CARDINALITIES OF g-DIFFERENCE SETS

Eric Schmutz
Department of Mathematics, Drexel University, Philadelphia, Pennsylvania
schmutze@drexel.edu

Michael Tait[1]
Department of Mathematics and Statistics, Villanova University, Villanova, Pennsylvania
michael.tait@villanova.edu

Received: 1/20/25, Revised: 2/27/25, Accepted: 8/5/25, Published: 8/15/25

Abstract

Let $\eta_g(n)$ be the smallest cardinality that $A \subseteq \mathbb{Z}$ can have if A is a g-difference basis for $[n]$, i.e, if, for each $x \in [n]$, there are *at least* g solutions to $a_1 - a_2 = x$. We prove that the finite, non-zero limit $\lim\limits_{n\to\infty} \frac{\eta_g(n)}{\sqrt{n}}$ exists, answering a question of Kravitz. We also investigate a similar problem in the setting of a vector space over a finite field. Let $\alpha_g(n)$ be the largest cardinality that $A \subseteq [n]$ can have if, for all nonzero x, $a_1 - a_2 = x$ has *at most* g solutions. We also prove that $\alpha_g(n) = \sqrt{gn}(1 + o_g(1))$ as $n \to \infty$.

1. Introduction

Suppose n and g are positive integers. We say that A *is a g-difference basis for* $[n]$ if A is a set of integers and, for each $x \in \{1, 2, 3 \ldots, n\}$, there are at least g solutions $(a_1, a_2) \in A \times A$ to the equation $a_1 - a_2 = x$. When $g = 1$ we will use "difference basis" to mean 1-difference basis. It is a natural question to ask what the minimum size of a g-difference basis for $[n]$ is. More generally, for a subset A of an abelian group G, one can define the representation function $r_{A-A} : G \to \mathbb{Z}$ by

$$r_{A-A}(d) = |\{(a, a') \in A \times A : d = a - a'\}|.$$

For g a natural number and $S \subseteq G$, define

$$\eta_g(S) = \min\{|A| : A \subseteq G, r_{A-A}(x) \geq g \text{ for all } x \in S\},$$

DOI: 10.5281/zenodo.16881835

[1]Research partially supported by NSF grant DMS-2245556.

and

$$\alpha_g(S) = \max\{|A| : A \subseteq S, r_{A-A}(x) \le g \text{ for all } x \in G, x \neq 0\}.$$

When $G = \mathbb{Z}$ and $S = \{1, \cdots, n\}$, we will use the notation $\eta_g(n)$ and $\alpha_g(n)$ to mean $\eta_g(\{1, \cdots, n\})$ and $\alpha_g(\{1, \cdots, n\})$, respectively, following the notation of [21] and [34].

In [21], Kravitz asserts that the main theorem in his paper can be strengthened, provided one can verify that the limit $\lim_{n\to\infty} \frac{\eta_g(n)}{\sqrt{n}}$ exists and is a finite non-zero real number. He asks whether this limit exists in Question 7.4 of [21] and writes that "it seems very likely" [that it does]. For the special case $g = 1$, Rédei and Rényi [28] proved that the limit does exist. Leonid Mirsky's helpful MathSciNet review [23] outlines the proof. We show that this proof can be modified in such a way that it works for any g. Hence we verify Kravitz's conjecture by proving the following theorem.

Theorem 1. *The limit* $\lim_{n\to\infty} \frac{\eta_g(n)}{\sqrt{n}}$ *exists and is a positive real number.*

It is also interesting to investigate $\eta_g(G)$ and $\alpha_g(G)$ for groups G other than the integers. Bounds in one direction for each function come from counting, but finding constructions seems highly dependent on the specific group. In particular, it is not known if there exists an $\epsilon > 0$ such that $\alpha_1(G) > \epsilon\sqrt{|G|}$ for any finite abelian group G (see the discussion after Problem 31 in [19]). In this paper we consider the case that $G = \mathbb{F}_p^n$. In this setting it seems that the parity of n plays a role, and we can only show that the "even limit" and "odd limit" each exist. We do this by proving the next theorem.

Theorem 2. *Fix an odd prime p and a natural number g. Then the limits*

$$L_e = \lim_{k\to\infty} \frac{\eta_g\left(\mathbb{F}_p^{2k}\right)}{\sqrt{p^{2k}}}$$

and

$$L_o = \lim_{k\to\infty} \frac{\eta_g\left(\mathbb{F}_p^{2k+1}\right)}{\sqrt{p^{2k+1}}}$$

both exist and are positive real numbers.

Contrary to the situation in the integers, we conjecture that these limits are not the same.

Conjecture 1. For L_e and L_o defined in Theorem 2,

$$L_e \neq L_o.$$

The densest constructions of Sidon sets in $\mathbb{F}_p^n$ that are currently known have a difference set that covers most elements of the group when n is even but not when n is odd. The best current constructions for n odd look like constructions in dimension $n-1$ with a smaller order term number of elements added on. Therefore it seems reasonable to guess that $L_o > L_e$.

Next we turn our attention to bounding the number of representations from above. From Corollary 1.4 in [34] (and the paragraph above it) one can easily deduce that that $\alpha_g(n) = \Theta\left(\sqrt{gn}\right)$ and that

$$\lim_{g\to\infty} \liminf_{n\to\infty} \frac{\alpha_g(n)}{\sqrt{gn}} = \lim_{g\to\infty} \limsup_{n\to\infty} \frac{\alpha_g(n)}{\sqrt{gn}}.$$

The results from [34] do not determine whether or not

$$\liminf_{n\to\infty} \frac{\alpha_g(n)}{\sqrt{gn}} = \limsup_{n\to\infty} \frac{\alpha_g(n)}{\sqrt{gn}}.$$

We therefore strengthen the result in [34] by proving the following theorem.

Theorem 3. *Let g be a positive integer. Then $\alpha_g(n) = (1 + o_g(1))\sqrt{gn}$.*

In this theorem, the $o_g(1)$ term goes to 0 as n goes to infinity and the subscript g indicates that the rate of convergence may depend on g. In other words, there is a function $\delta(g, n)$ such that

- $\frac{\alpha_g(n)}{\sqrt{gn}} = 1 + \delta(g, n)$, and
- for every g, $\lim_{n\to\infty} \delta(g, n) = 0$.

After writing this manuscript, Carlos Trujillo told us that Theorem 3 had been previously proved in [8] using a different argument for the upper bound and essentially the same construction for the lower bound.

In Sections 2, 3, and 4, we prove Theorems 1, 2, and 3, respectively. In Section 5 we discuss similar problems, but for sums instead of differences, and then we give some applications of these problems in coding theory and cryptography.

2. Difference Bases of Integers: Proof of Theorem 1

2.1. Crude Bounds

Before proving Theorem 1, we collect some necessary lemmas. The first step is to verify the "trivial" lower bound for $\eta_g(n)$ that is mentioned in [21]. We actually use this bound, so the proof is written out explicitly in Lemma 1 below. More or less the same argument is outlined in [5] for $g = 1$. Brauer made a similar argument for "restricted" difference bases in [7].

Lemma 1. *For all $n > 1$, $\frac{\eta_g(n)}{\sqrt{n}} \geq \sqrt{2g}$.*

Proof. Suppose that D is a finite set, that $f : D \to Y$, and that S is a finite subset of Y. The preimages $f^{-1}(y)$, $y \in S$ are disjoint, therefore $|D| \geq \sum_{y \in S} |f^{-1}(y)|$. If $|f^{-1}(y)| \geq g$ for all $y \in S$, then

$$|D| \geq g|S|. \tag{1}$$

Now consider the special case where

- $Y = \mathbb{Z}_+$, and $S = [n]$,
- A_n is a g-difference basis for $[n]$ with $|A_n| = \eta_g(n)$,
- $D = \binom{A_n}{2}$ is the set of all $\binom{|A_n|}{2}$ two-element subsets of A_n, and
- the function f is defined by $f(\{a_1, a_2\}) = |a_1 - a_2|$.

Observe that

$$\eta_g(n)^2 = |A_n|^2 > |A_n| (|A_n| - 1) = 2|D|.$$

Applying Inequality (1), we get $\eta_g(n)^2 \geq 2gn$. Taking square roots, the lemma is proved. □

The next lemma constructs a 1-difference basis that is essentially the same as that in [23, 28]. We need to make the n-dependence explicit.

Lemma 2. *Let $k_n = \lceil \sqrt{n} \, \rceil$. The $2k_n$-element set*

$$B_n := \{1, 2, 3, \ldots k_n - 1\} \cup \{k_n, 2k_n, 3k_n, \ldots, k_n^2, (k_n + 1)k_n\}$$

is a 1-difference basis for $[n]$.

Proof. Suppose $x \in [n]$. We need to write x as the difference of two elements of B_n.
`Case 1:` $x < k_n$. Observe that $x = k_n - (k_n - x)$. In this case, both k_n and $(k_n - x)$ are elements of B_n.
`Case 2:` $x = k_n$. Observe that $k_n = 2k_n - k_n$. Both k_n and $2k_n$ are elements of B_n.
`Case 3:` $x > k_n$. Divide x by k_n using the division algorithm. If r and q are the remainder and quotient, respectively, then $x = k_n q + r$, where $0 \leq r < k_n$ and $q = \lfloor \frac{x}{k_n} \rfloor$. Therefore $x = k_n(q+1) - (k_n - r)$. In this case, $(k_n - r)$ is an element of B_n. To verify that $k_n(q + 1)$ is also in B_n, we need only check that $q + 1 \leq k_n + 1$. Since the floor function is non-decreasing, we have

$$q = \left\lfloor \frac{x}{k_n} \right\rfloor \leq \left\lfloor \frac{n}{k_n} \right\rfloor \leq \frac{n}{k_n} \leq \frac{n}{\sqrt{n}} \leq \lceil \sqrt{n} \rceil = k_n.$$

□

Now, notice that translation does not change the distances between the elements of a set of integers: $(a_1+t)-(a_2+t)=(a_1-a_2)$ for any integer t. Therefore, by taking the union of g translates of a 1-difference basis, we get a g-difference basis. The following observation makes a specific choice.

Observation 1. Suppose that S_n is a 1-difference basis for $[n]$. If we define $\mathcal{T}_{n,g} = \{t+b : 0 \le t < g \text{ and } b \in S_n\}$, then $\mathcal{T}_{n,g}$ is a g-difference basis for $[n]$.

Corollary 1. *Let n be a natural number. Then* $\frac{\eta_g(n)}{\sqrt{n}} \le 2(1+\frac{1}{\sqrt{n}})g$.

Proof. Recall the 1-difference basis B_n in Lemma 2. Apply Observation 1 with $S_n = B_n$. By Boole's inequality, and the definition of $\eta_g(n)$, we have $\eta_g(n) \le |\mathcal{T}_{n,g}| \le |B_n|g = 2k_n g$. Since $k_n = \lceil \sqrt{n}\, \rceil < \sqrt{n}+1$, it follows that $\eta_g(n) < 2(\sqrt{n}+1)g$. □

Corollary 2. *For all $n > 1$ and all $g \ge 1$, we have*

$$\eta_1(n) \le \eta_g(n) \le g\eta_1(n).$$

2.2. Proof Outline

From Lemma 1 and Corollary 1, we know that $\frac{\eta_g(n)}{\sqrt{n}}$ is bounded above and below by positive real numbers. Therefore the limits $L = \liminf_{n\to\infty} \frac{\eta_g(n)}{\sqrt{n}}$ and $U = \limsup_{n\to\infty} \frac{\eta_g(n)}{\sqrt{n}}$ are positive real numbers:

$$\sqrt{2g} \le L \le U \le 4g. \tag{2}$$

The goal is to prove that $L = U$. Assume that $L < U$, and look for a contradiction.

Say that a number is *undersized* if it is less than $L + \frac{U-L}{3}$, and say that a number is *oversized* if it is greater than $U - \frac{(U-L)}{3}$. No number can be both undersized and oversized. From the definition of lim sup, we know that $\frac{\eta_g(n)}{\sqrt{n}}$ is oversized for infinitely many n. On the other hand, using ideas from Rédei-Rényi [28], we can prove that $\frac{\eta_g(n)}{\sqrt{n}}$ is undersized for all sufficiently large n. This means that infinitely many numbers are both undersized and oversized, which is the sought after contradiction.

2.3. Adapting Rédei-Rényi

For any $\delta > 0$, we can (by the definition of L) choose a positive integer v such that

$$\frac{\eta_g(v)}{\sqrt{v}} < L + \frac{\delta}{10g}. \tag{3}$$

In particular, we'll use $\delta = \frac{U-L}{3}$. We claim it is possible choose N_0 sufficiently large so that the following two conditions are both satisfied when $n > N_0$:

- $\frac{\eta_g(v)}{\sqrt{n}} < \frac{\delta}{10g}$,
- there is a prime q_n such that

$$\sqrt{\frac{n}{v}} \leq q_n \leq (1+\frac{\delta}{10g})\sqrt{\frac{n}{v}}. \tag{4}$$

It is clear that we can satisfy the first condition: if $N_0 > (\frac{10g\eta_g(v)}{\delta})^2$ then by elementary algebra $\frac{\eta_g(v)}{\sqrt{n}} < \frac{\delta}{10g}$ is satisfied for all $n > N_0$. The second condition is highly non-trivial, but it follows easily from well known results on the distribution of prime numbers. For example, Baker, Harman and Pintz [1] proved that, for all sufficiently large real numbers x, there is at least one prime number in the interval $[x - x^{.525}, x]$. Choose $x_n = (1+\frac{\delta}{10g})\sqrt{\frac{n}{v}}$, and $c = (1+\frac{\delta}{10g})^{-1}$. The interval $[cx_n, x_n] = [\sqrt{\frac{n}{v}}, (1+\frac{\delta}{10g})\sqrt{\frac{n}{v}}]$ has length $(1-c)x_n$. Since $x_n^{.525} = o(x_n)$, it is clear that $[x_n - x_n^{.525}, x_n] \subseteq [cx_n, x_n]$ for all sufficiently large n. Therefore the second condition is satisfied.

Following [23, 28], we use Singer's theorem to prove Lemma 4 below. It is sometimes convenient to adopt the more compact notation $v^{*,g}$ for $\eta_g(v)$.

Lemma 3 ([30])**.** *If q is a prime number, and $m = q^2 + q + 1$, then there are $q+1$ integers $a_0, a_1, \ldots, a_q$ such that the $q^2 + q$ differences $a_i - a_j$ are congruent modulo m to the numbers $1, 2, 3, \ldots, q^2 + q$ in some order.*

Without loss of generality, assume that the numbers a_i, a_j in Singer's theorem are elements of $[m]$.

Lemma 4. *If $\mathcal{B} = \{b_1, b_2, \ldots, b_{v^{*,g}}\}$ is a g-difference basis for $[v]$, then the set $\mathcal{F} = \{a_i + mb_j : 1 \leq i \leq q, 1 \leq j \leq v^{*,g}\}$ forms a g-difference basis for $[mv]$.*

Proof. Suppose $s \in [mv]$. We need to verify that s has at least g representations as the difference of two elements of $\mathcal{F}$.

`Case 1:`$s = mv$. Because $\mathcal{B}$ is a g-difference basis for $[v]$, we have $v = b_i - b_j$ for at least g pairs i, j. But in the special case $s = mv$, this means that

$$s = (a_i + b_i m) - (a_i + b_j m) \in \mathcal{F} - \mathcal{F},$$

for each of these pairs i, j.

`Case 2:` $s < mv$. If we divide s by m using the division algorithm, we get $s = km + r$ where $0 \leq r < m$. Singer's theorem guarantees that we have either $r = a_h - a_\ell$ for some h, ℓ, or $r - m = a_h - a_\ell$ for some h, ℓ. These two subcases are considered separately.

`Subcase 2a:` $s < mv$ `and` $r = a_h - a_\ell$. Recall that $s = km + r$ for a non-negative r, and that and $s < vm$. It follows that $k < v$. Because $\mathcal{B}$ is a g-difference basis for $[v]$, we have $k = b_i - b_j$ for at least g pairs i, j. Plugging into $s = km + r$, we get

$$s = (b_i - b_j)m + (a_h - a_\ell) = (a_h + b_i m) - (a_\ell + b_j m) \in \mathcal{F} - \mathcal{F}.$$

`Subcase 2b:` $s < mv$ `and` $r - m = a_h - a_\ell$. Plugging $m - r = a_h - a_\ell$ into $s = km + r$, we get

$$s = (k+1)m + a_\ell - a_h.$$

Because $s < mv$, we also have $k < v$. Because $k < v$, and $\mathcal{B}$ is a g-difference basis for $[v]$, we have $k + 1 = b_i - b_j$ for at least g pairs i, j. Therefore

$$s = (a_\ell + b_i m) - (a_h + b_j m) \in \mathcal{F} - \mathcal{F}.$$

□

We now prove the following corollary of Lemma 4.

Corollary 3. *Let g and n be natural numbers. Then with q_n and v defined as above, we have $\eta_g(n) \leq q_n \eta_g(v)$.*

Proof. Recall that $m = q_n^2 + q_n + 1$, where q_n satisfies the lower bound in (4), namely $\sqrt{n/v} < q_n$. Hence $n < q_n^2 v < (q_n^2 + q_n + 1)v = mv$. From the definition of η_g, it follows that $n^{*,g} \leq (mv)^{*,g}$. By Lemma 4, $(mv)^{*,g} \leq |\mathcal{F}| = q_n v^{*,g} = q_n \eta_g(v)$. The corollary now follows. □

We now have all the information that is needed to prove Kravitz's conjecture.

Proof of Theorem 1. Recall the proof outline in Subsection 2.2, and our choice of $\delta = \frac{U-L}{3}$. All that remains is to show that $\frac{\eta_g(n)}{\sqrt{n}}$ is undersized for all sufficiently large n. By Corollary 3 we have $\eta_g(n) \leq q_n \eta_g(v)$, and by Condition 2, we have $q_n \leq (1 + \frac{\delta}{10g})\sqrt{\frac{n}{v}}$. Therefore

$$\frac{\eta_g(n)}{\sqrt{n}} \leq (1 + \frac{\delta}{10g})\frac{\eta_g(v)}{\sqrt{v}}.$$

Now use the inequality $\frac{\eta_g(v)}{\sqrt{v}} < L + \frac{\delta}{10g}$ from Inequality (3) to get

$$\frac{\eta_g(n)}{\sqrt{n}} \leq \left(1 + \frac{\delta}{10g}\right)\left(L + \frac{\delta}{10g}\right) = L + \delta\left(\frac{L}{10g} + \frac{1}{10g} + \frac{\delta}{(10g)^2}\right).$$

By Inequality (2) we have $L < 4g$ and $\delta < 4g$. Therefore

$$\frac{\eta_g(n)}{\sqrt{n}} < L + \delta\left(\frac{4}{10g} + \frac{1}{10g} + \frac{4}{(10g)^2}\right) < L + \delta = L + \frac{U-L}{3}.$$

This shows that $\frac{\eta_g(n)}{\sqrt{n}}$ is undersized, and completes the proof. □

3. Difference Bases in $\mathbb{F}_p^n$

In this section, we prove Theorem 2. We will need a (slightly easier) version of Lemma 4. The proof follows from the definitions.

Lemma 5. *Let G_1 and G_2 be abelian groups and $A_i \subseteq G_i$ be g_i-difference bases of G_i. Then $A_1 \times A_2$ is a $g_1 g_2$-difference basis of $G_1 \times G_2$.*

Note that if p is prime we may take a difference basis of the integers up to p and consider them as field elements, and this will constitute a difference basis of $\mathbb{F}_p$. By Lemma 2, we have

$$\eta_1(\mathbb{F}_p) \leq 2\lceil \sqrt{p} \rceil. \tag{5}$$

Lemma 6. *Let q be an odd prime power. Then*

$$\eta_1(\mathbb{F}_q \times \mathbb{F}_q) \leq q + \eta_1(\mathbb{F}_q).$$

Proof. Let B be a difference basis of $\mathbb{F}_q$ with $|B| = \eta_1(\mathbb{F}_q)$. Let $A = \{(x, x^2) : x \in \mathbb{F}_q\}$. Then we claim that the set

$$A \cup \{(0, b) : b \in B\}$$

is a difference basis for $\mathbb{F}_q \times \mathbb{F}_q$. To see this, for any $a \neq 0$ and $b \in \mathbb{F}_q$, it is easy to check that $(a, b) \in A - A$. In particular, if $x = \frac{1}{2}\left(a + ba^{-1}\right)$ and $y = x - a$, then we have

$$(x, x^2) - (y, y^2) = (a, b).$$

By definition of B, for any $b \in \mathbb{F}_q$ there are two elements $b_1, b_2 \in B$ such that

$$(0, b_1) - (0, b_2) = (0, b).$$

Hence any $(a, b) \in \mathbb{F}_q \times \mathbb{F}_q$ may be written as a difference of two elements in the set. Since $|A| = q$ and $|B| = \eta_1(\mathbb{F}_q)$, the lemma follows. □

Lemma 7. *Let p be prime and q a power of p. Then*

$$\eta_1(\mathbb{F}_q) \leq 100\sqrt{q}.$$

Proof. Let $q = p^n$. We will prove the lemma by induction on n. When $n = 1$, we may use Inequality (5). When $n = 2$, by Lemma 6 and Inequality (5), we have that

$$\eta_1(\mathbb{F}_{p^2}) = \eta_1(\mathbb{F}_p \times \mathbb{F}_p) \leq p + 2\lceil \sqrt{p} \rceil < 100p.$$

Note that the inequality is trivial if $p^{n/2} \leq 100$, so we may also assume $p^{n/2} \geq 100$. We will do the case that n is even and odd separately. First, by Lemma 6 and the inductive hypothesis, we have that

$$\eta_1(\mathbb{F}_{p^{2k}}) = \eta_1(\mathbb{F}_{p^k} \times \mathbb{F}_{p^k}) \leq p^k + \eta_1(\mathbb{F}_{p^k}) \leq p^k + 100p^{k/2} < 100p^k.$$

For n odd, we use Inequality (5) to let A be a difference basis of $\mathbb{F}_p$ of size at most $3\sqrt{p}$. As in the last case, let B be a difference basis of $\mathbb{F}_{p^{2k}}$ of size at most $p^k + 100p^{k/2}$. Then by Lemma 5, $A \times B$ is a difference basis of $\mathbb{F}_p \times \mathbb{F}_{p^{2k}} \cong \mathbb{F}_{p^{2k+1}}$. Hence

$$\eta_1(\mathbb{F}_{p^{2k+1}}) \leq |A||B| \leq 3\sqrt{p}(p^k + 100p^{k/2}) < 100p^{k+1/2}.$$

□

By Lemmas 6 and 7, we have the following lemma.

Lemma 8. *Let p be prime. Then*

$$\eta_1(\mathbb{F}_{p^t} \times \mathbb{F}_{p^t}) = (1 + o(1))p^t,$$

where the $o(1)$ term goes to 0 as t goes to infinity.

We are now ready to prove the main result of the section, Theorem 2.

Proof of Theorem 2. The proof is similar to that of Theorem 1. We write the details for the odd case and omit those for the even case which follows in the same way. Let

$$L = \liminf_{k\to\infty} \frac{\eta_g\left(\mathbb{F}_p^{2k+1}\right)}{\sqrt{p^{2k+1}}}.$$

Let $\epsilon > 0$ be arbitrary and assume it is less than 1. We will show that for k large enough, we have

$$\frac{\eta_g\left(\mathbb{F}_p^{2k+1}\right)}{\sqrt{p^{2k+1}}} < L + \epsilon.$$

Choose any $\epsilon_1 > 0$ satisfying $\epsilon_1 + L\epsilon_1 + \epsilon_1^2 < \epsilon$. By definition of L, there is some k_0 such that $\eta_g\left(\mathbb{F}_{p^{2k_0+1}}\right) \leq (L + \epsilon_1)\sqrt{p^{2k_0+1}}$. By Lemma 8, there is an t_0 such that for all $t \geq t_0$ we have that

$$\eta_1\left(\mathbb{F}_{p^t} \times \mathbb{F}_{p^t}\right) \leq (1 + \epsilon_1)p^t.$$

Let $n \geq 2t_0 + 2k_0 + 1$ be odd. Let A be a g-difference basis of $\mathbb{F}_{p^{2k_0+1}}$ of size at most $(L + \epsilon_1)\sqrt{p^{2k_0+1}}$. Since $n - (2k_0 + 1) \geq 2t_0$, there is a difference basis B of $\mathbb{F}_{p^{n-(2k_0+1)}}$ of size at most $(1 + \epsilon_1)p^{n/2-k_0-1/2}$. By Lemma 5, we have that $A \times B$ is a g-difference basis of $\mathbb{F}_{p^n}$ and therefore

$$\eta_g(\mathbb{F}_{p^n}) \leq |A||B| \leq (L + \epsilon_1)\sqrt{p^{2k_0+1}}(1 + \epsilon_1)p^{n/2-k_0-1/2} < (1 + \epsilon)p^{n/2}.$$

□

4. Bounded Difference Representations

In this section, we prove Theorem 3. We first prove the upper bound

$$\alpha_g(n) \le (1 + o_g(1))\sqrt{gn}$$

in Proposition 1 and then the lower bound in Proposition 2. The upper bound essentially follows from an argument of Lindström [22], and the lower bound is essentially the same construction as in [8]. The upper bound also was proved in [8] in a different way. As noted by a referee, the upper bound could also be proved in a third way using a method of Cilleruelo (see [2] Section 3). If one wanted to understand the error term better, it might be possible to do so by analyzing the multiple proofs and seeing that they cannot be tight at the same time, as in [2].

4.1. Upper Bound for $\alpha_g(n)$

Proposition 1. *Fix a positive integer g. Then $\alpha_g(n) \le (1 + o(1))\sqrt{gn}$.*

Proof. Let $A = \{a_1, \cdots, a_k\}$ be a subset of $[n]$ with $r_{A-A}(x) \le g$ for all nonzero x. Without loss of generality assume that $a_1 < a_2 < \cdots < a_k$. A bound on k will be deduced by examining the constraints that the definition of A imposes on the sum

$$\sigma_\ell :\stackrel{\text{def}}{=} \sum_{t=1}^{\ell} \sum_{i=t+1}^{k} (a_i - a_{i-t}), \tag{6}$$

for an appropriately chosen ℓ. Let s be the number of terms $(a_i - a_{i-t})$ in the sum σ_ℓ. Note for future reference that, provided $\ell < k$, the inner sum has $k - t$ terms and

$$s = \sum_{t=1}^{\ell}(k - t) = k\ell - \binom{\ell+1}{2}. \tag{7}$$

As in [22], a simple upper bound for σ_ℓ can by obtained by exploiting cancellation in the inner sum on the right side of Equation (6). To simplify the calculation, impose the further restriction that $\ell \le \frac{k-1}{2}$. Since $1 \le t \le \ell \le \frac{k-1}{2}$, we have $k - t > t + 1$ and

$$\sum_{i=t+1}^{k} (a_i - a_{i-t}) \le a_k + a_{k-1} + \cdots + a_{k-t+1}. \tag{8}$$

Because $A \subseteq [n]$, we know that, on the right side of Inequality (8), each of the t terms is at most n. Therefore

$$\sigma_\ell < \sum_{t=1}^{\ell} tn = \binom{\ell+1}{2}n. \tag{9}$$

To obtain a lower bound for σ_ℓ, we will arrange the s differences from smallest to largest, then use the inequality $r_{A-A}(x) \leq g$. If d_i denotes the i'th smallest of the s differences, then $\sigma_\ell = \sum_{i=1}^{s} d_i$ and $d_1 \leq d_2 \leq \cdots \leq d_s$. Using the standard notation $\lfloor x \rfloor$ for the greatest integer less than or equal to x, we will insert parentheses so that the sum is partitioned into groups of g consecutive differences as follows:

$$\sigma_\ell \geq \sum_{b=0}^{\lfloor \frac{s}{g} \rfloor - 1} \left(\sum_{r=1}^{g} d_{bg+r} \right) \tag{10}$$

$$= (d_1 + d_2 + \cdots d_g) + (d_{g+1} + d_{g+2} + \cdots + d_{2g}) + (d_{2g+1} + d_{g+} + \cdots d_{3g}) + \ldots \tag{11}$$

If s is a multiple of g, then the inequality in Inequality (10) is an equality. If s is not a multiple then, g, we have have a strict inequality because the largest $s - g\lfloor \frac{s}{g} \rfloor$ differences have been omitted. By the definition of A, at least one of the the first $g+1$ terms $d_1, d_2, \ldots, d_{g+1}$ has to be greater than one. Therefore, inside the second pair of parentheses, all g differences must be greater than or equal to 2. In general, the b^{th} pair of parentheses encloses g differences that are each greater than or equal to b. Combining this fact with Inequality (9), we get

$$n\binom{\ell+1}{2} > \sigma_\ell \geq \sum_{b=1}^{\lfloor \frac{s}{g} \rfloor - 1} bg = \binom{\lfloor \frac{s}{g} \rfloor}{2} g \geq (\frac{s}{g} - 1)^2 \frac{g}{2}. \tag{12}$$

Provided $s > g$, we can solve for s and get

$$s < \sqrt{gn}\sqrt{\ell(\ell+1)} + g. \tag{13}$$

Recall that $s = k\ell - \binom{\ell+1}{2}$. Plugging this into Inequality (13), and solving for k we get

$$k < \sqrt{gn}\sqrt{1 + \frac{1}{\ell}} + \frac{\ell+1}{2} + \frac{g}{\ell}. \tag{14}$$

For each n, we can choose a $k(n)$-element set $A = A_n$, where $k(n) = \alpha_g(n)$. For this choice of k, define $\ell(n) = \lfloor k^{1/2} \rfloor$. We may assume that $k \to \infty$ and hence $\ell(n) \to \infty$ as $n \to \infty$, otherwise the conclusion would be immediate. Using calculus, it is easy to check that $\sqrt{1+u} < 1 + u$ for all $u > 0$. In particular, for $u = \frac{1}{\ell}$ we have $\sqrt{1 + \frac{1}{\ell}} \leq 1 + \frac{1}{\ell}$ Combining this with Inequality (14), we get

$$\begin{aligned} \alpha_g(n) &\leq \sqrt{gn} + \frac{\sqrt{gn}}{\ell} + \frac{(\ell+1)}{2} + \frac{g}{\ell} \\ &= \sqrt{gn}\left(1 + \frac{1}{\ell} + \frac{(\ell+1)}{2\sqrt{gn}} + \frac{g}{\ell\sqrt{gn}}\right). \end{aligned}$$

Using the asymptotic notation that was specified at the the end of Section 1 (after Theorem 3), we have $\alpha_g(n) \leq \sqrt{gn}\,(1 + o_g(1))$ as $n \to \infty$. $\square$

4.2. Lower Bound for $\alpha_g(n)$

To prove the lower bound in Theorem 3, We will construct a set $A \subset [n]$ of size $(1-o(1))\sqrt{gn}$ that satisfies $r_{A-A}(x) \le g$ for all nonzero x. Our set is the same set as was constructed in [8, 31], where it was shown that $r_{A+A}(x) \le 2g$. We include the proof that $r_{A-A}(x) \le g$ for completeness, although it is essentially the same as in [8, 31]. We start with a Bose-Chowla Sidon set [6]. Let g be a fixed integer and q be a prime power such that $q \equiv 1 \pmod{g}$. Let θ generate the multiplicative group of $\mathbb{F}_{q^2}$ and let $\mathbb{F}_q$ denote the subfield of order q in $\mathbb{F}_{q^2}$. Define the set

$$B_q = \{a \in \mathbb{Z}/(q^2-1)\mathbb{Z} : \theta^a - \theta \in \mathbb{F}_q\}.$$

Bose and Chowla [6] showed that B_q is a Sidon set of size q in $\mathbb{Z}/(q^2-1)\mathbb{Z}$. Let H be the subgroup of $\mathbb{Z}/(q^2-1)\mathbb{Z}$ generated by $\frac{q^2-1}{g}$ and note that H is a subgroup of order g. We will use the following lemma which is already known (see Lemma 2.2 of [32], Lemma 2.1 of [16], and Lemma 3.1 of [31]).

Lemma 9. *([16, 31]) Let g be fixed, q be a prime power congruent to 1 modulo g and B_q and H be defined as above. Then*

$$(B_q - B_q) \cap H = \{0\}.$$

Proposition 2. *There exists a set $A \subset [n]$ with $|A| = (1-o(1))\sqrt{gn}$ and $r_{A-A}(x) \le g$ for all nonzero x.*

Proof. First we note that if we can find $A \subset \mathbb{Z}/N\mathbb{Z}$ of size $(1-o(1))\sqrt{gn}$ with $r_{A-A}(x) \le g$ for all nonzero x and $N \le n$, then a corresponding set of integer representatives will prove the theorem. As above, let q be a prime power congruent to 1 modulo g, let B_q be the Bose-Chowla Sidon set in $\mathbb{Z}/(q^2-1)\mathbb{Z}$ and let H be the subgroup of $\mathbb{Z}/(q^2-1)\mathbb{Z}$ of order g which is generated by $\frac{q^2-1}{g}$. Consider the quotient group $\Gamma := [\mathbb{Z}/(q^2-1)\mathbb{Z}]/H$ and define a subset $A_H \subset \Gamma$ by $A_H = \{a+H : a \in B_q\}$. First we show that A_H is the same size as B_q. If $a+H = b+H$ with $a, b \in B_q$, then $a-b \in H$ and so by Lemma 9 we have that $a \equiv b \pmod{q^2-1}$. Therefore $|A_H| = |B_q| = q$.

By the Siegel-Walfisz theorem we may, for all sufficiently large n, choose a prime $q = q(n)$ such that $q \equiv 1 \pmod{g}$ and $\sqrt{(1-\epsilon)gn+1} \le q \le \sqrt{gn+1}$. Note that $(1-\epsilon)n \le \frac{q^2-1}{g} \le n$. Define $N = \frac{q^2-1}{g}$ and note that Γ is isomorphic to the cyclic group $\mathbb{Z}/N\mathbb{Z}$. Let $A \subset \mathbb{Z}/N\mathbb{Z}$ be the set corresponding to $A_H \subset \Gamma$. So $|A| = |A_H| = q = \sqrt{gN+1} \sim \sqrt{gn}$. Hence we are done as long as we can show that $r_{A_H - A_H}(d+H) \le g$ for any $d+H$ not the identity. Consider solutions to the equation

$$(a+H) - (b+H) = d+H, \tag{15}$$

where $d+H$ is not the identity and $(a+H), (b+H) \in A_H$. Equation (15) implies that there is an $h \in H$ such that $a - b \equiv d + h \not\equiv 0 \pmod{q^2-1}$. Since $(a+H), (b+H) \in$

A_H, we have that $a, b \in B_q$. Since B_q is a Sidon set, for each h there is at most 1 ordered pair $(a, b) \in B_q \times B_q$ such that $a - b \equiv d + h \pmod{q^2 - 1}$. Since H has order g, we know that Equation (15) has at most g solutions, completing the proof. $\square$

5. Concluding Remarks

We end the paper by discussing what is known when one considers sums instead of differences and by sketching some of the applications of these problems to other areas.

5.1. Sums

Naturally, one may also consider sums instead of differences and a large body of work has been done on this topic. Define the representation function

$$r_{A+A}(s) = |\{(a, a') \in A \times A : s = a + a; \}|.$$

Similarly to before, for g a natural number and $S \subseteq G$, define

$$\nu_g(S) = \min\{|A| : A \subseteq G, r_{A+A}(x) \geq g \text{ for all } x \in S\},$$

and

$$\beta_g(S) = \max\{|A| : A \subseteq S, r_{A+A}(x) \leq g \text{ for all } x \in G\},$$

and use $\nu_g(n)$ and $\beta_g(n)$ when $S = \{1, \cdots, n\}$. A set A satisfying $r_{A-A}(x) \leq 1$ for all nonzero x is equivalent to it satisfying $r_{A+A}(x) \leq 2$ for all x, and such sets are called *Sidon sets*. Sidon sets have been the subject of intensive study for almost a century since being introduced in [29]; see the surveys of O'Bryant [26] and Plagne [27] for an extensive history. In particular, it is known that $\beta_2(n) \sim \sqrt{n}$ and determining the error term is a 500 dollar Erdős question, see the papers [2, 11] for the current state of the art. For larger g, there is no such equivalent characterization of r_{A+A} in terms of r_{A-A} and perhaps surprisingly working with sums seems to be harder than working with differences. Much work has been done on this topic, and there are many open and intriguing problems. For example, proving a version of Theorem 1 but for $\beta_g(n)$ was conjectured in [12] and many papers have been written giving bounds on the liminf and limsup (Section 1.1 of [12] contains references to much of the progress on $\beta_2(n)$ over the years). The function $\nu_2(n)$ has also been considered extensively, for example the paper [18] has been cited over 500 times.

While most of the previous work has been done in the integers, considering other groups is also an intriguing and difficult problem. Just as we do in this paper, the first natural case to consider is when $G = \mathbb{F}_p^n$, and here there are many unanswered questions. When $n = 1$, it is unknown even for a sequence of values of p what the asymptotics of $\beta_2(\mathbb{F}_p)$ or $\nu_2(\mathbb{F}_p)$ are. In [19], it is asked in the discussion after Problem 31 and in Problem 33, respectively, whether an asymptotic formula for $\beta_2(\mathbb{F}_p)$ or $\nu_2(\mathbb{F}_p)$ can be determined for some sequence of values of p.

For odd p fixed and n growing, more is known (see for example [24, 33]). Along the same lines as Theorem 2 and Conjecture 1, it seems that the parity of n plays an important role. In particular, it is known that for n even,

$$\beta_2\left(\mathbb{F}_p^n\right) \sim p^{n/2}.$$

For n odd, the best general bounds are roughly

$$\left(\frac{1}{\sqrt{p}} + o(1)\right) p^{n/2} \leq \beta_2\left(\mathbb{F}_p^n\right) \leq (1 + o(1))p^{n/2}.$$

When $p = 2$, something similar is known, and closing these bounds is of great interest because of the problem's relationship to coding theory and cryptography, which we explain in the next subsection. Note that when $p = 2$ we have that $A + A = A - A$ and so the four main parameters collapse to two. Furthermore, one has $x + x = 0$ for any x, and so the definitions must be revised to exclude these trivial solutions from consideration. Once this is done the best known bounds are

$$(1 + o(1))\, 2^{n/2} \leq \beta_2\left(\mathbb{F}_2^n\right) \leq (\sqrt{2} + o(1))2^{n/2}$$

for n even, and

$$\left(\frac{1}{\sqrt{2}} + o(1)\right) 2^{n/2} \leq \beta_2\left(\mathbb{F}_2^n\right) \leq (\sqrt{2} + o(1))2^{n/2}$$

for n odd. Even less is known about $\nu_2\left(\mathbb{F}_p^n\right)$ and it would be interesting to understand this function better. Next we discuss some applications of this problem.

5.2. Coding Theory and Cryptography

5.2.1. Binary Codes with Minimum Distance 5

An $[n, k, d]_q$ code is a subspace of $\mathbb{F}_q^n$ of dimension k with the property that any nonzero vector has at least d entries which are nonzero. Such a code is linear (that is, it is a subspace) and has distance d (in the Hamming metric). Understanding how large a (linear) code may be with a fixed minimum distance is one of the fundamental problems in coding theory. There is an equivalence between Sidon sets in $\mathbb{F}_2^n$ and binary linear codes of minimum distance 5. In short, given a Sidon set $A \subseteq \mathbb{F}_2^t$ containing the 0 vector, the matrix H whose columns are the vectors of A except for the 0 vector is a parity check matrix (a matrix such that its null space is the code) for an $[|A| - 1, |A| - 1 - t, 5]_2$ code. The equivalence also goes in the other direction: given a parity check matrix H for an $[n, k, 5]_2$ code, the columns form a Sidon set in $\mathbb{F}_2^{n-k}$ of size n. This equivalence is explained in detail in [13] and they use coding theory techniques to give the best known upper bounds on $\beta_2(\mathbb{F}_2^n)$.

5.2.2. Covering Codes of Radius 2

Given a code $\mathcal{C} \subseteq \mathbb{F}_p^n$, the *covering radius* R of the code is the maximum distance of any vector in $\mathbb{F}_p^n$ to (the closest codeword in) the code [17]. One can check using the definitions that a linear $[n, n-r]_q$ code (a subspace of $\mathbb{F}_q^n$ of dimension $n-r$) with parity check matrix H has covering radius at most R if and only if any column $b \in \mathbb{F}_q^r$ can be written as a linear combination of at most R columns of H.

If H is the parity check matrix of an $[N, N-n]_2$ code that has covering radius 2, then the set A which is all of the columns of H and the 0 vector has the property that $A + A = \mathbb{F}_2^n$. That is, A is an additive basis for $\mathbb{F}_2^n$. Going the other direction, if A is a set containing the 0 vector and $A + A$ is an additive basis for $\mathbb{F}_2^n$, then the matrix H which has columns all of the nonzero vectors in A will be a parity check matrix for a linear code with covering radius 2.

By this equivalence, understanding $\nu_2(\mathbb{F}_2^n) = \eta_2(\mathbb{F}_2^n)$ is equivalent to understanding the smallest possible linear binary codes of covering radius 2. For p odd the problems are not equivalent, but there are implications. Given A which is either an additive or difference basis of $\mathbb{F}_p^n$, the matrix with columns from A is a parity check matrix for a code of covering radius 2. Conversely, given H a parity check matrix for a code of covering radius 2, the set A which is given by all scalar multiples of columns of H will be both an additive basis and a difference basis for $\mathbb{F}_2^n$. The problem of determining the smallest possible linear code with covering radius 2 has been studied extensively in coding theory, see for example [3, 4, 14, 15] and references therein.

5.2.3. Almost Perfect Nonlinear Functions

A function $f : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is called *almost perfect nonlinear* if for any $a, b \in \mathbb{F}_2^n$ with $a \neq 0$, the equation $f(x+a) + f(x) = b$ has at most 2 solutions. Note that if x is a solution then $x + a$ is also a solution, so the 2 cannot be reduced to 1. A function over $\mathbb{F}_2^n$ is almost perfect nonlinear if and only if the set $\{(x, f(x)) : x \in \mathbb{F}_2^n\}$ is a Sidon set (see for example [9]). Almost perfect nonlinear functions have applications in cryptography [10, 25] and so are studied in this context extensively.

Acknowledgements. The authors thank I-Hwa Lee for introducing them and Lorenzo Ambrogi for inspiration. We also thank a thorough referee and Carlos Trujillo for bringing [8] to our attention.

References

[1] R. C. Baker, G. Harman, and J. Pintz, The difference between consecutive primes. II, *Proc. Lond. Math. Soc. (3)* **83** (3) (2001), 532-562.

[2] J. Balogh, Z. Füredi, and S. Roy, An upper bound on the size of Sidon sets, *Amer. Math. Monthly* **130** (5) (2023), 437-445.

[3] D. Bartoli, A. A. Davydov, M. Giulietti, S. Marcugini, and F. Pambianco, New Bounds for Linear Codes of Covering Radius 2, in *Lecture Notes in Comput. Sci. Volume 10495*, Springer, Princeton, NJ, 2017, 1-10.

[4] D. Bartoli, A. A. Davydov, M. Giulietti, S. Marcugini, and F. Pambianco, New bounds for linear codes of covering radii 2 and 3, *Cryptogr. Commun.* **11** (2019), 903-920.

[5] A. Bernshteyn and M. Tait, Improved lower bound for difference bases, *J. Number Theory* **205** (2019), 50-58.

[6] R. C. Bose and S. Chowla, Theorems in the additive theory of numbers, *Comment. Math. Helv.* **37** (1962/63), 141-147.

[7] A. Brauer, A problem of additive number theory and its application in electrical engineering, *J. Elisha Mitchell Sci. Soc.* **61** (1945), 55-66.

[8] Y. Caicedo, C. A. Martos, and C. A. Trujillo, g-Golomb Rulers, *Rev. Integr. Temas Mat.*, **33** (2) (2015), 161-172.

[9] C. Carlet, On APN exponents, characterizations of differentially uniform functions by the Walsh transform, and related cyclic-difference-set-like structures, *Des. Codes Cryptogr.* **87** (2019), 203-224.

[10] C. Carlet, *Boolean functions for cryptography and coding theory*, Cambridge University Press, New York, 2020.

[11] D. Carter, Z. Hunter, and K. O'Bryant, On the diameter of finite Sidon sets, preprint, `arXiv: 2310.20032`

[12] J. Cilleruelo, I. Ruzsa, and C. Vinuesa, Generalized Sidon sets, *Adv. Math.* **225** (5) (2010), 2786-2807.

[13] I. Czerwinski and A. Pott, Sidon sets, sum-free sets and linear codes, *Adv. Math.Commun.* **18** (2) (2024), 549-566.

[14] A. A. Davydov, Constructions and families of nonbinary linear codes with covering radius 2, *IEEE Trans. Inform. Theory* **45** (5) (1999), 1679-1686.

[15] A. A. Davydov and P. Ostergard, Linear codes with covering radius $R = 2, 3$ and codimension tR, *IEEE Trans. Inform. Theory* **47** (1) (2001), 416-421.

[16] D. F. Daza, C. A. Trujillo, and F. A. Benavides, Sidon sets and C_4-saturated graphs, preprint, `arXiv: 1810.05262`

[17] R. Graham and N. Sloane, On the covering radius of codes, *IEEE Trans. Inform. Theory* **31** (3) (1985), 385-401.

[18] R. L. Graham and N. J. A. Sloane, On additive bases and harmonious graphs, *SIAM J. Algebraic Discrete Methods* **1** (4) (1980), 382-404.

[19] B. Green, 100 open problems, *manuscript, available on request to Professor Green (accessed November 2024)* (2024).

[20] G. Johnston, M. Tait, and C. Timmons, Upper and lower bounds on the size of $B_k[g]$ sets, *Australas. J. Combin.* **83** (1) (2022), 129-140.

[21] N. Kravitz, Generalized difference sets and autocorrelation integrals, *Acta Arith.* **199** (2) (2021), 199-219.

[22] B. Lindström, An inequality for B_2-sequences, *J. Combin. Theory* **6** (2) (1969), 211-212.

[23] L. Mirsky, MathSciNet Review MR003055. Mirsky's review outlines Rédei and Rényi's proof in English.

[24] G. P. Nagy, Thin Sidon sets and the nonlinearity of vectorial boolean functions, preprint, arXiv: 2212.05887

[25] K. Nyberg, Perfect nonlinear S-boxes, in *Lecture Notes in Comput. Sci.,* **547**, Springer, Berlin 1991, pp. 378-386.

[26] K. O'Bryant, A complete annotated bibliography of work related to Sidon sequences, preprint, arXiv: 0407117

[27] A. Plagne, Recent progress on finite $B_h[g]$ sets, *Congr. Numer.* (2001), 49-64.

[28] L. Rédei and A. Rényi, On the representation of the numbers $1, 2, \cdots, N$ by means of differences, *Mat. Sbornik N.S.* **24/66** (1949), 385-389.

[29] S. Sidon, Ein satz über trigonometrische polynome und seine anwendung in der theorie der Fourier-reihen, *Math. Ann.* **106** (1) (1932), 536-539.

[30] J. Singer, A theorem in finite projective geometry and some applications to number theory, *Trans. Amer. Math. Soc.* **43** (3) (1938), 377-385.

[31] G. Johnston, M. Tait, and C. Timmons, Upper and lower bounds on the size of $B_k[g]$ sets, *Australas. J. Combin.* **83** (1) (2022), 129-140.

[32] M. Tait and C. Timmons, Sidon sets and graphs without 4-cycles, *J. Comb.* **5** (2) (2014), 155–165.

[33] M. Tait and R. Won, Improved bounds on sizes of generalized caps in $AG(n, q)$, *SIAM J. Discrete Math.* **35** (1) (2021), 521-531.

[34] W. Xu, Popular differences and generalized Sidon sets, *J. Number Theory* **186** (2018), 103-120.