \frac{1}{\epsilon^2}\left(e^{(1-\alpha)H_\alpha(\widehat{f})} \cdot e^{\alpha H_{1-\alpha}(\widehat{f})} - 1\right) \\
&= \frac{1}{\epsilon^2}\left(\exp\left((1-\alpha)H_\alpha(\widehat{f}) + \alpha H_{1-\alpha}(\widehat{f})\right) - 1\right).
\end{aligned}$$

Thus, with

$$k \leq \frac{1}{\epsilon^2}\exp\left((1-\alpha)H_\alpha(\widehat{f}) + \alpha H_{1-\alpha}(\widehat{f})\right)$$

there is a deterministic choice of P such that

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

By Plancherel this gives $\|\widehat{f} - \widehat{P}\|_2 \leq \epsilon\|\widehat{f}\|_2$, and thus the desired claim follows from the fact that $\|\widehat{P}\|_0 \leq k$. ∎

Summarizing the above, we have the two bounds

$$k_\epsilon(f) \leq \frac{e^{H_\alpha(f)}}{\epsilon^{2\alpha/(1-\alpha)}}, \qquad k_\epsilon(f) \leq \frac{e^{(1-\alpha)H_\alpha(f)+\alpha H_{1-\alpha}(f)}}{\epsilon^2}$$

whenever $0 < \alpha < 1$.

Note though that neither of the above bounds can be converted to a bound in terms of the Shannon entropy $H_1(f)$, since both Theorem 3.9 and 3.10 degenerate as $\alpha \to 1$. For Theorem 3.9, the exponent of ϵ blows up in the denominator, while in Theorem 3.10 the numerator tends to simply $e^{H_0(f)} = \|f\|_0$.

Thus, we can use a distinct proof to give the following upper bound on $k_\epsilon(f)$ in terms of the Shannon entropy.

Theorem 3.11. *For $f : \mathbb{Z}_N^d \to \mathbb{C}$, we have*

$$k_\epsilon(f) \leq e^{H_1(f)/\epsilon^2},$$

in other words if $\|f\|_2 = 1$ then

$$\sigma_s(f)_2 \leq \left(\frac{H_1(f)}{\log s}\right)^{\frac{1}{2}}.$$

Proof. Let $p = p_f$ be the probability distribution $p(x) = \frac{|f(x)|^2}{\|f\|_2^2}$ defining $H_1(f)$. Note that for some subset S, we have $\|f\|_{\ell^2(S^C)} \leq \epsilon \|f\|_2$ if and only if

$$\sum_{x \notin S} p(x) \leq \epsilon^2.$$

Now, for $\eta > 0$, set

$$S = \{x \,:\, p(x) > e^{-\eta}\}.$$

We have $|S| < e^\eta$, for otherwise we would get

$$1 = \sum_x p(x) \geq \sum_{x \in S} p(x) > |S| \cdot e^{-\eta} \geq 1,$$

a contradiction. Moreover, note that $x \notin S$ if and only if $p(x) \leq e^{-\eta}$, which holds if and only if $\log \frac{1}{p(x)} \geq \eta$. Thus, by Markov's inequality we have

$$H_1(f) = \sum_x p(x) \log \frac{1}{p(x)} \geq \sum_{x \notin S} p(x) \log \frac{1}{p(x)} \geq \eta \sum_{x \notin S} p(x).$$

Rearranging, this gives

$$\sum_{x \notin S} p(x) \leq \frac{1}{\eta} H_1(f).$$

Choosing $\eta = \frac{1}{\epsilon^2} H_1(f)$, we conclude that

$$\sum_{x \notin S} p(x) \leq \epsilon^2,$$

with $|S| < e^{H_1(f)/\epsilon^2}$. By our first observation, this implies that

$$k_\epsilon(f) \leq e^{H_1(f)/\epsilon^2}$$

as wanted. The bound on $\sigma_s(f)_2$ follows from Remark 3.4. ∎

3.2 Approximate Sparsity Lower Bounds

The bounds in the previous section show that $\sigma_s(f)$ and $k_\epsilon(f)$ are controlled from above by Rényi entropies, and so the next question we investigate is whether one can prove corresponding lower bounds on $\sigma_s(f)$ and $k_\epsilon(f)$. This would thus give an equivalence between approximate sparsity as it is usually defined and small Rényi entropy, further justifying the notion that $H_\alpha(f)$ is a good measure of sparsity.

In the following result, we prove a lower bound on $k_\epsilon(f)$ in terms of the Shannon entropy of f, in other words $H_1(f)$. This demonstrates that small concentration implies small entropy, or equivalently that a function cannot be concentrated on a set that is much smaller than its entropy.

Theorem 3.12. *Let $f : \mathbb{Z}_N^d \to \mathbb{C}$ and $0 < \epsilon < \frac{1}{\sqrt{2}}$. Then*

$$k_\epsilon(f) \geq e^{H_1(f)} \exp\left(-\epsilon^2 \log N^d - h(\epsilon^2)\right),$$

where the function h is the binary entropy function

$$h(x) = x \log \frac{1}{x} + (1-x) \log \frac{1}{1-x}.$$

Proof. Choose any S with $|S| = k_\epsilon(f)$ and $\|f\|_{\ell^2(S^c)} \leq \epsilon \|f\|_2$. Then set $p = p_f$, noting that the concentration assumption on f amounts to

$$\sum_{m \notin S} p(m) = \sum_{m \notin S} \frac{|f(m)|^2}{\|f\|_2^2} \leq \epsilon^2.$$

Thus, set $c = \sum_{m \in S} p(m)$. We then have $1 - c \leq \epsilon^2$.

Now, write $H_1(p)$ as

$$\begin{aligned} H_1(p) &= \sum_m p(m) \log \frac{1}{p(m)} \\ &= \sum_{m \in S} p(m) \log \frac{1}{p(m)} + \sum_{m \notin S} p(m) \log \frac{1}{p(m)} \\ &= c \sum_{m \in S} \frac{p(m)}{c} \log \frac{c}{c \cdot p(m)} + (1-c) \sum_{m \notin S} \frac{p(m)}{1-c} \log \frac{1-c}{(1-c)p(m)} \\ &= c \left(\sum_{m \in S} \frac{p(m)}{c} \log \frac{c}{p(m)} \right) + (1-c) \left(\sum_{m \notin S} \frac{p(m)}{1-c} \log \frac{1-c}{p(m)} \right) + \left(c \log \frac{1}{c} + (1-c) \log \frac{1}{1-c} \right) \\ &= c H_1(q_1) + (1-c) H_1(q_2) + h(c), \end{aligned}$$

where $q_1 = \frac{1}{c} p|_S$ and $q_2 = \frac{1}{1-c} p|_{S^c}$. Note that both q_1, q_2 are probability distributions supported in S, S^c respectively, so that

$$H_1(q_1) \leq \log|\operatorname{supp}(q_1)| \leq \log|S|, \qquad H_1(q_2) \leq \log|\operatorname{supp}(q_2)| \leq \log(N^d - |S|) \leq \log N^d.$$

Since $h(x)$ is increasing for $x \leq \frac{1}{2}$ and symmetric about $\frac{1}{2}$, then by the fact that $1 - c \leq \epsilon^2$ we get

$$h(c) = h(1-c) \leq h(\epsilon^2).$$

Thus we can bound $H_1(p)$ above by

$$H_1(p) \leq c \log |S| + (1-c)\log N^d + h(\epsilon^2) \leq \log |S| + \epsilon^2 \log N + h(\epsilon^2).$$

Rearranging, we get

$$|S| \geq \exp\left(H_1(\widehat{f}) - \epsilon^2 \log N^d - h(\epsilon^2)\right)$$

as claimed. ∎

For general Rényi entropies, we can use a similar proof technique to obtain the following result.

Theorem 3.13. *Let* $f : \mathbb{Z}_N \to \mathbb{C}$, $\epsilon > 0$, *and* $0 < \alpha < 1$. *Then we have*

$$k_\epsilon(f) \geq \left(e^{(1-\alpha)H_\alpha(f)} - \epsilon^{2\alpha} N^{d(1-\alpha)}\right)^{\frac{1}{1-\alpha}}.$$

Proof. We proceed similarly to the previous proof, so choose any S with $|S| = k_\epsilon(f)$ and $\|f\|_{\ell^2(S^c)} \leq \epsilon \|f\|_2$. Setting $p = p_f$, as before the concentration assumption on f amounts to

$$\sum_{m \notin S} p(m) \leq \epsilon^2.$$

Again set $c = \sum_{m \in S} p(m)$, so that $1 - c \leq \epsilon^2$ and $c \leq 1$.

We have by definition that

$$\begin{aligned} e^{(1-\alpha)H_\alpha(f)} &= \sum_m p(m)^\alpha \\ &= \sum_{m \in S} p(m)^\alpha + \sum_{m \notin S} p(m)^\alpha \\ &= c^\alpha \sum_{m \notin S} \left(\frac{p(m)}{c}\right)^\alpha + (1-c)^\alpha \sum_{m \notin S} \left(\frac{p(m)}{1-c}\right)^\alpha \\ &= c^\alpha e^{(1-\alpha)H_\alpha(q_1)} + (1-c)^\alpha e^{(1-\alpha)H_\alpha(q_2)}, \end{aligned}$$

where as in the previous proof we have set $q_1 = \frac{1}{c} p|_S$ and $q_2 = \frac{1}{1-c} p|_{S^c}$. Since q_1, q_2 are probability distributions supported in S, S^c respectively we get

$$H_\alpha(q_1) \leq \log |S|, \qquad H_\alpha(q_2) \leq \log(N^d - |S|) \leq \log N^d.$$

Thus, using $c \leq 1$ and $1 - c \leq \epsilon^2$ we can bound $e^{(1-\alpha)H_\alpha(f)}$ by

$$e^{(1-\alpha)H_\alpha(f)} \leq e^{(1-\alpha)\log |S|} + \epsilon^{2\alpha} e^{(1-\alpha)\log N^d} = |S|^{1-\alpha} + \epsilon^{2\alpha} N^{d(1-\alpha)}.$$

Rearranging and using the fact that $|S| = k_\epsilon(f)$, we conclude that

$$k_\epsilon(f) \geq \left(e^{(1-\alpha)H_\alpha(f)} - \epsilon^{2\alpha} N^{d(1-\alpha)}\right)^{\frac{1}{1-\alpha}}$$

as wanted. ∎

3.3 Explicit Example

To further examine the above bounds, we compute $H_\alpha(f), \sigma_s(f), k_\epsilon(f)$ for the specific function $f : \mathbb{Z}_N \to \mathbb{C}$ given by

$$f(x) = \frac{1}{1+x}.$$

In the following, for quantities A_N, B_N depending on N, we write $A_N \approx B_N$ to mean there are absolute constants c, C such that

$$cB_N \leq A_N \leq CB_N.$$

We first note that

$$\|f\|_2^2 = \sum_{x=0}^{N-1} \frac{1}{(1+x)^2} = \sum_{x=1}^{N} \frac{1}{x^2} \approx \frac{\pi^2}{6},$$

so that namely $\|f\|_2 \approx 1$. Moreover, for $0 < \alpha < 1$,

$$\begin{aligned} \|f\|_{2\alpha}^{2\alpha} &= \sum_{x=0}^{N-1} \frac{1}{(1+x)^{2\alpha}} \\ &= \sum_{x=1}^{N} \frac{1}{x^{2\alpha}} \\ &\approx \int_1^N \frac{dx}{x^{2\alpha}} \\ &= \begin{cases} \frac{1}{1-2\alpha}\left(\frac{1}{N^{2\alpha-1}} - 1\right), & \alpha \neq \frac{1}{2}, \\ \log N, & \alpha = \frac{1}{2}. \end{cases} \end{aligned}$$

In particular, we get

$$\|f\|_{2\alpha}^{2\alpha} \approx \begin{cases} \frac{N^{1-2\alpha}}{1-2\alpha}, & 0 < \alpha < \frac{1}{2} \\ \log N, & \alpha = \frac{1}{2} \\ \frac{1}{2\alpha-1}, & \frac{1}{2} < \alpha < 1. \end{cases}$$

Consequently, by (1), we get

$$e^{H_\alpha(f)} \approx \begin{cases} N^{\frac{1-2\alpha}{1-\alpha}}, & 0 < \alpha < \frac{1}{2} \\ (\log N)^2, & \alpha = \frac{1}{2} \\ 1, & \frac{1}{2} < \alpha < 1, \end{cases}$$

provided that α is fixed and we are concerned with N large.

On the other hand, by Remark 3.3,

$$\sigma_s(f)_2^2 = \sum_{x=s}^{N} \frac{1}{x^2} \approx \int_s^N \frac{dx}{x^2} = \frac{1}{s} - \frac{1}{N},$$

so that for large N, $\sigma_s(f)_2 \approx \frac{1}{\sqrt{s}}$. In terms of $k_\epsilon(f)$, we get that

$$k_\epsilon(f) \approx \frac{1}{\epsilon^2 - \frac{1}{N}},$$

so that $k_\epsilon(f) \approx \frac{1}{\epsilon^2}$.

Now, our upper bound from Theorem 3.9 gives

$$k_\epsilon(f) \leq \frac{e^{H_\alpha(f)}}{\epsilon^{2\alpha/(1-\alpha)}} \approx \begin{cases} \frac{1}{\epsilon^{2\alpha/(1-\alpha)}} N^{(1-2\alpha)/(1-\alpha)}, & 0 < \alpha < \frac{1}{2} \\ \frac{1}{\epsilon^2}(\log N)^2, & \alpha = \frac{1}{2} \\ \frac{1}{\epsilon^{2\alpha/(1-\alpha)}}, & \frac{1}{2} < \alpha < 1. \end{cases}$$

In this case, we see that the bound using $H_{1/2}(f)$ is sharp in the dependence on ϵ, but the dependence in N is off by a polylogarithmic factor. On the other hand, for $\frac{1}{2} < \alpha < 1$, we get the correct constant dependence on N, but the dependence on ϵ has the exponent

$$\frac{2\alpha}{1-\alpha} > 2.$$

This suggests a trade-off between $H_\alpha(f)$ for varying α, giving either better decay in ϵ or with respect to N.

4 Entropic Uncertainty Principles

A classical idea in Fourier analysis is that a function and its Fourier transform cannot simultaneously be localized. In terms of support, this takes the form of the standard Fourier uncertainty principle, which can be stated in the following way.

Proposition 4.1. *If $f : \mathbb{Z}_N^d \to \mathbb{C}$ is supported in $E \subseteq \mathbb{Z}_N^d$ and $\widehat{f}$ is supported in $S \subseteq \mathbb{Z}_N^d$, then we have*

$$|E| \cdot |S| \geq N^d.$$

To justify the interpretation of $H_\alpha(f)$ as the sparsity of f, we should thus have an uncertainty principle in terms of $H_\alpha(f), H_\alpha(\widehat{f})$, and indeed we have the following result.

Theorem 4.2. *Let $\frac{1}{2} \leq \alpha \leq 1 \leq \beta$ with $\frac{1}{\alpha} + \frac{1}{\beta} = 2$. Then*

$$H_\beta(\widehat{f}) + H_\alpha(f) \geq \log N^d.$$

Proof. First, since Rényi entropy is scale invariant, we normalize so that $\|f\|_2 = \|\widehat{f}\|_2 = 1$. We thus have

$$H_\alpha(f) = \frac{1}{1-\alpha} \log \left(\|f\|_{2\alpha}^{2\alpha} \right).$$

Additionally, without loss of generality suppose that $\alpha < 1 < \beta$, since the case $\alpha = \beta = 1$ follows by taking the limit $\alpha \to 1$.

Next, recall the standard Hausdorff-Young inequality, which states that for $1 \leq p \leq 2 \leq q$ with $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\left(\sum_m |\widehat{f}(m)|^q \right)^{\frac{1}{q}} \leq N^{-\frac{d}{p} + \frac{d}{2}} \left(\sum_x |f(x)|^p \right)^{\frac{1}{p}}.$$

Setting $p = 2\alpha$, $q = 2\beta$, our assumption on α, β amounts to $1 \leq p < 2 < q$ and $\frac{1}{p} + \frac{1}{q} = 1$, so that by Hausdorff-Young we have

$$\left(\sum_m |\widehat{f}(m)|^{2\beta} \right)^{\frac{1}{2\beta}} \leq N^{-\frac{d}{2\alpha} + \frac{d}{2}} \left(\sum_x |f(x)|^{2\alpha} \right)^{\frac{1}{2\alpha}}.$$

Taking the logarithm of both sides gives

$$\begin{aligned} \frac{1}{2\beta} \log(\|\widehat{f}\|_{2\beta}^{2\beta}) &\leq \frac{\alpha - 1}{2\alpha} \log N^d + \frac{1}{2\alpha} \log\left(\|f\|_{2\alpha}^{2\alpha}\right) \\ \frac{1-\beta}{2\beta} H_\beta(\widehat{f}) &\leq \frac{\alpha - 1}{2\alpha} \log N^d + \frac{1-\alpha}{2\alpha} H_\alpha(f) \\ \frac{1-\beta}{2\beta} H_\beta(\widehat{f}) + \frac{\alpha-1}{2\alpha} H_\alpha(f) &\leq \frac{\alpha - 1}{2\alpha} \log N^d. \end{aligned}$$

Now, note that since $\alpha < 1$, we have $\frac{\alpha-1}{2\alpha} < 0$, and moreover since $\frac{1}{\alpha} + \frac{1}{\beta} = 2$ we have

$$\frac{\alpha-1}{2\alpha} = \frac{1}{2}\left(1 - \frac{1}{\alpha}\right) = \frac{1}{2}\left(1 - 2 + \frac{1}{\beta}\right) = \frac{1}{2}\left(\frac{1}{\beta} - 1\right) = \frac{1-\beta}{2\beta},$$

so we can multiply through by $\frac{2\alpha}{\alpha-1} = \frac{2\beta}{1-\beta} < 0$ to get that

$$H_\beta(\widehat{f}) + H_\alpha(f) \geq \log N^d$$

as desired. ∎

Taking the particular case $\alpha = \beta = 1$ above, we get

$$H_1(\widehat{f}) + H_1(f) \geq \log N^d.$$

Note though that this is this is the only case in which Theorem 4.2 gives a bound with $\alpha = \beta$. For the case $\alpha = \beta = \frac{1}{2}$, we have the following theorem.

Theorem 4.3. *For* $f : \mathbb{Z}_N^d \to \mathbb{C}$, *we have*

$$H_{1/2}(\widehat{f}) + H_{1/2}(f) \geq \log N^d.$$

In the notation from [1], this is

$$FR(f)^2 \cdot FR(\widehat{f})^2 \geq N^d.$$

Proof. Recall we can write

$$e^{H_{1/2}(f)} = \frac{\|f\|_1^2}{\|f\|_2^2},$$

so that the above equality is equivalent to

$$\frac{\|\widehat{f}\|_1^2}{\|\widehat{f}\|_2^2} \cdot \frac{\|f\|_1^2}{\|f\|_2^2} \geq N^d.$$

Now, first note that for any x,

$$|f(x)| = \left| N^{-d/2} \sum_m \widehat{f}(m)\chi(m \cdot x) \right| \leq N^{-d/2} \sum_m |\widehat{f}(m)| = N^{-d/2}\|\widehat{f}\|_1.$$

Thus, this implies that

$$\begin{aligned} \|f\|_2^2 &= \sum_x |f(x)|^2 \\ &= \sum_x |f(x)| \cdot |f(x)| \\ &\leq \sum_x |f(x)| N^{-d/2} \|\widehat{f}\|_1 \\ &= N^{-d/2} \|f\|_1 \|\widehat{f}\|_1, \end{aligned}$$

so that

$$N^{d/2} \geq \frac{\|\widehat{f}\|_1 \|f\|_1}{\|f\|_2^2},$$

and by Plancherel we are done. ∎

5 Pinsker's Inequality and Maximal Entropy

Next, we investigate the question of which functions maximize Rényi entropy, and more specifically what we can say about functions with near-maximal entropy. To do so, we first introduce some more information theoretic background.

A fundamental inequality in information theory is Pinkser's inequality, which relates total variation distance to Kullback–Leibler divergence, or more generally Rényi divergence.

Theorem 5.1 (Pinsker's inequality)**.** *If p, q are probability distributions, then*

$$d_{TV}(p, q) \leq \sqrt{\frac{1}{2} D_1(p\|q)},$$

where

$$d_{TV}(p, q) = \sup_A |p(A) - q(A)| = \frac{1}{2}\|p - q\|_1.$$

In terms of the ℓ^1 norms, we can restate Pinsker's inequality as

$$\frac{1}{2}\|p-q\|_1^2 \leq D_q(p\|q).$$

Pinsker's inequality has been improved to more general Rényi divergences, as in the following result from Gildaroni.

Theorem 5.2 (Corollary 6, [5])**.** *If p, q are probability distributions, then for $\alpha \leq 1$*

$$\frac{\alpha}{2}\|p-q\|_1^2 \leq D_\alpha(p\|q).$$

In [5], Gilardoni explores higher order inequalities relating total variation distance and various divergences. In particular if $V = \|p-q\|_1$ and $D = D_\alpha(p\|q)$, then the above states that $D \geq \frac{\alpha}{2}V^2$ over all probability distributions p, and Gildaroni shows that in fact we have

$$D \geq \frac{\alpha}{2}V^2 + \frac{\alpha(1+5\alpha-5\alpha^2)}{36}V^4.$$

For the special case of Kullback-Leibner divergence, i.e. $\alpha = 1$, inequalities bounding D below by even higher order polynomials in V^2 are known, but for the purposes here we apply just the second-order result above.

In order to apply Pinsker's inequality to the Fourier analytic setting, we first prove the following lemma, which shows that when q is uniform on the support of p_f, then $D_\alpha(p\|q)$ measures how close $H_\alpha(f)$ is to maximal.

Lemma 5.3. *Let p be a probability distribution supported in the set S, and let $q = \frac{1}{|S|}1_S$. Then*

$$\begin{aligned} D_\alpha(p\|q) &= \log|S| - H_\alpha(p) \\ &= H_0(p) - H_\alpha(p). \end{aligned}$$

Proof. Setting $q = \frac{1}{|S|}1_S$, we use the fact that p is supported in S to expand

$$\begin{aligned} D_\alpha(p\|q) &= \frac{1}{\alpha-1}\log\left(\sum_m \frac{p(m)^\alpha}{(\frac{1}{|S|}1_S(m))^{\alpha-1}}\right) \\ &= \frac{1}{\alpha-1}\log\left(\sum_{m\in S} p(m)^\alpha |S|^{\alpha-1}\right) \\ &= \log|S| - \frac{1}{1-\alpha}\log\left(\sum_m p(m)^\alpha\right) \\ &= \log|S| - H_\alpha(p). \end{aligned}$$

∎

Applying the generalized form of Pinsker's inequality to p_f, we get the following statement, which relates the maximality of $H_\alpha(f)$, to how flat f is on its support.

Corollary 5.4. *Let $f : \mathbb{Z}_N^d \to \mathbb{C}$, $0 \le \alpha \le 1$. Then if f is supported in S,*

$$\frac{\alpha}{2}\left\| \frac{|f|^2}{\|f\|_2^2} - \frac{1_S}{|S|} \right\|_1^2 \le H_0(f) - H_\alpha(f).$$

To further understand the implications of this corollary, suppose that $H_\alpha(f) = H_0(f)$. Then the right-hand side above is 0, so that we have $\frac{1}{\|f\|_2^2}|f|^2 = \frac{1}{|S|}1_S$. This implies $|f|$ is flat along its support. In particular, the functions f with $H_\alpha(f)$ maximal are exactly those which are flat along their support.

Moreover, this corollary implies that if $H_\alpha(f)$ is close to maximal, then $|f|$ must be close to flat along its support. In this way, we have a qualitative description of the functions maximizing $H_\alpha(f)$.

6 Boolean Fourier Entropy

One of the main contexts in which entropy has appeared in the study of discrete Fourier analysis is in the analysis of Boolean functions. In this context, different normalizations are typical, so we first review the definitions in the Boolean setting.

6.1 Fourier Analysis in the Boolean Setting

One of the most basic objects of study in combinatorics, theoretical computer science, social choice theory, and more are functions of the form $f : \mathbb{F}_2^n \to \{-1, 1\}$. In this case, one commonly identifies $\mathbb{F}_2$ with $\{-1, 1\}$ via $a \mapsto (-1)^a$, so we can view these functions as having co-domain either $\mathbb{F}_2$ or $\{-1, 1\}$.

We now endow $\mathbb{F}_2^n$ with the probability measure, and to distinguish this from the previous setting we have the following L^p norms and L^2 inner product for $f, g : \mathbb{F}_2^n \to \mathbb{R}$:

$$\|f\|_{L^p} = \left(\frac{1}{2^n}\sum_{x\in\mathbb{F}_2^n}|f(x)|^p\right)^{\frac{1}{p}}, \qquad \langle f, g\rangle_{L^2} = \frac{1}{2^n}\sum_{x\in\mathbb{F}_2^n} f(x)g(x).$$

For the case of the un-normalized norms, we define the ℓ^p norms and ℓ^2 inner product by

$$\|f\|_{\ell^p} = \left(\sum_{x\in\mathbb{F}_2^n}|f(x)|^p\right)^{\frac{1}{p}}, \qquad \langle f, g\rangle_{\ell^2} = \sum_{x\in\mathbb{F}_2^n} f(x)g(x).$$

As before, for $\gamma \in \mathbb{F}_2^n$ we still have the character function $\chi_\gamma : \mathbb{F}_2^n \to \mathbb{C}$, which in this case has co-domain $\{-1, 1\}$ and can be written

$$\chi_\gamma(x) = \chi(x \cdot \gamma) = e^{\pi i(x\cdot\gamma)} = (-1)^{x\cdot\gamma}.$$

Additionally, we can identify $\gamma \in \mathbb{F}_2^n$ with $S \subseteq [n] = \{1, 2, \ldots, n\}$ by taking $i \in S$ if and only if $\gamma_i = 1$. In this case, we have the characters

$$\chi_S(x) = (-1)^{\sum_{i \in S} x_i}.$$

As before, the collection $\{\chi_S\}_{S \subseteq [n]}$ are orthogonal, and in fact they are orthonormal with respect to the normalized L^2 inner product above. Thus, we have the expansion

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x)$$

for coefficients

$$\widehat{f}(S) = \langle f, \chi_S \rangle_{L^2} = \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} f(x)\chi_S(x).$$

Again under the identification between $\mathbb{F}_2^n$ and subsets $S \subseteq [n]$, we can view $\widehat{f}$ as a map $\widehat{f} : \mathbb{F}_2^n \to \mathbb{R}$ or $\widehat{f} : 2^{[n]} \to \mathbb{R}$.

Lastly, since the collection $\{\chi_S\}_{S \subseteq [n]}$ are orthonormal in L^2, we have the Parseval identity

$$\langle f, g \rangle_{L^2} = \langle \widehat{f}, \widehat{g} \rangle_{\ell^2},$$

so that in particular

$$\|f\|_{L^2} = \|\widehat{f}\|_{\ell^2}.$$

Note that unlike in the setting on $\mathbb{Z}_N^d$, we must take different norms on the space and Fourier side.

6.2 Influence and Fourier Entropy

Within Boolean analysis, one of the central objects of study is *influence*. For $f : \mathbb{F}_2^n \to \{-1, 1\}$ and $i \in [n]$, the influence of coordinate i is given by

$$I_i[f] = \Pr_{x \in \mathbb{F}_2^n}\left[f(x) \neq f(x^{\oplus i})\right] = \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} \left(\frac{f(x) - f(x^{\oplus i})}{2}\right)^2,$$

where $x^{\oplus i} = (x_1, \ldots, x_{i-1}, 1 - x_i, x_{i+1}, \ldots, x_n)$ is the vector obtained by flipping the ith entry of x. Thus, $I_i[f]$ measures the probability that the ith coordinate will affect the value of f. The total influence of f is given by

$$I[f] = \sum_{i=1}^{n} I_i[f].$$

One can show that $I[f]$ has the following representation in terms of the Fourier expansion of f.

Proposition 6.1 (Theorem 2.38, [8]). *For $f : \mathbb{F}_2^n \to \{-1, 1\}$, we have*

$$I[f] = \sum_{S \subseteq [n]} |\widehat{f}(S)|^2 |S|.$$

In this way, we see that $I[f]$ measures the average "height" of the Fourier coefficients, since we weight by the size $|S|$. This motivates another notion of sparsity, in which $\widehat{f}$ may be considered sparse if it is concentrated on low-degree frequencies. We thus say the Fourier spectrum of $f : \mathbb{F}_2^n \to \{-1, 1\}$ is ϵ-concentrated up to degree m if

$$\sum_{\substack{S \subseteq [n] \\ |S| > m}} \widehat{f}(S)^2 \leq \epsilon.$$

By using a Markov's inequality bound similar to Theorem 3.11, one gets the following proposition which shows that bounded total influence implies concentration up to some small degree.

Proposition 6.2 (Proposition 3.2, [8])**.** *For any $f : \mathbb{F}_2^n \to \{-1, 1\}$ and $\epsilon > 0$, the Fourier spectrum of f is ϵ-concentrated up to degree $I[f]/\epsilon$.*

Now, when considering the Rényi entropy of $\widehat{f}$, whenever $f : \mathbb{F}_2^n \to \{-1, 1\}$ we are guaranteed that

$$\|\widehat{f}\|_{\ell^2} = \|f\|_{L^2} = \left(\frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} f(x)^2 \right)^{\frac{1}{2}} = 1,$$

so that $\widehat{f}(S)^2$ is always a probability distribution. Thus, we can write the Shannon entropy of $\widehat{f}(S)^2$ as

$$H_1(\widehat{f}) = \sum_{S \subseteq [n]} \widehat{f}(S) \log \frac{1}{\widehat{f}(S)^2}.$$

As before, we can interpret this quantity as measuring the approximate sparsity of $\widehat{f}$.

In [4], Friedgut and Kalai posed the Fourier-Entropy-Influence (FEI) conjecture, which can be stated as follows.

Conjecture 6.3 (Fourier-Entropy-Influence (FEI))**.** *There exists some $C > 0$ such that for all $f : \mathbb{F}_2^n \to \{-1, 1\}$,*

$$\sum_{S \subseteq [n]} \widehat{f}(S) \log \frac{1}{\widehat{f}(S)^2} \leq C \sum_{S \subseteq [n]} \widehat{f}(S)^2 |S|. \tag{2}$$

Briefly, this may be stated as $H_1(\widehat{f}) \leq CI[f]$. Friedgut and Kalai initially posed this problem in 1996 while looking at thresholds for monotone graph properties, but the conjecture has come to be a longstanding open-problem about the Fourier structure of Boolean functions. Despite various attempts, the conjecture has resisted proof since it was originally posed in 1996.

Since we have the decreasing family of Rényi entropies $H_\alpha(\widehat{f})$ for $\alpha > 0$, one may ask the weaker question if $H_\alpha(\widehat{f}) \leq CI[f]$ for some $\alpha \geq 1$. In fact, we can take the limit $\alpha \to \infty$ to see that

$$\lim_{\alpha \to \infty} H_\alpha(\widehat{f}) = \min_{S \subseteq [n]} \log \frac{1}{\widehat{f}(S)},$$

where the minimum is taken over $\widehat{f}(S) \neq 0$. We define this quantity to be $H_\infty(\widehat{f})$, the min-entropy of $\widehat{f}$. Thus

$$H_\infty(\widehat{f}) = \min_{S \subseteq [n]} \log \frac{1}{\widehat{f}(S)^2}.$$

Hence, one can ask the weakened version of the FEI conjecture.

Conjecture 6.4 (Fourier-Min-Entropy-Influence (FMEI))**.** *There exists some $C > 0$ such that for all $f : \mathbb{F}_2^n \to \{-1, 1\}$,*

$$\min_{S \subseteq [n]} \log \frac{1}{\widehat{f}(S)^2} \leq C \sum_{S \subseteq [n]} \widehat{f}(S)|S|.$$

In terms of progress towards proving the FEI or FMEI conjecture, in 2019 the authors of [7] show that for $f : \mathbb{F}_2^n \to \{-1, 1\}$ with constant variance, we have

$$H_\infty(\widehat{f}) \leq C I[f] \log I[f].$$

This gives the bound in the FMEI conjecture up to the term $\log I[f]$. More recently, in 2025 Han ([6]) showed that

$$H_1(\widehat{f}) \leq C_1 I[f] + C_2 \sum_{i=1}^{n} I_i[f] \log \frac{1}{I_i[f]}.$$

Additionally, partial progress towards the FEI conjecture has been made over restricted subclasses of functions. Namely, in [9], O'Donnell, Wright, and Zhou show that (2) holds for symmetric functions with $C = 12.04$, where a function is symmetric if it is invariant under any permutation of its arguments. More generally, the authors show that (2) holds for functions which are symmetric under the action of subgroups of the form $S_{n_1} \times \cdots \times S_{n_d}$ for fixed d.

Aside from the above results, there has been limited progress towards resolving either the FEI or FMEI conjecture, and progress in either case would give considerable insight about the Fourier structure of Boolean functions.

7 Acknowledgments

I would first like to thank my advisor, Alex Iosevich, for his advice and support throughout my work on my thesis and throughout my entire undergraduate career, as well as for first introducing me to discrete harmonic analysis. I would like to thank Lillian Stolberg for her constant support ever since the days of TAing Math 171, and for pushing me to be the best mathematician and person I can be. Lastly, I am extremely thankful for the support and company of my family and my friends Karam Aldahleh, Sasha Lifchez, and Nathan Skerrett.

References

[1] K. Aldaleh et al. *The Fourier Ratio and complexity of signals.* 2025. arXiv: `2511.19560 [math.CA]`. URL: `https://arxiv.org/abs/2511.19560`.

[2] David L. Donoho and Philip B. Stark. "Uncertainty Principles and Signal Recovery". In: *SIAM Journal on Applied Mathematics* 49.3 (1989), pp. 906–931. ISSN: 00361399. URL: `http://www.jstor.org/stable/2101993` (visited on 04/06/2026).

[3] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing.* Birkhäuser Basel, 2013. ISBN: 0817649476.

[4] Ehud Friedgut and Gil Kalai. "Every Monotone Graph Property has a Sharp Threshold". In: *Proceedings of the American Mathematical Society* 124.10 (1996), pp. 2993–3002. ISSN: 00029939, 10886826. URL: `http://www.jstor.org/stable/2161839` (visited on 04/07/2026).

[5] Gustavo L. Gilardoni. "On Pinsker's and Vajda's Type Inequalities for Csiszár's f-Divergences". In: *IEEE Transactions on Information Theory* 56.11 (Nov. 2010), pp. 5377–5386. ISSN: 0018-9448. DOI: `10.1109/tit.2010.2068710`. URL: `http://dx.doi.org/10.1109/TIT.2010.2068710`.

[6] Xiao Han. "A new bound for the Fourier-entropy-influence conjecture". en. In: *Combinatorica* 45.1 (Feb. 2025).

[7] Esty Kelman et al. *Towards a Proof of the Fourier–Entropy Conjecture?* 2020. arXiv: `1911.10579 [cs.DM]`. URL: `https://arxiv.org/abs/1911.10579`.

[8] Ryan O'Donnell. "Analysis of Boolean Functions". In: *CoRR* abs/2105.10386 (2021). arXiv: `2105.10386`. URL: `https://arxiv.org/abs/2105.10386`.

[9] Ryan O'Donnell, John Wright, and Yuan Zhou. "The Fourier Entropy–Influence Conjecture for Certain Classes of Boolean Functions". In: *Automata, Languages and Programming.* Ed. by Luca Aceto, Monika Henzinger, and Jiří Sgall. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 330–341. ISBN: 978-3-642-22006-7.