RESULTS ON SUM-FREE SEQUENCES

AARON KAUFER, DAVID LIN KEWEI, JONATHAN AKABA

ABSTRACT. A sequence or set of numbers is said to be sum-free if no element is the sum of two other elements. In this paper, we survey a variety of results about sum-free sequences. Additionally, we prove an original result that gives a criteria under which the indicator sequence of a sum-free sequence is eventually periodic.

1. INTRODUCTION

Definition 1.1. A subset S of the natural numbers (or, more generally, of any abelian group) is said to be **sum-free** if no element of S is a sum of two (not necessarily distinct) elements of S. Formally, S is sum-free if $S \cap (S + S) = \varnothing$. An increasing sequence of natural numbers is said to be **sum-free** if the set of its elements is sum-free.

The study of sum-free sets dates back to at least 1916 when Schur proved that the set of positive integers cannot be partitioned into a finite number of sum-free sets. Further studies of sum-free sets gave rise to vertex-transitive triangle-free graphs with applications in fields ranging from Ramsey Theory to extremal Graph Theory (Haviv and Levy [2018]).

There is a natural bijection between the set $\{0, 1\}^{\mathbb{N}}$ of infinite binary sequences and the set of sum-free sequences of natural numbers that will be described in Subsection 2.2. A long standing question about sum-free sequences posed by Cameron [1987] asks if the periodicity of a sum-free sequence's associated binary sequence implies the regularity of the sum-free sequence itself. While the converse of this conjecture has been shown to be true by Cameron [1987], the question of whether the binary sequence's periodicity implies the sum-free sequence's regularity remains open. Despite this, there is evidence to suggest that the conjecture is false (Cameron [1987]).

One natural method of constructing sum-free sequences is by taking a greedy approach: begin with an arbitrary finite increasing sequence of numbers, and then inductively choose the next number to be the smallest natural number that is not a sum of two previous numbers. A striking empirical observation is that for nearly every choice of initial data in a greedy sum-free sequence, the sequence eventually takes on a "period". Specifically, we have the following definition:

Definition 1.2. An increasing sequence of natural numbers S is said to be **regular** if there exists some positive $q \in \mathbb{N}$ such that $n \in S$ if and only if $n+q \in S$ for sufficiently large N.

If we consider the indicator sequence

$$1_S(x) := \begin{cases} 1 & x \in S \\ 0 & x \notin S \end{cases}$$

then S being regular is equivalent to saying that 1_S is eventually periodic with period q. Motivated by this, we shall call q a **period** of S.

While it was initially conjectured that all greedy sum-free sequences are regular, computational evidence now seems to suggest that this is false. This phenomenon is explored further in Section 2.

Given the nature of sum-free sets, a natural quantity to explore is the amount of ways any given number can be expressed as a sum of two elements in the set. Specifically, we make the following definition:

Definition 1.3. Suppose H is a sum-free sequence of natural numbers. Then, for $k \in \mathbb{N}$, we define

$$R_H(k) := \left|\left\{(a,b) \in H^2 \mid a+b=k\right\}\right|.$$

Much of the results about sum-free sequences will have analogues in $\mathbb{Z}/N\mathbb{Z}$, where the notion of a sum-free sequence gets replaced by that of a sum-free subset of $\mathbb{Z}/N\mathbb{Z}$. For these analogues, we have the following definition:

Definition 1.4. Suppose $A \subseteq \mathbb{Z}/N\mathbb{Z}$ is sum-free. Then, for $k \in \mathbb{Z}/N\mathbb{Z}$, we define

$$r_A(k) = \left|\left\{(a,b) \in A^2 \mid a+b=k\right\}\right|.$$

These two quantities end up being crucial to the study of sum-free sets. In fact, the main result of this paper, given in Section 5, is a proof that if R_H grows at least linearly outside of H, then H must be regular, which can be seen as a weaker version of the (likely false) conjecture which states that all greedy sum-free sequences are regular. We also prove an analogous result about sum-free subsets of $\mathbb{Z}/N\mathbb{Z}$. The proof of our main result will rely heavily on some discrete Fourier analysis, so we will develop the necessary tools in Section 3 and explore some of their uses in Section 4, before giving the proof of the main result in Section 5.

2. Greedy Sum-Free Sequences

We begin with a formal definition of a greedy sum-free sequence:

Definition 2.1. An increasing sum-free sequence of natural numbers $H = (a_1, a_2, \ldots)$ is said to be **greedy** if for all sufficiently large n, we have that a_n is the smallest number greater than a_{n-1} which cannot be expressed as a sum of two elements in $\{a_1, a_2, \ldots, a_{n-1}\}$.

It is straightforward to see that the following definition is equivalent:

Definition 2.2 (alternative)**.** An increasing sum-free sequence of natural numbers H is said to be greedy if for all sufficiently large $k \notin H$, we have $R_H(k) > 0$.

As mentioned in the introduction, the following conjecture served as the motivation for much of this paper, despite being likely false:

Conjecture 2.3 (Strong regularity)**.** *All greedy sum-free sequences are regular.*

Despite still being an open problem, there is ample evidence from Calkin and Finch [1996] suggesting that this conjecture is in fact false. In fact, the greedy sum-free sequences beginning with any of the following triples:

$$(8, 18, 30), (8, 27, 32), (9, 16, 29), (9, 26, 32)$$

do not appear to be periodic despite being analyzed up to 10^7.

2.1. Finite Greedy Sum-Free Sequences.

Although Conjecture 2.3 is likely false, there is a variant that is easily seen to be true. In particular, we introduce the notion of a finite-greedy sum-free sequence:

Definition 2.4. An increasing sequence of natural numbers $a_1, a_2, \ldots$ is said to be **finite-greedy** if there is some finite $L > 0$ such that for all $i, j \in \mathbb{N}$ with $j \leq L$, we have $a_i + a_j$ is not an element of the sequence.

This is essentially a greedy sum-free sequence, with the key difference being that instead of avoiding numbers that are sums of any two numbers in the sequence, we only avoid numbers that are sums of two numbers where at least one number is among the first L terms of the sequence. If we allow L to be infinity, then we recover the definition of a greedy sum-free sequence. The analogue of 2.3 is then:

Theorem 2.5 (Calkin and Finch [1996])**.** *All finite-greedy sum-free sequences are regular.*

Proof. Let $S = (a_1, a_2, \ldots)$ be a finite-greedy sum-free sequence and let L be an in the definition of finite-greedy. Consider a block of length a_L within the indicator sequence 1_S of S.

Note that every term that appears after such a block depends only on the terms within the block. To see why this is true, note that to test whether some natural number z belongs to S, we need only check whether $z - k$ belongs to S for $k \in \{a_1, ..., a_L\}$. If z appears after a given block, then all such $z - k$ will either fall within the block, or will fall in some later block which by induction also depends only on our initial block. Thus, to prove the regularity of S, it suffices to show that there are two identical blocks in 1_S of length a_L.

There are 2^{a_L} possible arrangements of a block in 1_S of size a_L. Therefore, by the pigeonhole principle, if we consider $2^{a_L} + 1$ (possibly overlapping) blocks of this size in the indicator sequence, we must be able to find at least two matching blocks, completing the proof. In fact, this proof gives an upper bound to the period of 1_S of $2^{a_L} + 1$. □

Despite proving that all finite greedy sum-free sequences are regular, the above proof gives no insight into the finer structure of these sequences. For instance, other than the above proof showing that the smallest period of the greedy finite sum-free sequence is bounded by $2^{a_L} + 1$, it gives us no information on what specific periods are attained by these finite greedy sum-free sequences. From empirical results, the bound of $2^{a_L} + 1$ seems too high. In fact, when $L = 1$ and $L = 2$, we have proven a significantly stronger upper bound on the period of S.

Theorem 2.6. *Suppose $S = (a_1, a_2, \ldots)$ is a finite-greedy sum-free sequence with $L = 1$. Then $2a_1$ is a period of S.*

Proof. Consider large n. Then, if $n \in S$, we necessarily have $n + a_1 \notin S$. Since $n + a_1 = (n + 2a_1) - a_1 \notin S$, it follows that $n + 2a_1 \in S$. A similar argument shows that $n \notin S$ implies $n + 2a_1 \notin S$. □

Theorem 2.7. *Suppose $S = (a_1, a_2, \ldots)$ is a finite-greedy sum-free sequence with $L = 2$. Then $a_1 + a_2$ is a period of S.*

Proof. For large enough n, n being in S only depends on $n - a_1$ and $n - a_2$. Thus, suppose $n \in S$. Then $n + a_1, n + a_2 \notin S$. Thus, $(n + a_1 + a_2) - a_1 \notin S$ and $(n + a_1 + a_2) - a_2 \notin S$, so it follows that $n + a_1 + a_2 \in S$. A similar argument shows $n \notin S$ implies $n + a_1 + a_2 \notin S$. □

Unfortunately, for $L > 2$, there does not appear to be any general simple expression for a period of S.

2.2. **Random Sum-Free Sequences.**

Let $\mathcal{S}$ denote the set of all sum-free sequences. Cameron [1987] describes a natural way to set up a probability measure on $\mathcal{S}$ by constructing a natural bijection between $\{0, 1\}^{\mathbb{N}}$ and $\mathcal{S}$.

An informal description of this bijection is as follows: consider for example, a sequence of bits $s \in \{0,1\}^{\mathbb{N}}$. Then, we construct a sum-free sequence as follows. Beginning with 1, we iterate through the natural numbers in order. If a number is already the sum of two previous numbers, we skip over it, and if not, then we include it if and only if the next bit of s is 1. It is straightforward to see that this association defines a bijection.

This bijection allows us to equip $\mathcal{S}$ with a probability measure via the Bernoulli distribution on $\{0,1\}^{\mathbb{N}}$.

An interpretation of a random sample of this probability measure is as follows: beginning with 1, we iterate over the natural numbers. For each element that is a sum of two previously chosen numbers, we skip it, and for all other elements, we flip a fair coin to determine whether or not to include it.

For the remainder of this section, we let S denote a random sum-free sequence, we let $[k] = \{1, \dots, k\}$, and we let $S_k = S \cap [k]$.

A counter-intuitive property of random sum-free sequences, originally discovered by Cameron, is that there is a positive probability of all terms being odd. While this can be empirically confirmed, there is also a heuristic argument for this: there is a positive probability that the first N terms of a random sum-free sequence consists of exactly the first N odd numbers (i.e. $1, 3, 5, ..., 2N-1$) for some fixed large N. This automatically excludes the even numbers up to $4N-2$, which means that the odd numbers $2N+1, ..., 4N-1$ are selected independently by coin flips. Since $4N$ is automatically excluded if any of the pairs

$$(1, 4N-1), (3, 4N-3), \dots, (2N-1, 2N+1)$$

are included, we expect $4N$ (and by vague analogy all larger even numbers) to be included with an exponentially small probability (in this case, 2^{-N}).

More generally, we may pose the following question for any sum-free sequence H:

Question. For what sum-free $H \subset \mathbb{N}$ do we have $P(S \subset H) > 0$?

As it turns out, the answer to this question is heavily tied to the R_H function defined earlier. In particular, we have the following partial answer, due to Calkin [1998]:

Theorem 2.8 (Calkin [1998])**.** *Suppose $H \subseteq \mathbb{N}$ is sum-free and that*

$$\sum_{k \notin H} \left(\frac{3}{4}\right)^{R_H(k)} < \infty.$$

Then $P(S \subseteq H) > 0$.

Proof. To aid with the proof, we first define $R'_H(k) = \{(a,b) \in H^2 \mid a+b=k, a \geq b\}$. It is clear that the hypothesis of the theorem is satisfied if and only if the same statement is satisfied with R'_H instead of R_H, so we will work with R'_H instead.

Define the event $E_k = \{S_{k-1} \subseteq H \text{ and } k \in S\}$. Then, for S to not be a subset of H, there must be some $k \notin H$ for which E_k is true. Thus, $P(S \not\subseteq H) \leq \sum_{k \notin H} P(E_k)$. Thus, if we can show that $\sum_{k \notin H} P(E_k) = 1 - \varepsilon$ for some $\varepsilon > 0$, then we have $P(S \subseteq H) > \varepsilon$ as desired. It suffices to show that $\sum_{k \notin H} P(E_k) < \infty$, because then we may replace each event E_k with some $E'_k = E_k \cap \{S_N = H \cap [N]\}$ for some large enough N (since H is infinite, this will decrease $\sum_{k \notin H} P(E'_k)$, so for large enough N, it will be less than 1.

Since $\sum_{k \notin H} (\frac{3}{4})^{R'_H(k)} < \infty$, it thus clearly suffices to show that $P(E_k) < (\frac{3}{4})^{R'_H(k)}$ for all $k \notin H$. For this, suppose A is a sum-free subset of $[k]$ for which $A \cap [k-1] \subseteq H$ and $k \in A$ (such subsets are simply subsets that S_k could be to satisfy E_k). Then, since A is sum-free, one can easily see that $P(S_k = A) = 2^{-m(A)}$ where $m(A) = |[k] - (A + A)|$

Then, as events, we have

$$E_k = \bigcup_{\substack{A \subseteq [k] \text{ sum-free} \\ A \cap [k-1] \subseteq H \\ k \in A}} \{S_k = A\}$$

so

$$P(E_k) \leq \sum_{\substack{A \subseteq [k] \text{ sum-free} \\ A \cap [k-1] \subseteq H \\ k \in A}} P(S_k = A) = \sum_{\substack{A \subseteq [k] \text{ sum-free} \\ A \cap [k-1] \subseteq H \\ k \in A}} 2^{-m(A)}$$

Now let $l = |H \cap [k]|$. Then, by construction we have $l \leq m(A)$ for all A above. Thus, $P(E_k) \leq \sum_A 2^{-l}$, where l does not depend on A. Thus, it suffices to count the number of such A. For thus, we note that since A is sum-free, of every pair (a, b) such that $a+b = k$, at most one of a, b can be included. For each pair, at most three scenarios can happen: neither are included, only a is included, and only b is included. The number of such pairs is precisely $R'_H(k)$. Then, for any given pair, there are at most $2^{l-2R'_H(k)}$ inclusions of the remaining elements, so there are at most $3^{R'_H(k)} 2^{l-2R'_H(k)}$ possible A. Thus, we have

$$P(E_k) \leq (3^{R'_H(k)} 2^{l-2R'_H(k)}) 2^{-l} = \left(\frac{3}{4}\right)^{R'_H(k)}$$

as desired. □

With this theorem, we can rigorously prove the initial claim:

Corollary 2.9. *If S is the random sum-free sequence, then with positive probability, all terms of S are odd.*

Proof. If we let H denote the sequence of all odd numbers, then we can see that R_H grows linearly outisde H (approximately $k/2$ for even k), so $\sum_{k \notin H} \left(\frac{3}{4}\right)^{R_H(k)}$ clearly converges. □

In fact, this corollary generalizes to other regular sum-free sequences:

Corollary 2.10. *Suppose $A \subseteq \mathbb{Z}/N\mathbb{Z}$ is sum-free and complete in the sense that $A \sqcup (A+A) = \mathbb{Z}/N\mathbb{Z}$. Then if we let $H = \left\{k \in \mathbb{N} \mid \overline{k} \in A\right\}$ where $\overline{k}$ denotes the residue of k mod N, we have $P\left(S \subseteq H\right) > 0$.*

Proof. Suppose $k \in \mathbb{N}$ with $k \notin H$. Then $\overline{k} \notin A$, so $\overline{k} \in A + A$. Then, if $\overline{k} = \overline{a} + \overline{b}$ for $\overline{a}, \overline{b} \in A$, then we have $k = (m_1 N + a) + (m_2 N + b)$ for roughly $\frac{k}{N}$ choices of pairs (m_1, m_2) with $m_1 N + a, m_2 N + b > 0$. By definition, each of $m_1 N + a, m_2 N + b \in H$, so we have that $R_H(k)$ grows at least linearly. Thus $\sum_{k \notin H} \left(\frac{3}{4}\right)^{R_H(k)}$ converges. □

Intuitively, Theorem 2.8 says that if R_H grows fast enough, then $P\left(S \subseteq H\right) > 0$.

A partial converse to Theorem 2.8, also due to Calkin [1998], is as follows:

Theorem 2.11 (Calkin [1998])**.** *Let $H \subseteq \mathbb{N}$ be sum-free. Suppose there exists a sequence $n_1, n_2, \ldots \notin H$ such that $n_{k+1} > 2n_k$ and $\sum_k 2^{-R_H(n_k)} = \infty$. Then $P(S \subseteq H) = 0$.*

Proof. Again, for convenience we work with R'_H instead of R_H.

Let E_k be the event that $S_{n_k} \subseteq H$. Then, as E_{k+1} clearly implies E_k, we have $P(E_k) = P(E_k \mid E_{k-1}) P(E_{k-1} \mid E_{k-2}) \cdots P(E_1)$.

We now note that $P(a \notin S \mid E) \geq \frac{1}{2}$ as long as E is any event that only depends on (at most) the first $a-1$ elements of S. To see this, note that depending on S_{a-1}, we have that either a is forbidden from inclusion into S, in which case the probability is 1, or a is permitted, in which case the probability is $1/2$, so in every case, the probability is $\geq \frac{1}{2}$.

We note that for E_k to happen, it is necessary for n_k to be excluded from S, so we have $P(E_{k+1} \mid E_k) \leq P(n_{k+1} \notin S \mid E_k)$. Let $c = R'_H(n_{k+1})$, so that $n_{k+1} = a_1 + b_1 = \cdots = a_c + b_c$ are the only c ways of expressing n_{k+1} as a sum of two elements in H. Without loss of generality, we assume $a_1 > a_2 > \cdots > a_c$, and $a_i \geq b_i$. Then, since $n_{k+1} > 2n_k$, we have $a_i > n_k$. Then, if all of the a_i are excluded from S, then n_{k+1} is permitted to be included in S, in which case it has a $\frac{1}{2}$ probability of inclusion. Thus, we have

$$P(n_{k+1} \in S \mid E_k) \geq \frac{1}{2} P(a_1 \notin S \wedge \cdots \wedge a_c \notin S \mid E_k)$$

By chaining conditional probabilities, we get

$$\begin{aligned}
P(a_1 \notin S \wedge \cdots \wedge a_c \notin S \mid E_k) &= P(a_1 \notin S \wedge \cdots \wedge a_{c-1} \notin S \mid E_k \wedge a_c \notin S) P(a_c \notin S \mid E_k) \\
&\geq \frac{1}{2} P(a_1 \notin S \wedge \cdots \wedge a_{c-1} \notin S \mid E_k \wedge a_c \notin S) \\
&= \frac{1}{2} P(a_1 \notin S \wedge \cdots \wedge a_{c-2} \notin S \mid E_k \wedge a_c \notin S \wedge a_{c-1} \notin S) \\
&\quad \times P(a_{c-1} \notin S \mid E_k \wedge a_c \notin S)
\end{aligned}$$

$$\geq \left(\frac{1}{2}\right)^2 P(a_1 \notin S \wedge \cdots \wedge a_{c-2} \notin S \mid E_k \wedge a_c \notin S \wedge a_{c-1} \notin S)$$
$$\vdots$$
$$\geq \left(\frac{1}{2}\right)^c$$

Thus, we have $P(n_{k+1} \in S \mid E_k) \geq 2^{-R'_H(n_{k+1})-1}$, so in particular $P(E_{k+1} \mid E_k) \leq 1 - 2^{-R'_H(n_{k+1})-1}$. Thus, we get $P(E_k) \leq \prod_{r\leq k}\left(1 - 2^{-R'_H(n_r)-1}\right)$, and since $\sum 2^{-R'_H(n_k)}$ diverges, it follows that $P(E_k) \to 0$, so in particular $P(S \subseteq H) = \lim P(S_{n_k} \subseteq H) = 0$. □

Although Theorem 2.11 is not a full converse to 2.8, it does have the following important corollary:

Corollary 2.12. *Suppose $H \subseteq \mathbb{N}$ is sum-free and that R_H does not tend to infinity outside of H. Then $P(S \subseteq H) = 0$.*

Proof. If R_H does not tend to infinity outside of H, then there exists some N such that $R_H(k) \leq N$ for infinitely many $k \notin H$. In particular, we choose $n_k \notin H$ such that $n_{k+1} > 2n_k$ and $R_H(k) \leq N$. Then $\sum 2^{-R_H(n_k)} \geq \sum 2^{-N} = \infty$. □

3. The Fourier Transform of a Sequence

For this section, we define $e(x) = e^{2\pi i x}$

Definition 3.1. Let $a_1, a_2, \ldots$ be an increasing sequence of natural numbers. Then, the **(Discrete) Fourier Transform** of (a_n) with T terms is defined as:

$$f_T(x) = \sum_{n=1}^{T} e(a_n x)$$

There is also a natural analogue for subsets of $\mathbb{Z}/N\mathbb{Z}$:

Definition 3.2. Let $A \subseteq \mathbb{Z}/N\mathbb{Z}$. Then, the **Fourier Transform** of A is defined for $x \in \mathbb{Z}/N\mathbb{Z}$ as

$$f(x) = \sum_{a\in A} e\left(\frac{ax}{N}\right)$$

Clearly f_T is always periodic with period 1. Since $e(a_n x)$ only depends on the fractional part $\{a_n x\}$ of $a_n x$, and since $e(a_n x)$ is the point on the unit circle with angle $2\pi\{a_n x\}$, it follows that the more evenly distributed $\{a_n x\}$ is in the interval $[0, 1]$, the more cancellation the sum $f_T(x)$ sees. In particular, the following result was proved in Weyl [1916]:

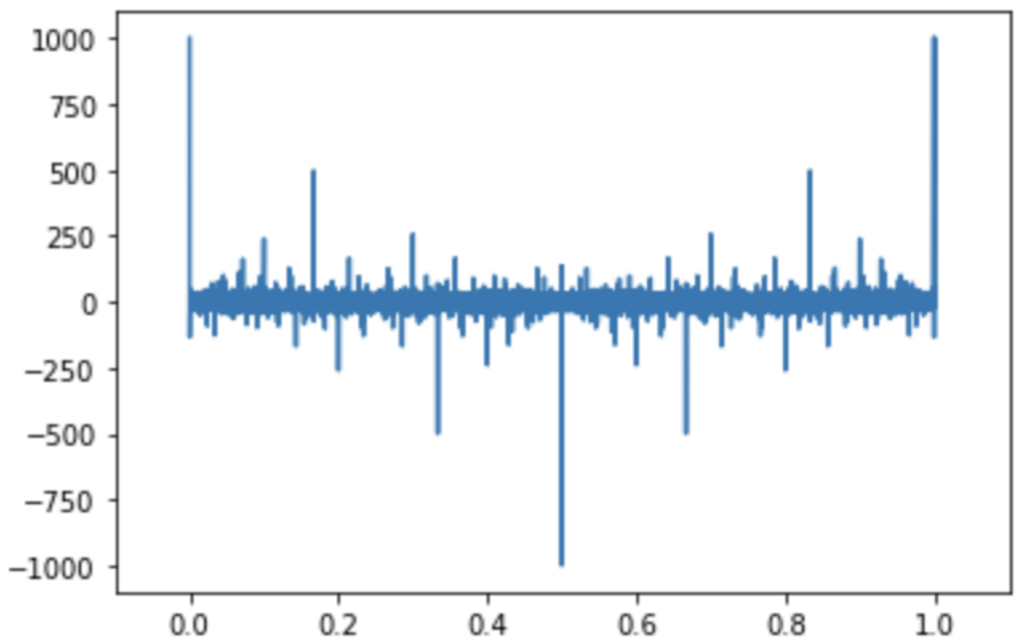
Theorem 3.3 (Weyl [1916]). *Let a_n be any increasing sequence of natural numbers. For almost all $x \in [0,1]$, we have $f_T(x) = o(T)$.*

This theorem motivates the definition:

Definition 3.4. Suppose a_n is an increasing sequence of natural numbers. A number $\alpha \in [0,1]$ is called a **signal** for the sequence (a_n) if $f_T(\alpha) \neq o(T)$

In order for cancellation to not occur in the sum for $f_T(\alpha)$, the numbers $\{a_n\alpha\}$ must not be evenly distributed in $[0,1]$, so there must be some subset of $[0,1]$ for which there is a higher concentration of the $\{a_n\alpha\}$. This idea of studying sequences by analyzing their signals and the corresponding distributions of $\{a_n\alpha\}$ is based on Steinerberger [2017].

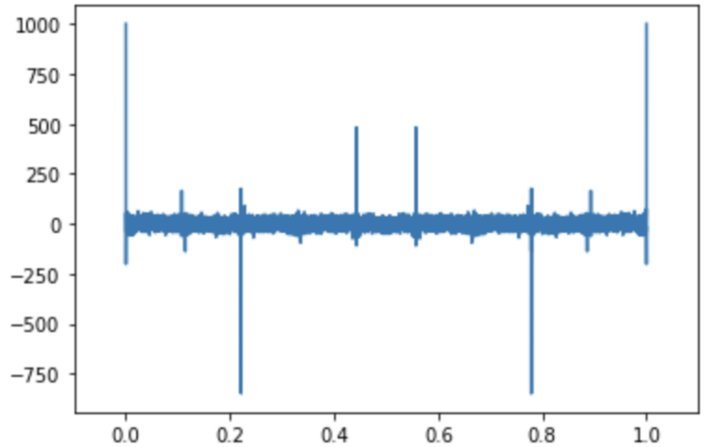
Example 3.5. Let p_n denote the sequence of prime numbers. Then, the plot of the fourier transform of p_n with 1000 terms is shown below:



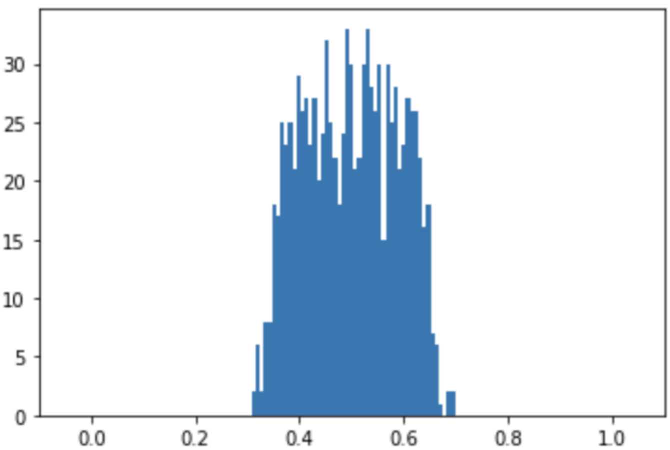
The most prominent signals (up to symmetry) occur at $x = \frac{1}{2}$, $x = \frac{1}{3}$, and $x = \frac{1}{6}$. This is unsurprising, since, for example, $e\left(\frac{1}{2}p_n\right) = -1$ for all but $p_1 = 2$, and $e\left(\frac{1}{3}p_n\right) = e\left(\frac{1}{3}\right)$ or $e\left(\frac{2}{3}\right)$, both of which have negative real part.

As can be seen from this example, rational signals p/q detect some sort of "regularity" of the sequence mod q.

Example 3.6. Consider a random sum-free sequence (chosen randomly using the probability measure constructed earlier) which starts with $a_n = 2, 3, 7, 11, 15, 20, 21, 25, 29, 34, \ldots$. The plot of its fourier transform with 1000 terms is shown below:



Its primary signal occurs at $\alpha \approx 0.221$. When plotting a histogram of $\{\alpha a_n\}$, we get the following distribution:



This histogram is a plot of all the values of $\{\alpha a_n\}$ for $n \in \{1, \ldots, 10000\}$ rounded down to the nearest hundredth.

As can be seen from the distribution, almost every value of $\{a_n \alpha\}$ lies in the interval $\left(\frac{1}{3}, \frac{2}{3}\right)$. This phenonenon actually seems to occur with rather high probability. More formally, we have the following conjecture:

Conjecture 3.7. *Let $S = (a_1, a_2, \ldots)$ denote a random sum-free sequence. Then, with positive probability, there exists $\alpha \in [0, 1]$ such that $\{\alpha a_n\} \in \left(\frac{1}{3}, \frac{2}{3}\right)$ for almost all a_n (where almost all means density 1)*

In fact, $\left(\frac{1}{3}, \frac{2}{3}\right)$ is not the only subset of $[0, 1]$ that appears to have this property, but it does appear to be the only sum-free (mod 1) subset of $[0, 1]$ for which the probability of the event in the conjecture is greater than $\frac{1}{2}$. This phenomenon will be explored more closely in the section on sum-free complete subsets of the torus.

4. Sum-Free Complete Subsets of the Torus

In this chapter, we examine a mysterious relationship between sum-free sequences with some signal α, and sum-free subsets of the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$.

4.1. **Supporting subsets of a sequence.**

Example 3.6 displayed an interesting phenomenon: a randomly chosen sum-free sequence turned out to have a signal α for which almost all of $\{\alpha a_n\}$ lied in the interval $\left(\frac{1}{3}, \frac{2}{3}\right)$. In fact, for most such randomly generated sum-free sequences, this seems to be the case, which begs the question: what is special about $\left(\frac{1}{3}, \frac{2}{3}\right)$. One important property is that it is sum-free as a subset of $\mathbb{T}$. In fact, we may consider the reverse construction:

Definition 4.1. Let $\alpha \in \mathbb{R}$. Then we define the sequence $H_\alpha = \left\{n \in \mathbb{N} \mid \{n\alpha\} \in \left(\frac{1}{3}, \frac{2}{3}\right)\right\}$

Then, precisely because $\left(\frac{1}{3}, \frac{2}{3}\right)$ is sum-free subset of the torus, it follows that H_α is a sum-free sequence. In particular, it is easy to see from its definition that α is a signal of H_α, and the distribution of $\{\alpha a_n\}$ (here $\{\}$ means fractional part) for $a_n \in H_\alpha$ be entirely contained in the interval $\left(\frac{1}{3}, \frac{2}{3}\right)$. In fact, for α irrational, $\{\alpha a_n\}$ will be equidistributed in $\left(\frac{1}{3}, \frac{2}{3}\right)$.

Before continuing to explore precisely why $\left(\frac{1}{3}, \frac{2}{3}\right)$ has this property, we give a definition to help make this notion precise:

Definition 4.2. Given a sum-free sequence $H = \{a_1 < a_2 < ...\}$ and a sum-free subset $I \subset \mathbb{T}$ we say that I **supports** H (or equivalently, H **is supported by** I) if there exists some $\alpha \in [0, 1]$ such that

$$\lim_{N \in \infty} \frac{\#\{n : \{a_n\alpha\} \in I, n \leq N\}}{N} = 1$$

Additionally, we say that a sum-free subset $I \subseteq \mathbb{T}$ **exactly supports** H if it supports H, yet no subset of I of strictly smaller measure also supports H.

Then, Conjecture 3.7 says that with positive probability, a random sum-free sequence is supported by $\left(\frac{1}{3}, \frac{2}{3}\right)$. Additionally, the stronger statement that a random sum-free sequence is exactly supported by $\left(\frac{1}{3}, \frac{2}{3}\right)$ with positive probability appears to be true.

Interestingly, $\left(\frac{1}{3}, \frac{2}{3}\right)$ is not the only subset of the torus on which a random sum-free sequence is supported with positive probability. Computational evidence seems to suggest that $\left(\frac{1}{7}, \frac{2}{7}\right) \cup \left(\frac{5}{7}, \frac{6}{7}\right)$ also has this property. This set is of course sum-free, but it begs the question as to why other sum-free subsets of the torus, such as $\left(\frac{1}{4}, \frac{2}{4}\right)$ don't seem to occur with positive probability. Specifically, we wish to answer the question: which subsets of the torus occur with positive probability as the support of a random sum-free sequence?

The answer seems to have to do with the following definition:

Definition 4.3. A sum-free set $I \subseteq \mathbb{T}$ is **complete** if $I \sqcup (I+I)$ has measure 1.

In other words, I is complete if $I \sqcup (I+I) = \mathbb{T}$ up to a set of measure 0. It is readily seen that sets like $\left(\frac{1}{3}, \frac{2}{3}\right)$ and $\left(\frac{1}{7}, \frac{2}{7}\right) \cup \left(\frac{5}{7}, \frac{6}{7}\right)$ are complete while $\left(\frac{1}{4}, \frac{2}{4}\right)$ is not.

From this, we conjecture the following:

Conjecture 4.4. *If a sum-free subset $I \subseteq \mathbb{T}$ is such that a random sum-free sequence is exactly supported by I with positive probability, then I is complete.*

As there are an uncountable number of complete sum-free subsets of the torus (for example, $(x, 2x) \cup (1-2x, 1-x)$ works for any $\frac{1}{8} < x < \frac{1}{6}$), it follows that it cannot be true that all such subsets occur as the exact support of a random sum-free sequence with positive probability. Instead, the following conjecture is the current best guess as to classifying when this phenomenon occurs.

Conjecture 4.5. *Let $I \subset \mathbb{T}$ be a complete sum-free subset, which is expressible as the union of finitely many intervals with rational endpoints. Then for a random sum-free sequence S, I supports S with positive probability.*

4.2. **Sequences induced by torus subsets.**

We can generalize the construction of H_α to other sum-free subsets of the torus. Specifically, for a positive real number α and a sum-free subset $I \subset \mathbb{T}$, we may consider the sequence

$$H_\alpha^I = \{n : \{n\alpha\} \in I\}$$

It is clear that this sequence is sum-free.

We may also ask if the induced sequence H_α^I is ever greedy. Necessarily, I would have to be complete for this to happen. In Calkin and Erdős [1996], it was proved that $H_\alpha^{(1/3,2/3)}$ was not greedy. The proof proceeds by considering the best rational convergents $\frac{p_n}{q_n}$ of 3α and considering various cases depending on $(q_{2n}, q_{2n+1}) \bmod 3$. In particular, the argument only applies specifically to the interval $(1/3, 2/3)$.

It is not clear how to generalize this to other examples of I such as $(1/7, 2/7) \cup (5/7, 6/7)$, but an example of I for which similar approaches must fail is if I was the union of intervals with irrational endpoints. For instance, if $\frac{1}{8} < x < \frac{1}{6}$ is irrational, then this approach must fail for the subset $I_x = (x, 2x) \cup (1-2x, 1-x)$.

In general, we conjecture the following:

Conjecture 4.6. *For sum-free $I \subset \mathbb{T}$, $H = H_\alpha^I$ is not greedy. In other words, there exists infinitely many $k \notin H$ where $R_H(k) = 0$.*

However, under the assumption that I is the union of finitely many intervals, we can prove that R_H is small infinitely often:

Theorem 4.7. *Let I be a sum-free subset of $\mathbb{T}$ expressible as the union of intervals. Then for $H = H_\alpha^I$, there exists some $c > 0$ such that there are infinitely many $n \notin S$ satisfying $R_H(n) < c\sqrt{n}$.*

Proof. By Dirichlet approximation, it follows that

$$\liminf_{n\to\infty} \|n\alpha\| = 0$$

and clearly $\|n\alpha\| > 0$.

Pick a sufficiently large M where

$$\|M\alpha\| = \min_{1\le n\le M} \|n\alpha\|$$

It follows that for distinct a, b differing by at most M:

$$\|a\alpha - b\alpha\| = \||b-a| \cdot \alpha\| \ge \|M\alpha\|$$

Denote $\varepsilon = \|M\alpha\|$. By the above, $\alpha, 2\alpha, ..., M\alpha$ are mutually at least ε away from each other along $\mathbb{T}$, which implies that $\varepsilon < 1/M$.

We claim that any open interval of length ε in $\mathbb{T}$ has at most M elements from the set $\{\alpha, 2\alpha, ..., \lfloor \frac{M}{\varepsilon} \rfloor \alpha\}$. Suppose that there were distinct $n_1, ..., n_{M+1} \le M/\varepsilon$ such that $n_i\alpha \in (x, x+\varepsilon)$ for every i, then there exists distinct i, j such that $|n_i - n_j| \le 1/\varepsilon < M$, so $\|n_i\alpha - n_j\alpha\| \ge \varepsilon$, a contradiction.

Express $I = \bigsqcup_{i=1}^k I_k$ where $\{I_k\}$ are disjoint open intervals. Select x to be the endpoint of any interval I_k. Since I is sum-free, this implies that $I \cap (x - I)$ has measure 0, so for a small $\delta > 0$ (depending only on I), $I \cap (x + \delta - I)$ is the union of at most K intervals (where K depends on only I) of length at most δ. Heuristically, we can now bound the total number of points in $I \cap (x + \delta - I)$, which leads to a bound on R_H if $x + \delta$ is a multiple of α (since if $y\alpha + z\alpha = x + \delta$ then $y\alpha \in I \cap (x + \delta - I)$).

Hence, we want to pick a point (on $\mathbb{T}$) of the form $n\alpha$ that comes very slightly after the point x. To do this, we start from 0 and take steps of size $\|M\alpha\|$.

Specifically, pick $n = \lceil x\varepsilon^{-1} \rceil \cdot M$, then $n\alpha = x + \delta$ for some $0 < \delta < \varepsilon$. Supposing that M is sufficently large so that δ is sufficiently small for the given I. then

$$\begin{aligned} R_H(n) &\le \#\{m : m \le n, m \in I \cap (x + \delta - I)\} \\ &\le K \cdot M \\ &\le K\sqrt{M\varepsilon^{-1}} \\ &< Kx^{-1/2} \cdot n^{1/2} \end{aligned}$$

□

5. A Sufficient Condition for Regularity of a Sum-Free Sequence

For this section, H will be an infinite sum-free subset of $\mathbb{N}$.

The strong regularity conjecture that all greedy sum-free sequences are regular can be restated as:

Conjecture 5.1 (Strong regularity, restated)**.** *If $R_H(k) > 0$ for all $k \notin H$ sufficiently large, then H is regular.*

As an weaker version of this conjecture, we have proven the following result:

Theorem 5.2. *Suppose there exists $c > 0$ such that $R_H(k) > ck$ for all $k \notin H$. Then H is regular.*

In other words, we wish to find $q \in \mathbb{N}$ with $n \in H \iff n + q \in H$ for large n. To do this, we first prove that there is an $\varepsilon-$"almost period", in the sense that $n \in H \iff n + q \in H$ for all but an ε fraction of numbers in the sequence.

Lemma 5.3. *For all $\varepsilon > 0$, there exists $Q, M \in \mathbb{N}$ such that for all $m > M$, there exists $q \leq Q$ (dependent on m) such that*

$$\frac{|\{n \leq m \mid n \in H, n - q \notin H\}|}{m} < \varepsilon$$

Proof. Let $f_N(x)$ denote the fourier transform of H as defined earlier in the paper. We begin by isolating the "signals" of f_N.

Fix $\varepsilon > 0$ and $N > 0$. Construct a sequence $\alpha_{N,1}, \alpha_{N,2}, \ldots \in [0,1]$ and a sequence of intervals $I_{N,1}, I_{N,2}, \ldots$ inductively as follows: Pick $\alpha_{N,j} \in [0,1] - I_{N,1} \cdots - I_{N,j-1}$ such that $|f_N(\alpha_{N,j})| > \varepsilon N$ (if there are no such $\alpha_{N,j}$, then terminate the process), and let $I_{N,j} = [\alpha_{N,j} - \frac{\varepsilon}{4\pi N}, \alpha_{N,j} + \frac{\varepsilon}{4\pi N}]$ (here we are thinking of $[0,1]$ as S^1, so intervals wrap around). This process must terminate in a finite number of steps. To see this, we first note that for $x \in I_{n,j}$, we have $|f_n(x)| \geq \frac{\varepsilon N}{2}$. In particular, we have that

$$\begin{aligned}
|f_N(x) - f_N(\alpha_{N,j})| &= \left|\sum_{a \in H_n} (e(ax) - e(a\alpha_{N,j}))\right| \leq \sum_{a \in H_n} |e(ax) - e(a\alpha_{N,j})| \\
&\leq \sum_{a \in H_n} 2\pi a \,|x - \alpha_{N,j}| \leq N 2\pi N \left(\frac{\varepsilon}{4\pi N}\right) = \frac{\varepsilon N}{2}
\end{aligned}$$

and since $|f_N(\alpha_{N,j})| > \varepsilon N$, it follows that $|f_N(x)| > \frac{\varepsilon N}{2}$. Then, we have:

$$\int_{I_{N,j}} |f_N(x)|^2 \, dx \geq \left(\frac{\varepsilon N}{2}\right)^2 |I_{N,j}| = \left(\frac{\varepsilon N}{2}\right)^2 \left(\frac{\varepsilon}{2\pi N}\right) = \frac{\varepsilon^3 N}{8\pi}$$

Then, given that the $I_{N,j}$ are a sequence of intervals of same length for which the center of one is never inside another, each $x \in [0,1]$ can belong to at most two of the $I_{N,j}$. Thus, we have:

$$\sum_j \int_{I_{N,j}} |f_N(x)|^2 \, dx \leq 2 \int_0^1 |f_N(x)|^2 \, dx = 2\,|H_N| \leq 2N$$

but

$$\sum_j \int_{I_{N,j}} |f_N(x)|^2 \, dx \geq \sum_j \frac{\varepsilon^3 N}{8\pi}$$

Thus, as the latter quantity is bounded by $2N$, it follows that the amount of terms in the sum must be finite. Additionally, if d is the amount of terms in the sum, we have $d\left(\frac{\varepsilon^3 N}{8\pi}\right) \leq 2N$, so $d < \frac{16\pi}{\varepsilon^3}$. In particular, let $D = \left\lfloor \frac{16\pi}{\varepsilon^3} \right\rfloor$, so for any N, the number of $\alpha_{N,j}$ is at most D. Importantly, this bound does not depend on N.

Now, let $Q > \varepsilon^{-2D}$ and let $\alpha_{N,1}, \ldots, \alpha_{N,d}$ where $d \leq D$ be chosen as above. For $x \in \mathbb{R}$, let $\|x\|$ be the distance from x to the nearest integer to x. For $q \in \mathbb{N}$, consider the points $P_q = (\|\alpha_{N,1}q\|, \ldots, \|\alpha_{N,d}q\|) \in [0,1]^d$. Partition $[0,1]^d$ into hypercubes of side length at most ε^2. There are at approximately ε^{-2d} such hypercubes, so by the pigeonhole principle, there must be some P_{q_1}, P_{q_2} that lie in the same hypercube with $Q > q_2 > q_1$. Then, with $q := q_2 - q_1 \leq Q$, we have that $\|\alpha_{N,j}q\| < \varepsilon^2$ for each j. Then, we have $|\sin(\pi q \alpha_{N,j})| \leq \|q\alpha_{N,j}\| \leq \varepsilon^2$, so for $x \in I_{N,j}$, we have $|\sin(\pi q x)| \leq 2\varepsilon^2$.

Now, let $I_N = \bigcup_j I_{N,j}$. Then, we have

$$\begin{aligned}\int_{I_N} |f_N(x)|^4 \sin(\pi q x)^2 dx &\leq |I_N|\, N^4 (2\varepsilon^2)^2 \leq D\left(\frac{2\varepsilon}{4\pi N}\right) N^4 4\varepsilon^4 \\ &= D(2\pi^{-1})\varepsilon^5 N^3 \leq \left(\frac{16\pi}{\varepsilon^3}\right)(2\pi^{-1})\varepsilon^3 N^3 = 32\varepsilon^2 N^3\end{aligned}$$

On the other hand, when $x \notin I_N$, by construction we have $|f_N(x)| < 2\varepsilon N$. Thus, we have:

$$\int_{[0,1]-I_N} |f_N(x)|^4 \sin(\pi q x)^2 dx \leq \sup_{x \notin I_N} |f_N(x)|^2 \int_{[0,1]} |f_N(x)|^2 \, dx \leq 4\varepsilon^2 N^2 \,|H_N| \leq 4\varepsilon^2 N^3$$

Thus,

$$\int_0^1 |f_N(x)|^4 \sin(\pi q x)^2 dx \leq 32\varepsilon^2 N^3 + 4\varepsilon^2 N^3 = 36\varepsilon^2 N^3$$

Thus far, we have shown that for all $\varepsilon > 0$ and $N > 0$, there exists q such that $\int_0^1 |f_N(x)|^4 \sin(\pi q x)^2 dx < \varepsilon N^3$

Let $H_N = H \cap [N]$. We now express this integral in terms of the R_H function from before. We first note that

$$|f_N(x)|^4 = \sum_{a,b,c,d \in H_N} e(((a+b)-(c+d)x)$$

so in particular, we have:

$$\begin{aligned}
\int_0^1 |f_N(x)|^4 \sin(\pi q x)^2 dx &= \frac{1}{4}\int_0^1 \left(\sum_{a,b,c,d\in H_N} e(((a+b)-(c+d)x)\right)(2-e(qx)-e(-qx)) \\
&= \frac{1}{4}\left(\sum_{\substack{a,b,c,d\in H_N \\ a+b=c+d}} 2 - \sum_{\substack{a,b,c,d\in H_N \\ a+b=c+d+q}} 2\right) \\
&= \frac{1}{2}\left(\sum_k R_{H_N}(k)^2 - \sum_k R_{H_N}(k)R_{H_N}(k-q)\right) \\
&= \frac{1}{4}\left(\sum_k R_{H_N}(k)^2 - 2\sum_k R_{H_N}(k)R_{H_N}(k-q) + \sum_k R_{H_N}(k)^2\right) \\
&= \frac{1}{4}\left(\sum_k R_{H_N}(k)^2 - 2\sum_k R_{H_N}(k)R_{H_N}(k-q) + \sum_k R_{H_N}(k-q)^2\right) \\
&= \frac{1}{4}\sum_k \left(R_{H_N}(k) - R_{H_N}(k-q)\right)^2
\end{aligned}$$

where the third line comes from parameterizing k over all possible values of $a+b$, and the penultimate line comes from translation invariance of the sum (it can be viewed as a sum over all k in $\mathbb{Z}$).

Next, let $F_{m,q} = \{d \in H_m \mid d - q \notin H\}$.

Then, we have:

$$\begin{aligned}
\frac{1}{4}\sum_k \left(R_{H_N}(k) - R_{H_N}(k-q)\right)^2 &\geq \frac{1}{4}\sum_{k\in F_{N,q}} \left(R_{H_N}(k) - R_{H_N}(k-q)\right)^2 \\
&= \frac{1}{4}\sum_{k\in F_{N,q}} \left(R_{H_N}(k-q)\right)^2 \\
&= \frac{1}{4}\sum_{k\in F_{N,q}} \left(R_{H}(k-q)\right)^2
\end{aligned}$$

Then, when $k > q$, and $k - q \notin H$, we have $R_H(k-q) > c(k-q)$, so we get:

$$\begin{aligned}
\frac{1}{4}\sum_{k\in F_{N,q}} (R_H(k-q))^2 &\geq \frac{1}{4}\sum_{\substack{k\in F_{N,q}\\ k>q}} (R_H(k-q))^2 \geq \sum_{\substack{k\in F_{N,q}\\ k>q}} (c(k-q))^2 \\
&= \sum_{k\in F_{N,q}-[q]} c^2(k-q)^2 \geq \sum_{k=q}^{q+\left|F_{N,q}-[q]\right|} c^2(k-q)^2 \\
&= \sum_{k=0}^{\left|F_{N,q}-[q]\right|} c^2k^2 \geq c'\left|F_{N,q}-[q]\right|^3 \geq c'\left(|F_{N,q}|-q\right)^3 \geq c'\left(|F_{N,q}|-Q\right)^3
\end{aligned}$$

for some c' dependent on c (but not on any other variable).

We now conclude the proof. In particular, fix $\varepsilon > 0$, and pick Q as above. Then, let $M = 2Q/\varepsilon$. Let $N > M$. Then, we may pick $q \leq Q$ such that $\int |f_N(x)|^4 \sin(\pi q x)^2 dx < \frac{c'\varepsilon^3}{8}N^3$, and thus $\frac{1}{4}\sum_k \left(R_{H_N}(k) - R_{H_N}(k-q)\right)^2 < \frac{c'\varepsilon^3}{8}N^3$. On the other hand, however, we have

$$\frac{1}{4}\sum_k \left(R_{H_N}(k) - R_{H_N}(k-q)\right)^2 \geq c'\left(|F_{N,q}|-Q\right)^3$$

so combining these results gives $c'\left(|F_{N,q}|-Q\right)^3 < \frac{c'\varepsilon^3}{8}N^3$ and thus $|F_{N,q}| < \frac{\varepsilon N}{2} + Q$. Since $N > M > \frac{2Q}{\varepsilon}$, we have $\frac{\varepsilon N}{2} + Q < \varepsilon N$, so $|F_{N,q}| < \varepsilon N$ as desired. □

We may now complete the proof that H is regular:

Proof. Fix $\varepsilon < c^2/10$, and pick M, Q from the above lemma. Then, fix some $N > 100Q$, and let q be the almost-period obtained by the above lemma (so q depends on N, but is bounded by Q).

We begin by showing that for $d \in \mathbb{N}$ with $\frac{c}{10}N < d < N$, we have $d \in H$ implies $d + q \in H$. In particular, suppose for contradiction that such a d satisfied $d \in H$ but $d + q \notin H$. Then, we have $r := R_H(d+q) > c(d+q) > cd > \frac{c^2}{10}N > \varepsilon N$.

From the definition of r, there exists $a_1, \ldots, a_r \in H$ with $d + q - a_1 \in H$ and $a_1 \leq d + q - a_1$. Since $r > \varepsilon N$, and from our choice of q we have $|\{k \leq N \mid k \in H, k - q \notin H\}| < \varepsilon N$, so it follows that at least one of the a_i must satisfy $a_i - q \in H$. However, we have $(a_i - q) + (d + q - a_i) = d$ with $a_i - q, d + q - a_i, d \in H$, contradicting the fact that H is sum-free.

By a similar reasoning, if $d \in \mathbb{N}$ satisfies $\frac{c}{10}N < d < N$, then $d \notin H$ implies $d + q \notin H$, so in fact H is regular with period q within the interval $\left[\frac{c}{10}N, N\right]$. In particular, let $S \subseteq \mathbb{Z}/q$ be the subset of resudues mod q of $H \cap \left[\frac{c}{10}N, N\right]$. By definition, for $d \in \left[\frac{c}{10}N, N\right] \cap \mathbb{N}$, we have $d \in H$ if and only if $\overline{d} \in S$.

We first claim that S is complete in the sense that $S \sqcup (S+S) = \mathbb{Z}/N\mathbb{Z}$. Since S clearly must be sum-free, it suffices to show that $\mathbb{Z}/N\mathbb{Z} - S \subseteq S + S$. In particular, suppose there exists $a \notin S$ with $a \notin S + S$. Then, we have that $R_H(\ell q + a) > c(\ell q + a)$ for all ℓ such that $cN/10 < (\ell q + a) < N$ (of which there are at least 90 since $N > 100Q$). However, if we pick large enough ℓ (still with $(\ell q + a) < N$), then any representation of $\ell q + a$ as a sum of two numbers in H cannot have both terms belong to the interval $\left[\frac{c}{10}N, N\right]$ (because then we would have $a \in S + S$), so we have $R_H(q\ell + a) < \frac{c}{10}N$, contradicting $R_H(\ell q + a) > c(\ell q + a)$. (because ℓ is chosen to make $\ell q + a$ close to N).

We now claim that for all $n > N$, we have $n \in H$ if and only if $\overline{n} \in S$. We prove this by induction. Indeed, suppose it has been shown for all integers between N and $n - 1$. Then, if $\overline{n} \notin S$, we have $\overline{n} \in S + S$, so we may write $\overline{n} = \overline{a} + \overline{b}$ for $0 \leq a, b < q$. Now note that $n = (\ell_1 q + a) + (\ell_2 q + b)$, where we may pick ℓ_1, ℓ_2 such that $\ell_1 q + a, \ell_2 q + b \in \left[\frac{c}{10}N, n-1\right]$ (because $n > N$ is sufficiently large). Thus, as $\overline{a}, \overline{b} \in S$, we have $\ell_1 q + a, \ell_2 q + b \in H$, and thus $n \notin H$. Now suppose Now suppose $\overline{n} \in S$, and suppose for contradiction that $n \notin H$. Then $R_H(n) > cn$. As $\overline{n} \notin S + S$, it follows that any representation of n as a sum of two elements in H cannot have its terms both belong to $\left[\frac{c}{10}N, n\right]$, so there are at most $\frac{c}{10}N$ such representations, contradicting $R_H(n) > cn$. Thus $n \in H$, and it follows that the period of q persists past N. □

Additionally, there is a variant of Theorem 5.2 for sum-free subsets of $\mathbb{Z}/N\mathbb{Z}$. Morally, Theorem 5.2 says that if R_H grows is large enough outside of H, then H must have a period. To generalize to $\mathbb{Z}/N\mathbb{Z}$, we shall say that a subset $A \subseteq \mathbb{Z}/N\mathbb{Z}$ has period q if $a \in A \iff a + q \in A$ holds for all $a \in \mathbb{Z}/N\mathbb{Z}$. Clearly N is a period of any such subset. The appropriate analogue of Theorem 5.2 is below, and its proof is remarkably similar.

Theorem 5.4. *Let $\varepsilon > 0$ and suppose N is sufficiently large. Then, if $A \subseteq \mathbb{Z}/N\mathbb{Z}$ is sum-free such that $r_A(k) > \varepsilon N$ for all $k \notin A$, then A has a period $q < N$.*

Proof. The proof is exceedingly similar to that of the lemma above.

In particular, suppose $N > \varepsilon^{-4\varepsilon^{-4}} + 2$.

Let $f(x) = \sum_{a \in a} e\left(\frac{ax}{N}\right)$ for $x \in \mathbb{Z}/N\mathbb{Z}$. Then, making use of the fact that for $a \in \mathbb{Z}/N\mathbb{Z}$,

$$\sum_{x \in \mathbb{Z}/N\mathbb{Z}} e\left(\frac{ax}{N}\right) = \begin{cases} N & a = 0 \\ 0 & a \neq 0 \end{cases}$$

we have that an identical computation to as in the lemma above gives us :

$$\sum_{x \in \mathbb{Z}/N\mathbb{Z}} |f(x)|^4 \sin\left(\pi q \frac{x}{N}\right)^2 = \frac{N}{4} \sum_{k \in \mathbb{Z}/N\mathbb{Z}} (r_A(k) - r_A(k-q))^2$$

We now note that $r_A(k) = \sum_{x\in\mathbb{Z}/N\mathbb{Z}} 1_S(x)1_S(k-x)$ and likewise, for any q, we have $r_S(k-q) = \sum_{x\in\mathbb{Z}/N\mathbb{Z}} 1_S(x-q)1_S(k-x)$, so

Suppose now that $\alpha_1, \ldots, \alpha_d$ are the elements of $\mathbb{Z}/N\mathbb{Z}$ for which $|f(\alpha_i)| > \varepsilon^2 N$. Then, we have:

$$d\varepsilon^4 N^2 < \sum_{i=1}^{d} |f(\alpha_i)|^2 \leq \sum_{x\in\mathbb{Z}/N\mathbb{Z}} |f(x)|^2 = N\,|A| \leq N^2$$

so we have $d < \varepsilon^{-4}$, which importantly does not depend on N. As before, we let $\|x\|$ denote the distance between x and the nearest integer.

For $x \in \mathbb{Z}/N\mathbb{Z}$ with $x' \in \{0, \ldots, N-1\}$ chosen such that $x \equiv x' \pmod N$, we now define $\|x\| = \min(x', N-x')$.

Consider the points $P_q = (\left\{\frac{q\alpha_1}{N}\right\}, \ldots, \left\{\frac{q\alpha_d}{N}\right\}) \in [0,1]^d$ for $q = 1, 2, \ldots, \varepsilon^{-4d}+2$. Since $N > \varepsilon^{-4\varepsilon^{-4}} + 2 > q$, we have that each of these possibilities of q are distinct mod N. Then, we divide $[0,1]^d$ into cubes of side length ε^4, of which there are approximately ε^{-4d}, so by the pigeonhole principle, there exists $q_1 < q_2 < N$ such that for $q = q_2 - q_1$, we have $\left\|\frac{q\alpha_i}{N}\right\| < \varepsilon^4$ for all i. In particular, we have $\left|\sin\left(\frac{\pi q\alpha_i}{N}\right)\right| < \varepsilon^4$. Then, with $B = \{\alpha_1, \ldots, \alpha_d\}$ we have:

$$\begin{aligned}\sum_{x\in\mathbb{Z}/N\mathbb{Z}} |f(x)|^4 \sin\left(\frac{\pi q x}{N}\right)^2 &= \sum_{x\in B} |f(x)|^4 \sin\left(\frac{\pi q x}{N}\right)^2 + \sum_{x\in\mathbb{Z}/N\mathbb{Z}-B} |f(x)|^4 \sin\left(\frac{\pi q x}{N}\right)^2 \\ &\leq d\,|S|^4\,\varepsilon^8 + \sup_{x\notin B} |f(x)|^2 \sum_{x\in\mathbb{Z}/N\mathbb{Z}} |f(x)|^2 \\ &\leq \varepsilon^4 N^4 + (\varepsilon^2 N)^2 N\,|S| \\ &\leq 2\varepsilon^4 N^4\end{aligned}$$

We now define $S\,(q) = |\{k \in A \mid k+q \notin A\}|$. We note that

$$\begin{aligned}2\varepsilon^4 N^4 &\geq \sum_{x\in\mathbb{Z}/N\mathbb{Z}} |f(x)|^4 \sin\left(\pi q \frac{x}{N}\right)^2 \\ &= \frac{N}{4} \sum_{k\in\mathbb{Z}/N\mathbb{Z}} (r_A(k) - r_A(k-q))^2 \\ &\geq \frac{N}{4} \sum_{\substack{k\in\mathbb{Z}/N\mathbb{Z} \\ k\in A \\ k+q\notin A}} (r_A(k) - r_A(k-q))^2\end{aligned}$$

$$= \frac{N}{4} \sum_{\substack{k \in \mathbb{Z}/N\mathbb{Z} \\ k \in A \\ k+q \notin A}} (r_A(k-q))^2$$
$$\geq \frac{N}{4} \sum_{\substack{k \in \mathbb{Z}/N\mathbb{Z} \\ k \in A \\ k+q \notin A}} \varepsilon^2 N^2$$
$$= \frac{1}{4} S(q)\, \varepsilon^2 N^3$$

so we have $S(q) \leq 8\varepsilon^2 N < \varepsilon N$ for small ε.

We now show that $S(q) < \varepsilon N$ implies $S(q) = 0$. In particular, suppose $k \notin A$. Then, we have $\ell := r_A(k) > \varepsilon N > S(q)$. We write $k = a_1 + b + 1 = \cdots = a_\ell + b_\ell$ for $a_i, b_i \in A$ with distinct a_i. Then, since there are fewer than ℓ elements x in $\mathbb{Z}/N\mathbb{Z}$ such that $x \in A$ and $x + q \notin A$, it follows that at least one of the ℓ elements $a_1 + q, \ldots, a_\ell + q$ belongs to A, say $a_i + q \in A$. Then $k + q = (a_i + q) + b_i$ is a sum of two elements of A, so since A is sum-free, we have $k + q \notin A$. Thus $k \notin A \implies k + q \notin A$, and a symmetric argument shows the reverse implication (alternatively, as A is finite, one direction suffices, since we necessarily have $|\{k \in A \mid k + q \notin A\}| = |\{k \in A \mid k - q \notin A\}|$) and so q is a period and by construction $q < N$. □

6. Acknowledgements

This research was conducted during the Stanford Undergraduate Research In Mathematics (SURIM) summer program. The authors would like to thank Professor Chris Ohrt and the SURIM program for its support throughout the research process. Additionally, the authors would like to thank our project mentor Rodrigo Sanches Angelo who came up with the project, provided invaluable assistance throughout, and laid the groundwork for a significant portion of our results.

References

Neil J. Calkin. On the structure of a random sum-free set of positive integers. *Discrete Math.*, 190(1-3):247–257, 1998. ISSN 0012-365X. doi: 10.1016/S0012-365X(98)00021-1. URL https://doi.org/10.1016/S0012-365X(98)00021-1.

Neil J. Calkin and Paul Erdős. On a class of aperiodic sum-free sets. *Math. Proc. Cambridge Philos. Soc.*, 120(1):1–5, 1996. ISSN 0305-0041. doi: 10.1017/S0305004100074600. URL https://doi.org/10.1017/S0305004100074600.

Neil J. Calkin and Steven R. Finch. Conditions on periodicity for sum-free sets. *Experiment. Math.*, 5(2):131–137, 1996. ISSN 1058-6458. URL http://projecteuclid.org/euclid.em/1047565644.

P. J. Cameron. Portrait of a typical sum-free set. In *Surveys in combinatorics 1987 (New Cross, 1987)*, volume 123 of *London Math. Soc. Lecture Note Ser.*, pages 13–42. Cambridge Univ. Press, Cambridge, 1987.

Ishay Haviv and Dan Levy. Symmetric complete sum-free sets in cyclic groups. *Israel J. Math.*, 227(2):931–956, 2018. ISSN 0021-2172. doi: 10.1007/s11856-018-1754-5. URL https://doi.org/10.1007/s11856-018-1754-5.

Stefan Steinerberger. A hidden signal in the Ulam sequence. *Exp. Math.*, 26(4):460–467, 2017. ISSN 1058-6458. doi: 10.1080/10586458.2016.1204638. URL https://doi.org/10.1080/10586458.2016.1204638.

Hermann Weyl. Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.*, 77(3):313–352, 1916. ISSN 0025-5831. doi: 10.1007/BF01475864. URL https://doi.org/10.1007/BF01475864.