Problems in Additive Number Theory

by

Brooke Orosz

A dissertation submitted to the Graduate Faculty in Mathematics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2009

This manuscript has been read and accepted for the Graduate Faculty in Mathematics in satisfaction of the dissertation requirements for the degree of Doctor of Philosophy.

Dr. Melvyn Nathanson

Date | Chair of Examining Committee

Dr. Józef Dodziuk

Date | Executive Officer

Dr. Joseph Malkevitch

Dr. Kevin O'Bryant

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Problems in Additive Number Theory

by

Brooke Orosz

Advisor: Professor Melvyn Nathanson

The first chapter deals with the following problem: Let $f(n)$ be a growth function, and A be a sequence with $f(n) \leq a_n \leq Uf(n)$, U constant. Under what conditions is it possible to construct a sequence B, with $b_k \sim \beta f(k)$, which has A as a subsequence?

The next two chapters deal with the possible sizes of $f(A) = \{\sum_{i=1}^{n} u_i a_i \;\Big|\; a_i \in A\}$ on different sets A, for certain forms f. The final chapter discusses counting relatively prime subsets of the natural numbers.

Acknowledgments

My parents, for their faith, my husband for his support, and my advisor and research group, for tempting my curiosity.

Contents

Chapter 1

Well-behaved sequences with worse-behaved subsequences

A basis of order h for the integers is a set of integers such that hA, the h-fold sum set, $(A + A + A + \cdots + A5$ h times) is $\mathbb{N}$. Since $|hA| < |A|^h$, then $|A \cap [1, n]| \geq n^{1/h}$ for any such basis. A basis of order h is called *thin* if $|A \cap [1, n]| \leq \alpha n^{1/h}$ for some constant α. An asymptotic basis for the integers is a set A such that for some N, hA contains all integers larger than N.

The first examples of thin bases were discovered by Raikov [15] and Stohr [17], and recent constructions are due to Blomer [1], Hofmeister [7], and Jia and Nathanson [8]. Cassels [2] constructed a beautiful family of asymptotic bases of order h such that $\lim_{n\to\infty} a_n/n^h = \alpha$.

This fact led Nathanson [10] to a study of strictly increasing sequences of integers asymptotic to various functions.

Definition: A *supersequence* of A is any sequence which has A as a subsequence.

Definition: A *growth function* is a function which is strictly increasing, unbounded, and continuous on $x \geq 1$.

Definition: A function is *asymptotically stable* if $\lim_{n\to\infty} \frac{f(n+\Delta)}{f(n)} = 1$ for any Δ. Polynomial functions are asymptotically stable, as is $e^{\sqrt{x}}$. Exponential functions are asymptotically unstable.

1.1 Necessary and sufficient conditions on a sequence

Theorem 1.1. *Let $f(n)$ be a growth function and let A be a sequence such that $f(n) \leq a_n \leq Uf(n)$, where U is constant. Then it is possible to find a supersequence B with $\lim_{k\to\infty} \frac{b_k}{f(k)} = \beta$ only if $\beta \leq \liminf_{n\to\infty} a_n/f(n)$ and A meets the following condition:*

Given any $\delta > 0$, there exists some N such that for all $n > N$ and for all $j \in \mathbb{N}$

$$a_{n+j} \geq (\beta - \delta) f\left(f^{-1}\left(\frac{a_n}{\beta}\right) + j\right). \tag{1.1}$$

If f is asymptotically stable, the conditions are sufficient.

Proof. First, I prove that the condition is sufficient when f is asymptotically stable.

Assume A satisfies Condition (1.1), and construct $B = b_k$ as follows:

Define

$$k(n) = \begin{cases} 1, & k = 1 \\ \max\{k(n-1)+1, \lceil f^{-1}(a_n/\beta)\rceil\}, & \text{elsewhere.} \end{cases}$$

Thus $k(n)$ is a strictly increasing integer-valued function with $k(n) \geq n$. Let

$$b_k = \begin{cases} a_1, & k = 1 \\ a_n, & k = k(n) \text{ for some } n \in \mathbb{N} \\ \beta f(k), & \text{elsewhere.} \end{cases}$$

Then B is a well-defined supersequence of A, so it suffices to show that $b_k/f(k)$ approaches β. By definition, $b_k/f(k) = \beta$ outside the range of $k(n)$, but we need to prove it for values of k in the range of $k(n)$.

Choose any $n_1 > N$. Then there are two cases, either $k(n_1) = \lceil f^{-1}(a_{n_1}/\beta) \rceil$, or $k(n_1) > \lceil f^{-1}(a_{n_1}/\beta) \rceil$.

In the first case,

$$\frac{b_k}{f(k)} = \frac{a_{n_1}}{f(\lceil f^{-1}(a_{n_1}/\beta) \rceil)}.$$

Since a_{n_1} grows without bound and f is asymptotically stable, this expression approaches β as $k \to \infty$.

Now, I consider the second case. Assume for the moment that the sequence $\frac{a_n}{f(n)}$ has multiple limit points. (The alternative, that $\frac{a_n}{f(n)}$ converges, will be dealt with at the end).

Let n_0 be the greatest integer such that $n_0 < n_1$ and $k(n_0) = \lceil f^{-1}(a_{n_0}/\beta) \rceil$ (Case 1 holds for n_0.) Since $\frac{a_n}{f(n)}$ has multiple limit points, and since $\beta \leq \liminf_{n\to\infty} \frac{a_n}{f(n)}$, $\frac{a_n}{f(n)}$ has a limit point which is greater than β. Therefore, there are arbitrarily large values of n for which $\frac{a_n}{f(n)} > \beta$.

Let $j = n_1 - n_0$. Because n_0 is the *greatest* integer less than n_1 for which Case 1 holds, $k(n_1) = k(n_1 - 1) + 1 = k(n_1 - 2) + 2 = ... = k(n_1 - j) + j$. Then

$$k(n_1) = k(n_0 + j) = k(n_0) + j = \lceil f^{-1}(a_{n_0}/\beta) \rceil + j > \lceil f^{-1}(a_{n_1}/\beta) \rceil,$$

or else $k(n_1)$ would be $\lceil f^{-1}(a_{n_1}/\beta)\rceil$. Then:

$$\begin{aligned} f^{-1}(a_{n_0}/\beta)+j+1 &> f^{-1}(a_{n_1}/\beta) \\ f(f^{-1}(a_{n_0}/\beta)+j+1) &> a_{n_1}/\beta \\ \beta f\left(f^{-1}\left(\frac{a_{n_0}}{\beta}\right)+j+1\right) &> a_{n_1}=b_{k(n_1)}. \end{aligned} \tag{1.2}$$

So, for $k=k(n_1)$,

$$\frac{b_k}{f(k)}=\frac{a_{n_1}}{f(k(n_0)+j)}<\frac{\beta f\left(f^{-1}\left(\frac{a_{n_0}}{\beta}\right)+j+1\right)}{f(\lceil f^{-1}\left(\frac{a_{n_0}}{\beta}\right)\rceil\ +j)}.$$

Since, by assumption, f is asymptotically stable, as $n_0\to\infty$,

$$\frac{f\left(f^{-1}\left(\frac{a_{n_0}}{\beta}\right)+j+1\right)}{f(\lceil f^{-1}\left(\frac{a_{n_0}}{\beta}\right)\rceil\ +j)}\to 1,$$

By Condition (1.1),

$$a_{n_1}\geq(\beta-\delta)f\left(f^{-1}\left(\frac{a_{n_0}}{\beta}\right)+j\right),$$

where $\delta\to 0$ as $n_1\to\infty$. Thus we run up against the same limit β from above and below, and the result is proven.

In the case where $\frac{a_n}{f(n)}$ converges to some limit L, whenever f is asymptotically stable and $\beta\leq L$, we can construct a supersequence $B=\{b_k\}$ asymptotic to $\beta f(n)$. Simply set $b_k=a_n$, where $k=k(n)=\lceil f^{-1}(\frac{L}{\beta}f(n))\rceil$, and $b_k=\beta f(k)$ elsewhere. Then

$$\frac{b_k}{f(k)}=\frac{a_n}{f(\lceil f^{-1}(L/\beta f(n))\rceil)}\sim\frac{\beta}{L}\frac{a_n}{f(n)}\sim\beta.$$

I will now prove the other part of the theorem, that Condition (1.1) is necessary, whether f is asymptotically stable or not. (It is plainly necessary that $\beta \leq \liminf_{n\to\infty} \frac{a_n}{f(n)}$.)

Assume that there is some $\alpha > 0$ such that for any number $N, \exists n > N, j \in \mathbb{N}$ such that

$$a_{n+j} < (\beta - \alpha) f\left(f^{-1}\left(\frac{a_n}{\beta}\right) + j\right).$$

Also assume that A has a supersequence $B = \{b_k\}$, with $\lim_{k\to\infty} \frac{b_k}{f(k)} = \beta$.

Since n may be arbitrarily large, we can choose n such that $|\beta - \frac{a_n}{f(k(n))}| < \epsilon$ so that $k = \lceil f^{-1}\left(\frac{a_n}{\beta \pm \epsilon}\right) \rceil$. Then

$$\begin{aligned}
\frac{b_k}{f(k)} &= \frac{a_{n+j}}{f(k(n+j))} \\
&< \frac{(\beta-\alpha) f(f^{-1}(\frac{a_n}{\beta}) + j)}{f(\lceil f^{-1}(\frac{a_n}{\beta \pm \epsilon}) \rceil + j)} \\
&< \frac{(\beta-\alpha) f(f^{-1}(\frac{a_n}{\beta}) + j)}{f(f^{-1}(\frac{a_n}{\beta - \epsilon}) + j)}.
\end{aligned}$$

This expression approaches $\beta - \alpha$, not β. Hence we have a contradiction, and Condition (1.1) is necessary to find a convergent supersequence B. □

Theorem 1.2. *Let $f(n)$ be an asymptotically stable growth function and let A be a strictly increasing sequence of integers such that $f(n) \leq a_n \leq Uf(n)$, where U is constant. Then it is possible to find a strictly increasing supersequence $B \subset \mathbb{N}$ with $\lim_{k\to\infty} \frac{b_k}{f(k)} = \beta$ if and only if $\beta f(n+1) - \beta f(n) \geq 1$ for all sufficiently large n, and A meets the conditions of Theorem 1.*

Proof. By Theorem 1, the conditions are necessary.

Proving that they are sufficient requires a bit of extra work. We construct B in the

same manner as in the proof of Theorem 1, with a slight modification to $B \setminus A$: If $k \neq k(n)$ for any n, let $b_k = \lfloor \beta f(k) \rfloor$.

Then B is a sequence of integers, asymptotic to $\beta f(k)$. Since A is strictly increasing, it suffices to show that B is strictly increasing at each $b_i \in B \setminus A$.

Let $k(n) < i < k(n+1)$, so that b_i is an arbitrary element of $B - A$. I will show that $b_{i-1} < b_i < b_{i+1}$, by breaking into three cases.

I) If all three of these terms are in $B - A$, by assumption $\lfloor \beta f(i-1) \rfloor < \lfloor \beta f(i) \rfloor < \lfloor \beta f(i+1) \rfloor$, and B is strictly increasing at i.

II) If $i - 1 = k(n)$, then $b_{k(n)} = a_n$, and

$$\begin{aligned} k(n) &\geq \lceil f^{-1}(a_n/\beta) \rceil \\ k(n) &\geq f^{-1}(a_n/\beta) \\ f(k(n)) &\geq a_n/\beta \\ \beta f(k(n)) &\geq a_n. \end{aligned}$$

Then

$$b_{i-1} = a_n \leq \beta f(k(n)) < \lfloor \beta f(k(n)+1) \rfloor = b_i.$$

III) If $i+1 = k(n+1)$, $b_{k(n+1)} = a_{n+1}$. The need for an "extra" term i means $k(n+1) >$

$k(n)+1$. Therefore:

$$\begin{aligned} k(n+1) &= \lceil f^{-1}(a_{n+1}/\beta) \rceil \\ k(n+1)-1 &< f^{-1}(a_{n+1}/\beta) \\ f(k(n+1)-1) &< a_{n+1}/\beta \\ \beta f(k(n+1)-1) &< a_{n+1} \\ &= b_{i+1}. \end{aligned}$$

Then we have:

$$\begin{aligned} b_{i+1} &> \beta f(k(n+1)-1) = \beta f(i) \geq \lfloor \beta f(i) \rfloor = b_i \\ b_{i+1} &> b_i. \end{aligned}$$

Thus B is strictly increasing. □

1.2 Explicit examples

A Negative Example. Let $f(n) = n^2$. Below is a strictly increasing sequence of integers with $n^2 \leq a_n \leq 2n^2$ which cannot have a supersequence asymptotic to βn^2 for any constant β.

Let $a_1 = 2$, and let

$$a_n = \begin{cases} a_{n-1}+1, & a_{n-1}+1 \geq n^2 \\ 2n^2, & \text{otherwise.} \end{cases}$$

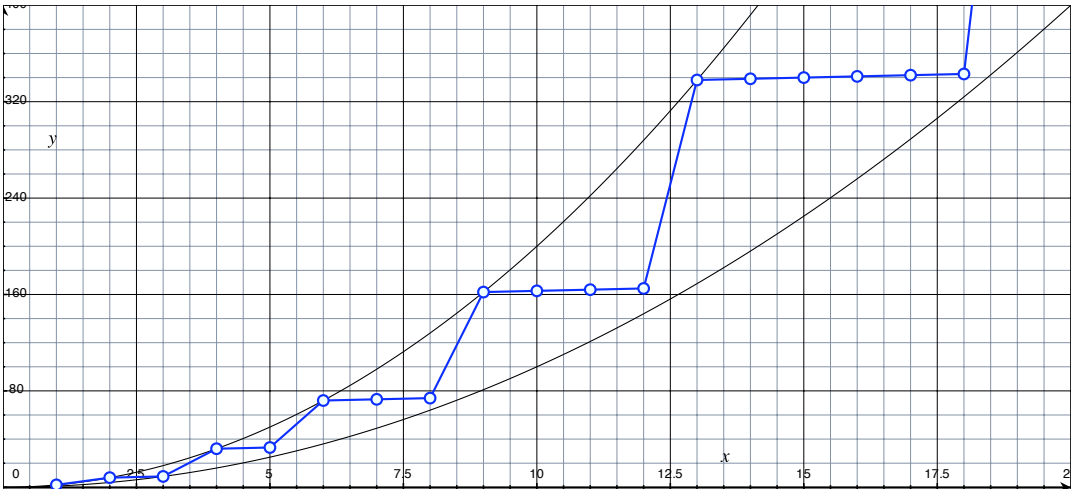

Figure 1.1: The sequence which fails

The first few terms are:

$$\{2, 8, 9, 32, 33, 72, 73, 74, 162, 163, 164, 165, 338, 339, 340, 341, 342, 343, 761\}.$$

The first terms are also plotted between the curves $y = x^2$ and $y = 2x^2$ in Figure 1.1.

We will now show it cannot have a supersequence asymptotic to βn^2.

Let n be a large number with $a_n = 2n^2$, and let j be the greatest integer such that

$$a_{n+j} = 2n^2 + j \geq (n + j)^2$$

That is, j is the distance between successive $2n^2$ terms. Then we find:

$$0 \geq j^2 + (2n - 1)j - n^2$$

$$j \leq \frac{1}{2} - n + \sqrt{2n^2 - n + \frac{1}{4}}$$

For sufficiently large n, $j \approx (\sqrt{2} - 1)n$.

Assume B is a supersequence of A. Then two terms of B are $b_k = a_n = 2n^2$ and $b_{k+j+i} = a_{n+j} = 2n^2 + j$, for $j \approx (\sqrt{2} - 1)n = \alpha n$. If $b_k/f(k)$ is asymptotic to a constant β, the following expression should approach 0 as $n \to \infty$.

$$\begin{aligned}
&\left| \frac{b_k}{f(k)} - \frac{b_{k+j+i}}{f(k+j+i)} \right| \\
=&\left| \frac{2n^2}{k^2} - \frac{2n^2 + j}{(k+j)^2} \right| \\
=&\left| \frac{2n^2(k+j+i)^2 - (2n^2 + j)k^2}{k^2(k+j+i)^2} \right| \\
=&\left| \frac{4n^2k(j+i) + (j+i)^2 - k^2j}{k^2(k+j+i)^2} \right|.
\end{aligned}$$

By assumption, $j = \alpha n$. Also, if B is asymptotic to βn^2, we must have $k = mn$. Then the above expression becomes

$$\frac{4m\alpha n^4 + 4min^3 + \alpha^2 n^2 + 2i\alpha n + i^2 - m^2\alpha n^3}{m^2n^2(mn + \alpha n + i)^2}.$$

Since α, m are constants and i grows no faster than n, for large n it is asymptotically

$$\frac{4m\alpha n^4}{m^2(m+\alpha)^2n^4},$$

which does not approach 0 as $n \to \infty$.

The reason this sequence fails is that $\frac{a_n}{f(n)}$ decreases too rapidly, so that the value of j is too small. Here is a similar sequence which works.

A Positive Example. Let $f(n) = n^2$, $U = 2$, and $\beta = \frac{1}{2}$. Define the sequence A as

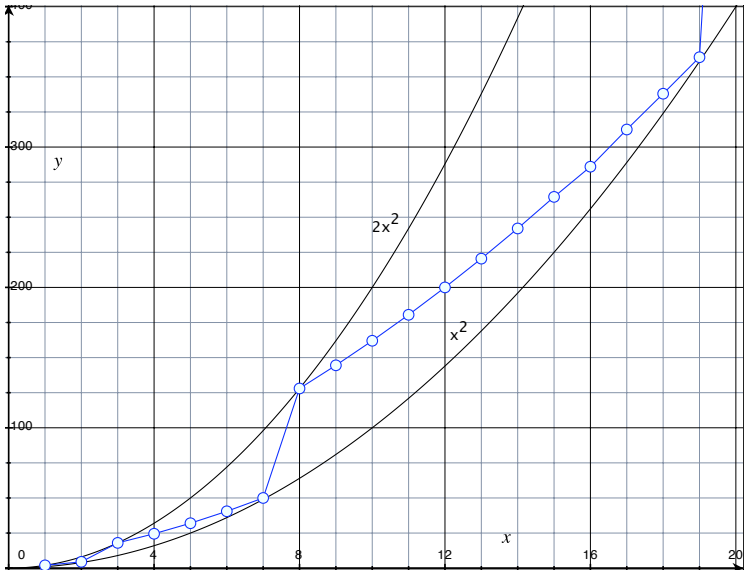

Figure 1.2: The sequence that works

follows:

$$a_{n_i} = 2n^2$$

$$n_1 = 1$$

$$n_{i+1} = n_i + \lceil n_i\sqrt{2} \rceil$$

For $0 < j < n_i\sqrt{2}$,

$$a_{n_i+j} = \frac{1}{2}(j + 2n_i)^2$$

Thus, $a_{n_2} = a_3 = 18$, $n_4 = 8$, $n_5 = 20$, and the first few terms of the sequence are:

$$\{2, 4.5, 18, 24.5, 32, 40.5, 50, 128, 144.5, 162, 180.5, 200, 220.5, 242, 264.5, 286, 312.5, 338, 364.5, 800\}.$$

The graph above shows the first few points, plotted between $y = x^2$ and $y = 2x^2$.

Then 1 and 2 are both limit points of a_n/n^2, and A obeys Condition (1.1) for $\beta = \frac{1}{2}$.

In fact, if we choose B correctly, then $b_k = \frac{k^2}{2}$ for all k.

To check the last statement, observe that for all i, $2n_i^2 = a_{n_i} = b_{k(n_i)}$, and let

$$k(n_i) = \lceil \sqrt{2a_{n_i}} \rceil = \lceil \sqrt{4n_i^2} \rceil = 2n_i$$

$$b_{2n_i} = 2n_i^2 = \frac{1}{2}(2n_i)^2.$$

This deals with the terms in the subsequence a_{n_i}. For the other terms, a similar logic applies, with $k(n_i + j) = k(n_i) + j$, so that

$$b_{k(n_i)+j} = b_{2n_i+j} = \frac{1}{2}(j + 2n_i)^2.$$

As before, for any k not in the range of $k(n)$, just define $b_k = \frac{k^2}{2}$.

1.3 Which functions and constants work?

The next question is, for a particular asymptotically stable growth function $f(n)$, constant U, and sequence A, for which values of β can we find a supersequence B asymptotic to $\beta f(n)$?

Theorem 1.3. *Given any asymptotically stable growth function f and constants $U > 1$ and $\beta < 1$, there exists a sequence A such that $f(n) \leq a_n \leq Uf(n)$ and 1, U are both limit points of $a_n/f(n)$, which has a supersequence asymptotic to $\beta f(n)$.*

Proof. If f is asymptotically stable, then for any integer n,

$$\lim_{j\to\infty} \frac{\beta f\left(f^{-1}\left(\frac{uf(n)}{\beta}\right)+j\right)}{f(n+j)} = \beta.$$

Let $\beta < 1 < U$. Then given any $n_i, \exists j_i > 0$ s.t.

$$\frac{\beta f\left(f^{-1}\left(\frac{uf(n)}{\beta}\right)+j\right)}{f(n+j)} \leq 1.$$

Then we can construct a sequence $A = \{a_n\}$ with $\text{limsup}\frac{a_n}{f(n)} = U$, $\text{liminf}\frac{a_n}{f(n)} = 1$, and super-sequence $B \sim \beta f(n)$, as follows:

Let $n_1 = 1$. For any $i > 0$, let j_i be as in (2), and for $i > 1$, $n_{i+1} = j_i + 1$. Then let $a_{n_i} = Uf(n_i)$, and for $n_i < n < j_i$, let

$$a_n = \beta f\left(f^{-1}\left(\frac{uf(n_i)}{\beta} + n - n_i\right)\right).$$

Then $\frac{a_n}{f(n)}$ has both U and 1 as limit points, and it obeys condition 1. □

Theorem 1.4. *Let a number β be acceptable to a sequence A with $f(n) \leq a_n \leq Uf(n)$ if there exists a supersequence of A asymptotic to $\beta f(n)$. Assume that 1, U are both limit points of $a_n/f(n)$.*

Let $F = \{\beta \in \mathbb{R}$ s.t. β is acceptable to $A\}$.

Then for any f, A, either F is empty, or F is an interval contained in $(0, 1)$.

Proof. The statement $B \sim \beta f$ would have no meaning if β were 0, and negative values of β cannot possibly be acceptable, since all but finitely many terms of A are positive. Further, since 1 is a limit point of $a_n/f(n)$, then $\beta \leq 1$.

Hence, $F \subset (0, 1]$.

Additionally, let $1 \in F$. Then, by Theorem 1 and the fact that 1, U both limit points of $a_n/f(n)$, we have $j \in \mathbb{N}$, n arbitrarily large, satisfying:

$$f(n+j) \approx a_{n+j} \geq f(f^{-1}(a_n)+j) \approx f(f^{-1}(Uf(n))+j)$$

$$n \geq f^{-1}(Uf(n)) - \epsilon.$$

But since f is a growth function defined on all positive integers, its inverse also increases without bound, so this is impossible. Then $\beta \neq 1$, and $F \subset (0,1)$.

In the proof of Theorem 1, it was shown that if $\beta_0 \in F$, and $\beta < \beta_0$, then $\beta \in F$. □

Further, it is possible to find a sequence, having both 1 and U as limit points of $\frac{a_n}{f(n)}$, which has a supersequence asymptotic to $\beta f(n)$ for any $\beta \in (0,1)$. The construction is as follows:

Let $a_1 = Uf(1)$, and $a_{n_i} = Uf(n_i)$.

For $n_i < n < n_{i+1}$,

$$a_n = \frac{i-1}{i} f\left(f^{-1}\left(\frac{a_{n_i}}{(i-1)/i}\right) + n - n_i\right).$$

Just as in the proof of Theorem 3, when the above equation yields a value less than $f(n)$, set $n = n_{i+1}$.

Then Condition (1.1) is satisfied for any $\beta < \frac{i-1}{i}$, for all $n > n_i$.

Theorem 1.5. *Given any asymptotically unstable growth function f and constants $U > 1$ and $\beta < 1$, and any sequence A with $f(n) \leq a_n \leq Uf(n)$ and 1, U both limit points of $a_n/f(n)$, it is impossible to find a supersequence of A asymptotic to $\beta f(n)$.*

Proof. Let f be an asymptotically unstable growth function, $U > 1$ a constant, and choose

a sequence A with $\liminf \frac{a_n}{f(n)} = 1$ and $\limsup \frac{a_n}{f(n)} = U$. Assume that B is a supersequence of A asymptotic to $\beta f(n)$, $0 < \beta < 1$ a constant.

By Theorem 1, $\exists n$ arbitrarily large, $j \in \mathbb{N}$ s.t. $a_n \approx Uf(n)$ and $a_{n+j} \approx f(n+j)$. Then

$$f(n+j) - \epsilon \geq (\beta - \delta) f(f^{-1}(a_n/\beta) + j) \geq (\beta - \delta) f\left(f^{-1}\left(\frac{U(f(n)+\epsilon)}{\beta}\right) + j\right)$$

$$f(n+j) \geq (\beta - \epsilon) f\left(f^{-1}\left(\frac{U}{\beta}(f(n)+\epsilon)\right) + j\right) + \epsilon.$$

Choose $\beta' < \beta$ so that we may drop the error term for large n, then

$$f(n+j) \geq \beta' f\left(f^{-1}\left(\frac{U}{\beta}(f(n)+\epsilon)\right) + j\right).$$

Let $g = f^{-1}(\frac{U}{\beta} f(n)) - n$. Then g increases without bound, since $\frac{U}{\beta} > 1$, and

$$1 < \lim_{n+j\to\infty} \frac{f(j+n+g)}{f(j+n)} < 1/\beta'.$$

Since f is asymptotically unstable, $\lim_{n+j\to\infty} \frac{f(n+j+1)}{f(n+j)} \geq \alpha$, for some $\alpha > 1$.

Then $\lim_{n+j\to\infty} \frac{f(n+j+g)}{f(n+j)} \geq \alpha^g$ and is less than a fixed $1/\beta$. Hence, there is a contradiction, and A cannot satisfy Condition (1.1) for any β. □

Additional observation. Let f be a growth function with $\lim_{n\to\infty} \frac{f(n+1)}{f(n)} = \infty$, ($f(n) = e^{(n^2)}$ for example), and let A be as in Theorem 1. If A has a supersequence B which is asymptotic to $\beta f(n)$, then $a_n = b_n$ for all n.

Assume A is a proper subsequence of B. Then $\exists N$ s.t. $\forall n > N$, $k(n) \geq n+1$.

Then

$$\lim_{n\to\infty} \frac{b_k}{f(k)} = \lim_{n\to\infty} \frac{a_n}{f(k)} \leq \lim_{n\to\infty} \frac{Uf(n)}{f(n+1)} = 0.$$

Then $b_k \nsim \beta f(k)$.

Chapter 2

Binary Linear Forms

Some classical problems in additive number theory involve the size and structure of sum sets. The sum set is defined as $s(A) = \{x + y \mid x, y \in A\}$. Thus if $A = \{2, 3, 7\}$, then $s(A) = \{4, 5, 6, 9, 10, 14\}$.

The difference set is defined as $d(A) = \{x - y \mid x, y \in A\}$. Since addition is commutative and subtraction is not, the difference set is most often larger than the sum set. In fact, examples of sets for which $|s(A)| > |d(A)|$ were at first found only by computer search. Recently, however, many families of such sets have been described. Sets with more sums than differences will be discussed a bit further in the next chapter.

In this chapter, we consider a generalization of the sum set. For a set $A \subset \mathbb{Z}$, and coefficients $u, v \in \mathbb{Z}$, define $f(A) = f(x, y) = \{ux + vy \mid x, y \in A\}$. We can then consider the size and structure of the set $f(A)$. Clearly $|f(A)| \leq |A|^2$, but in many cases $|f(A)| < |A|^2$.

For all linear forms discussed henceforth, assume they are normalized in the following fashion: $\gcd(u, v) = 1$, the first coefficient u is positive, and $u \geq |v|$. (If not, simply reorder or divide by a constant, as neither will change the value of $|f(A)|$ on a given set A.)

We will now examine the following question: Let $f(x, y) = u_1x + v_1y$ and $g(x, y) =$

$u_2x + v_2y$ be normalized binary linear forms with nonzero integer coefficients $(u_1, v_1) \neq (u_2, v_2)$. Do there exist finite sets of integers A and B such that $|f(A)| > |g(A)|$ and $|f(B)| < |g(B)|$?

In most cases, there are small sets with this property which we can describe explicitly.

2.1 Pairs of Forms with $u_1 \geq 2$

Theorem 2.1. *For $u > |v| \geq 1$ and $(u, v) = 1$, consider the normalized binary linear form*

$$f(x, y) = ux + vy.$$

i)If $|A| = 2$, then $|f(A)| = 4$.

ii)If $u \geq 3$ and $|A| = 3$, then $|f(A)| = 8$ or 9, and $|f(A)| = 8$ if and only if A is affinely equivalent to one of the two sets

$$\{0, v, u\} \text{ and } \{0, v, u + v\}.$$

iii)If $u = 2$ and $|A| = 3$, then $|f(A)| < 9$ if and only if A is affinely equivalent to one of the two sets

$$\{0, 1, 2\} \text{ and } \{0, 1, 3\}.$$

Moreover, $|f(\{0, 1, 2\})| = 7$ and $|f(\{0, 1, 3\})| = 8$.

iv)If $f(x, y) = ux + vy$, and $g(x, y) = ux - vy$, then $|f(A)| = |g(A)|$ for every set A with $|A| = 3$.

Proof. i) Since $f(c * A) = c * f(A)$, $|f(A)|$ is constant under affine transformation on A. Then $|f(A)| = |f(0, 1)| = |0, v, u, u + v|$ whenever $|A| = 2$.

ii)If $|A| = 3$ and $|f(A)| < 9$, then we have a nontrivial (meaning $(x_1, y_1) \neq (x_2, y_2)$) solution to:

$$ux_1 + vy_1 = ux_2 + vy_2$$

$$u(x_1 - x_2) = v(y_2 - y_1)$$

$$\frac{u}{v} = \frac{y_2 - y_1}{x_1 - x_2}.$$

Let us normalize so that A is relatively prime and $0 \in A$. Then, since u and v are also relatively prime, we have $y_2 - y_1 = cu$ and $x_1 - x_2 = cv$, $c = \pm 1$. Since there are only three elements in the set, at least two of $\{x_1, y_1, x_2, y_2\}$ must be the same.

Let $x_1 = y_1$, $c = 1$. Then

$$y_2 - x_1 = u$$

$$x_1 - x_2 = v.$$

Which implies

$$x_1 = v + x_2$$

$$y_2 = u + v + x_2.$$

Assume $x_2 = 0$, then

$$A_1 = \{0, v, u + v\}.$$

Let $x_1 = y_1$, $c = -1$, then

$$y_2 - x_1 = -u$$

$$x_1 - x_2 = -v$$

$$y_2 + u = x_1$$

$$y_2 + u + v = x_2$$

Then we have again

$$A_1 = \{0, v, u + v\}.$$

Let $x_1 = y_2$, $c = 1$

$$x_1 - y_1 = u, x_1 - x_2 = v$$

$$x_1 = 0, x_2 = -v, y_1 = -u$$

$$A_2 = \{0, v, u\}$$

Let $x_1 = y_2$, $c = -1$

$$x_1 - y_1 = -u, x_1 - x_2 = -v$$

$$x_1 + u = y_1, x_1 + v = x_2.$$

Then we have again

$$A_2 = \{0, v, u\}.$$

If $x_1 = x_2$, then $y_1 = y_2$ and the solution is trivial.

These two sets A_1, A_2 are distinct under affine transformation.

We will now show that if $u > 2$, $|f(A_1)| = |f(A_2)| = 8$.

If $v > 0$, then $f(A_1) = \{0, v^2, uv, u^2, uv + v^2, 2uv, u^2 + v^2, u^2 + uv\}$. Each element is a homogeneous quadratic polynomial in two variables. Then if two of these elements are equal to each other, their difference must have a rational root other than $\{\pm 1, \pm 2, \pm 1/2\}$. Therefore, the difference must have a coefficient which is at least 3. But the largest coefficient is 2, and there are no negative coefficients, so these eight elements are distinct whenever $u > 2$.

If $v < 0$, then $f(A_1) = \{0, -v^2, -uv, u^2, -uv - v^2, uv, u^2 - v^2, u^2 + uv\}$. There are no coefficients greater than 2, hence the difference of two elements will have no coefficients of 3 or more, and $|f(A_1)| = 8$.

If $v > 0$, $f(A_2) = \{0, uv, uv + v^2, u^2, u^2 + uv, u^2 + uv + v^2, u^2 + 2uv, u^2 + 2uv + v^2\}$. The largest coefficient is 2, and there are no negative coefficients, hence the difference of two elements will not have a coefficient of 3 or more, and $|f(A_2)| = 8$.

If $v < 0$, then $f(A_2) = \{0, uv, uv - v^2, u^2, u^2 + uv, u^2 + uv - v^2, u^2 - uv, u^2 - v^2\}$. There are no coefficients greater than 2, hence $|f(A_2)| = 8$ again.

iii) As we just showed, A must be equivalent to $\{0, |v|, u\}$ or $\{0, u, u + |v|\}$. Then if $u = 2$, A is $\{0, 1, 2\}$ or $\{0, 2, 3\}$, and a straightforward calculation will show the size of $f(A)$ in each case. If we have a nontrivial solution to

$$ux_1 + vy_1 = ux_2 + vy_2,$$

Then we have a nontrivial solution to

$$ux_1 - vy_2 = ux_2 - vy_1.$$

□

Hence, the size of one form is smaller than $|A|^2$ if and only if the other is. Since $|A| = 3$ implies $|f(A)| = 8$ or 9, (except in the case $u = 2$, which was just dealt with) we cannot distinguish $ux + vy$ and $ux - vy$ using three-element sets.

However, we can distinguish them using four-element sets.

Theorem 2.2. *Let $f(x, y) = ux + vy$ and $g(x, y) = ux - vy$ be normalized binary linear forms with $u > v$.*

1: For $u = 2$, if

$$A = \{0, 3, 4, 6\}$$

$$B = \{0, 4, 6, 7\}$$

then

$$|f(A)| = 13 > 12 = |g(A)|$$

$$|f(B)| = 13 < 14 = |g(B)|.$$

2: For $u \geq 3$, if

$$A = \{0, u^2 - v^2, u^2, u^2 + uv\}$$

$$B = \{0, u^2 - uv, u^2 - v^2, u^2\}$$

then

$$|f(A)| = 14 > 13 = |g(A)|$$

$$|f(B)| = 13 < 14 = |g(B)|.$$

Proof. The following addition table shows the elements of $f(A) = \{ux + vy \mid x, y \in A\}$.

$f(A)$	0	u^2-v^2	u^2	u^2+uv
0	0	u^2v-v^3	u^2v	u^2v+uv^2
u^2-v^2	u^3-uv^2	$u^3+u^2v-uv^2-v^3$	$u^3+u^2v-uv^2$	$\boxed{u^3+u^2v}$
u^2	u^3	$u^3+u^2v-v^3$	$\boxed{u^3+u^2v}$	$u^3+u^2v+uv^2$
u^2+uv	$\boxed{u^3+u^2v}$	$u^3+2u^2v-v^3$	u^3+2u^2v	$u^3+2u^2v+uv^2$

Since u^3+u^2v appears three times, $|f(A)| \leq 14$. As I argued in the proof of the second part of the previous theorem, if two of these elements were equal to each other, their difference would have a rational solution, which means it would need a coefficient of 3 or more. Some of these polynomials have a $2u^2v$ term, but none have a negative u^2v term. Then they are all distinct for $u > 2$, and $|f(A)| = 14$.

$g(A)$	0	u^2-v^2	u^2	u^2+uv
0	0	$-u^2v+v^3$	$-u^2v$	$-u^2v-uv^2$
u^2-v^2	$\boxed{u^3-uv^2}$	$u^3-u^2v-uv^2+v^3$	$\boxed{u^3-u^2v-uv^2}$	$u^3-u^2v-2uv^2$
u^2	$\boxed{u^3}$	$u^3-u^2v+v^3$	u^3-u^2v	$\boxed{u^3-u^2v-uv^2}$
u^2+uv	u^3+u^2v	u^3+v^3	$\boxed{u^3}$	$\boxed{u^3-uv^2}$

Since u^3-uv^2, u^3, and $u^3-u^2v-uv^2$ appear twice, $|g(A)| \leq 13$. By the same rational root argument, if two elements are equal, one of them must be $u^3-u^2v-2uv^2$. Since this table increases from top to bottom and right to left, $u^3-u^2v-uv^2 > u^3-u^2v-2uv^2 > u^2v$. Then since there are no nontrivial rational solutions to

$$u^3-u^2v-2uv^2=0$$

or

$$u^3-u^2v-2uv^2=-u^2v+v^3,$$

$|g(A)| = 13$.

$f(B)$	0	u^2-uv	u^2-v^2	u^2
0	0	u^2v-uv^2	u^2v-v^3	u^2v
u^2-uv	u^3-u^2v	$\boxed{u^3-uv^2}$	u^3-v^3	$\boxed{u^3}$
u^2-v^2	$\boxed{u^3-uv^2}$	$u^3+u^2v-2uv^2$	$u^3+u^2v-uv^2-v^3$	$\boxed{u^3+u^2v-uv^2}$
u^2	$\boxed{u^3}$	$\boxed{u^3+u^2v-uv^2}$	$u^3+u^2v-v^3$	u^3+u^2v

The only polynomial in $f(B)$ with a coefficient greater than 2 is $u^3+u^2v-2uv^2$, and since the rows and columns are strictly increasing, $u^3+u^2v-2uv^2$ cannot be equal to any other polynomial containing a uv^2 term. Then $|f(B)|=13$.

$g(B)$	0	u^2-uv	u^2-v^2	u^2
0	0	$-u^2v+uv^2$	$-u^2v+v^3$	$-u^2v$
u^2-uv	$\boxed{u^3-u^2v}$	$u^3-2u^2v+uv^2$	$u^3-2u^2v+v^3$	u^3-2u^2v
u^2-v^2	u^3-uv^2	$\boxed{u^3-u^2v}$	$u^3-u^2v-uv^2+v^3$	$u^3-u^2v-uv^2$
u^2	u^3	$u^3-u^2v+uv^2$	$u^3-u^2v+v^3$	$\boxed{u^3-u^2v}$

The elements which are polynomials with coefficients greater than 2 are $u^3-2u^2v+uv^2, u^3-2u^2v+v^3$, and u^3-2u^2v. Since none of the other elements have a positive u^2v term, we cannot have rational values for $u\geq 2$ such that two elements are equal to each other, and $|g(B)|=14$. □

2.2 Comparisons with the Difference Set

We will now examine the case where one of the forms is the difference set.

Theorem 2.3. *Let u and v be relatively prime positive integers with $u>v$, and $u>2$, and consider the linear forms*

$$f(x,y)=ux+vy \text{ and } d(x,y)=x-y.$$

Let

$$A = \{0, v^3, v^3 + v^2u, v^3 + v^2u + vu^2, v^3 + v^2u + vu^2 + u^3\}.$$

Then

$$|f(A)| \leq 19 \text{ and } |d(A)| = 21.$$

Proof. First, we find $d(A)$. The only difference that appears more than once is 0, which appears 5 times. Also, $d(A)$ consists entirely of homogeneous cubic polynomials with coefficients in $\{-1, 0, 1\}$, hence (by the same argument used in the previous section) the difference of any two such elements has no rational roots with $u > 2$ and the distinct polynomials are distinct elements. Then $|d(A)| = 21$ whenever $u > 2$.

Let

$$\begin{aligned}
a_0 &= 0 \\
a_1 &= v^3 \\
a_2 &= v^3 + v^2u \\
a_3 &= v^3 + v^2u + vu^2 \\
a_4 &= v^3 + v^2u + vu^2 + u^3.
\end{aligned}$$

Now, we will show that at least 6 elements of $f(A)$ have multiple representations:

$$\begin{aligned}
n_1 &= ua_1 + va_1 = ua_0 + va_2 \\
n_2 &= ua_2 + va_1 = ua_0 + va_3 \\
n_3 &= ua_2 + va_2 = ua_1 + va_3 \\
n_4 &= ua_3 + va_1 = ua_0 + va_4 \\
n_5 &= ua_3 + va_2 = ua_1 + va_4 \\
n_6 &= ua_3 + va_3 = ua_2 + va_4.
\end{aligned}$$

Then $|f(A)| \leq 19$.

□

We now handle the case $u = 2$ separately.

Let $B = \{1, 2, 4, 8, 16\}$.

Then $|d(B)| = 21$ and $|f(B)| = |2B + B| = 19$.

Theorem 2.4. *Let u and v be relatively prime positive integers with $u > v$, and $u > 2$, and consider the linear forms*

$$f(x, y) = ux - vy \text{ and } d(x, y) = x - y.$$

Let

$$A = \{0, u^2v - uv^2, u^2v - uv^2 + v^3, u^2v, u^3\}.$$

Then

$$|f(A)| \leq 19 \text{ and } |d(A)| = 21.$$

Proof. As in the previous theorem, $d(A)$ consists entirely of homogeneous cubic polynomials, and all coefficients are ± 1. The only repeated element is 0, which appears 5 times. Thus $|d(A)| = 21$.

We will also show that there are at least 6 elements with multiple representations in $f(A)$. Let:

$$
\begin{aligned}
a_0 &= 0 \\
a_1 &= u^2v - uv^2 \\
a_2 &= u^2v - uv^2 + v^3 \\
a_3 &= u^2v \\
a_4 &= u^3.
\end{aligned}
$$

Then

$$
\begin{aligned}
0 = ua_0 - va_0 &= ua_3 - va_4 \\
n_2 = ua_0 - va_3 &= ua_1 - va_4 \\
n_3 = ua_0 - va_1 &= ua_2 - va_4 \\
n_4 = ua_2 - va_3 &= ua_1 - va_1 \\
n_5 = ua_3 - va_3 &= ua_1 - va_0 \\
n_6 = ua_2 - va_0 &= ua_3 - va_1.
\end{aligned}
$$

□

Here is a specific example for $u = 2$: $B = \{0, 2, 6, 7, 10\}$.

Then $|d(B)| = 19$ and $|f(A)| = |2A - A| = 18$.

Hence, we have answered the question posed at the beginning of the chapter, except in the case where one of the sets is the sum set.

2.3 Comparisons with the sum set.

Let $s(x, y) = x + y$, and let $A = \{0, 1\}$. Let $f(x, y) = ux + vy$, $u \geq 2$.

Then $|s(A)| = |d(A)| = 3$ and $|f(A)| = |\{0, v, u, u + v\}| = 4$.

Let $B = \{0, 1, 3\}$. Then $|s(A)| = 6 < |d(A)| = 7$.

Let $f(x, y) = ux + vy$, $u \geq 2$. Finding a set with $|f(A)| < |s(A)|$ proved considerably more difficult. Various computer searches have failed to find any examples of such sets, and I believe no small sets with this property exist. However, a proof that some very large sets with the property do exist is described in [13].

Chapter 3

h-Fold Sum and difference Sets

In the last chapter, we considered one generalization of the sum set, to the set $u*A+v*A$. In this chapter, we generalize in another way, and study linear forms such as $f = \sum_{i=1}^{h} \epsilon_i * A$, where $\epsilon_i \in \{-1, 1\}$. I will provide partial answers to two questions.

1) If f, g are linear forms with h terms, and all coefficients are ± 1, do there always exist sets A, B, C such that $|f(A)| > |g(A)|$, $|f(B)| < |g(B)|$ and $|f(C)| = |g(C)|$?

2) If such sets exist, how plentiful are they?

In the case where $h = 2$, comparing the sum set and the difference set, there is a body of literature concerning sets with more sums than differences. Nathanson [12] and Hegarty [6] both constructed many infinite families of sets with more sums than differences (MSTD sets). Nathanson's were constructed by forming a symmetric set that was close to an arithmetic progression, then breaking the symmetry by adding one more element. Hegarty proved that there are no MSTD sets with fewer than 8 elements, and that there is only one MSTD set with 8 elements, up to affine transformation. He also described additional families of nearly symmetric MSTD sets.

Recent work has also examined the prevalence of MSTD sets. O'Bryant and Martin

[9] determined that a positive proportion of the subsets of any sufficiently long arithmetic progression are MSTD sets, but Hegarty and Miller, [5] using a different counting technique, found that the asymptotic density of MSTD sets was 0.

However, with linear forms having more than two terms, very little is known.

I will use the following notation: h_1 for the number of positive coefficients of f, and h_2 for the number of negative coefficients, so that $h_1+h_2 = h$. I will also use $f = h_1A-h_2A$ to describe such a form.

3.1 Upper bounds

Since we are concerned only with the cardinality of $f(A)$, we can assume without loss of generality that at least half of the coefficients of f are positive, that $h_1 \geq h_2$.

For an interval of length r, such as $C = [1, r]$, then for any form f with h terms, $|f(C)| = |[h_1(1) - h_2(r), h_1(r) - h_2(1)]| = hr - h + 1$. In fact, this holds for any arithmetic progression with h terms.

With $h = 2$, the difference set is "usually" greater than the sum set, especially with thin sets. We will now show that the forms with more positive terms are still "normally" smaller.

If f has h terms, with h_1 coefficients equal to 1 and h_2 equal to -1, let $|A| = r$, and define $S_{h_1,h_2}(r) = \sup_{|A|=r}|f(A)|$.

In the $h = 2$ case, $S_{h_1,h_2}(r)$ is simple and well-known. For the sum set, we have $S_{2,0} = \frac{r^2+r}{2}$ and for the difference set, $S_{1,1} = r^2 - r + 1$ (because 0 appears r times and it should be counted only once.) Sets which have $|A + A| = \frac{r^2+r}{2}$ and $|A - A| = r^2 - r + 1$ are known as Sidon sets.

We will now compute $S_{h_1,h_2}(r)$ for any form with coefficients ± 1. By elementary

combinatorics,

$$|h_1 A| = r \text{ multichoose } h_1 = \binom{r + h_1 - 1}{h_1}.$$

Define

$$T_{h_1,h_2}(r) = |h_1 A||h_2 A| = \binom{r + h_1 - 1}{h_1}\binom{r + h_2 - 1}{h_2}$$

Then $T_{h_1,h_2}(r)$ is an upper bound for the size of $S_{h_1,h_2}(r)$. However, it counts some elements more than once.

In particular, each $x \in (h_1 - 1)A - (h_2 - 1)A$ is counted at least once for each $a \in A$, as $a + x - a$. (In fact, it's counted more than r times, as $(h_1 - 1)A - (h_2 - 1)A$ suffers from the same multiple-counting problem as $h_1 A - h_2 A$.)

So, we have x appearing r times the number of *representatives* of x in the difference set $(h_1 - 1)A - (h_2 - 1)A$. Then, since each of the multiply represented elements should be counted once, we then need to add back the number of unique elements in $(h_1 - 1)A - (h_2 - 1)A$, which is $S_{h_1-1,h_2-1}(r)$. Thus:

$$S_{h_1,h_2}(r) = T_{h_1,h_2}(r) - rT_{h_1-1,h_2-1}(r) + S_{h_1-1,h_2-1}(r)$$

$$S_{h_1,h_2}(r) = T_{h_1,h_2}(r) - (r-1)\sum_{i=1}^{h_2} T_{h_1-i,h_2-i}(r)$$

$$= \binom{r + h_1 - 1}{h_1}\binom{r + h_2 - 1}{h_2} - (r-1)\sum_{i=1}^{h_2}\binom{r + h_1 - i - 1}{h_1 - i}\binom{r + h_2 - i - 1}{h_2 - i}$$

$$\sim \frac{r^{h_1}}{h_1!}\frac{r^{h_2}}{h_2!} - r\sum_{i=1}^{h_2}\frac{r^{h_1-i}}{(h_1-i)!}\frac{r^{h_2-i}}{(h_2-i)!}$$

$$= \frac{r^h}{h_1!\, h_2!} - \sum_{i=1}^{h_2} \frac{r^{h-2i+1}}{(h_1-i)!\,(h_2-i)!}$$

$$\sim \frac{r^h}{h_1!\, h_2!}.$$

By assumption, $h_2 \leq h_1$. Then for thin sets with h fixed, $|f(A)|$ increases with h_2, so that, for example, $|A+A+A+A| < |A+A+A-A| < |A+A-A-A|$ when $|A|$ is thin.

So, we easily have sets to answer two of the three parts of question 1. Now, we will see a family of sets which answers the third part in the special case that the number of terms h is odd and the form has no more than two negative terms.

3.2 Families of nearly symmetric sets

Theorem 3.1. *For any odd $h > 1$ and $k > h$, let $B = \{-2hk-2, -2hk+2, 2hk-2, 2hk+2\} \cup \{2ih+h \mid k \leq i \leq k-1\}$ and $A = \{2h-2\} \cup B$.*

Let $f_1(A)$ = the h-fold sumset, hA.

Let $f_2(A) = (h-1)A - A$.

For $h > 3$, let $f_3(A) = (h-2)A - 2A$.

Then $|f_1(A)| = |f_2(A)| + 1 = |f_3(A)| + 3$.

We will show that, except on a narrow fringe, $f_i(A)$ is nearly an interval, with all missing elements multiples of $2h$.

First, we move into $\mathbb{Z}/2h$ and consider the set $\bar{B} = \{h, 2, -2\}$ and $f_1(\bar{B})$.

Lemma 3.1. *The only representations of 0 in $f_1(\bar{B})$ are $h(2)$ and $h(-2)$.*

Proof. A representation of 0 is a solution to

$$a(2) + b(-2) + c(h) \equiv 0 (\text{mod } 2h)$$

with $a, b, c \geq 0$, $a + b + c = h$.

Since $\gcd(2,\ h) = 1$, we know that $h|(a + b)$. Then $a + b = 0$ or h.

If $a + b = 0,\ c = h$. But $h^2 \equiv h (\mathrm{mod}\ 2h)$, so this is not a solution.

If $a + b = h,\ c = 0$ then we have $a(2) + (h - a)(-2) \equiv 0 (\mathrm{mod}\ 2h)$, which implies $2a \equiv 0$, so $a = h$ or 0. □

Lemma 3.2. *All other elements of $\mathbb{Z}/2h$ have representations in $f_1(\bar{B})$ containing at least one h term.*

Proof. Since $\{-2, 2\}$ is an arithmetic progression with common difference 4, the $(h-1)$-fold sumset of $\{-2, 2\}$ is an h-term arithmetic progression with common difference 4. Since 4 does not divide $2h$, the AP includes all even elements in $\mathbb{Z}/2h$. Then adding h to the AP yields all the odd elements.

By a similar argument, $(h - 2)\{-2, 2\} + 2h$ contains all the even elements except 0. □

We now leave finite groups and return to the integers.

Lemma 3.3. *All elements of $[-2kh^2, 2kh^2]$ not contained in $f_1(B)$ are multiples of $2h$.*

Proof. In other words, except on a narrow fringe, $f_1(B)$ is almost an interval, with all "missing" elements multiples of $2h$.

As we argued in the proof of Lemma 2.2, $f_1(B)$ contains the set of all sums with exactly one odd term:

$$\{w(-2kh - 2) + x(-2kh + 2) + y(2kh - 2) + z(2kh + 2) + l(2h + 1)$$
$$w + x + y + z = h - 1,\ -k \leq l \leq k - 1\}$$
$$= \{(2kh)(z + y - w - x) + (2)(x + z - y - w) + l(2h + 1)\}.$$

With the only constraint on w, x, y, z that they must add up to an even $h-1$, we have:

$$= \{i(4kh) + j(4) + l(2h) + h \;\Big|\; -\frac{h-1}{2} \leq i, j \leq \frac{h-1}{2}, -k \leq l \leq k-1\}.$$

We re-index to make the parameters non-negative, producing:

$$= \{a(4kh) + b(2h) + c(4) + h - 2kh^2 - 2h + 2 \;\Big|\; 0 \leq a, c \leq h-1,\ 0 \leq b \leq 2k-1\}.$$

Next we have almost a division algorithm argument, with a small modification to take into account problems with the c term.

For any odd x, define:

$$\begin{aligned} x_1 &= x - (h - 2kh^2 - 2h + 2), \\ a &= \left[\frac{x_1}{4kh}\right], \\ x_2 &= x_1 - 4khha, \\ b_1 &= \left[\frac{x_2}{2h}\right], \\ x_3 &= x_2 - 2hb. \end{aligned}$$

Then if $\frac{x_3}{4} \in \mathbb{Z}$, let $c = \frac{x_3}{4}$ and $b = b_1$. If not, let $c = \frac{x_3 - 2h}{4}$ and $b = b_1 + 1$. (If $b_1 = 2k-1$, add 1 to a and let $b = 0$.) Since h is odd, either x_3 or $x_3 - 2h$ is divisible by 4.

The range of b is sufficient to cover the gaps between successive values of a, and c to cover the gaps between values of b. By taking the extreme values of a, b, c we see that all odd elements in the interval $[-2kh^2 - h + 2, 2kh^2 + h - 2]$ are covered, and thus that the odd elements of $[-2kh^2, 2kh^2]$ are.

A similar argument on the set of all elements containing exactly two h terms will

show that it contains all the even elements not multiples of $2h$. $\square$

Lemma 3.4. *Since B is symmetric about 0, $f_1(B) = f_2(B) = f_3(B)$.*

Proof of Theorem :

Now we must consider the effect of breaking the symmetry of the set B by adding $\{2h-2\}$. Since $(h-1)(2kh+2)+(2h-2) = 2kh^2-2kh-4 < 2kh^2$, by Lemmas 3 and 4, any elements of $f_i(A)$ not contained in $f_1(B)$ are multiples of $2h$. Then they fall into one of two sets. The first is:

$$\{i(2hk-2)+j(2h-2)+(h-h_2-i-j)(-2hk-2)-l(2hk+2)-(h_2-l)(-2hk-2) \Big|$$

$$0 \le i,j,l,\ \ l \le n,\ \ i+j \le h-n\}$$

$$= \{2h(2ki+kj+2h_2k-2lk-hk+j-1)\}$$

$$= \{2h(ak+b) \mid a = 2i+2h_2-2l-h+j,\ b = j-1\}.$$

We see $-h \le a \le h$. Since h is odd, b is even if and only if a is. Further, since $j = h-a-2i-2h_2+2l$ and $i \le h_1 - j$, we find $0 \le j \le \min\{h_1, h-|a|\}$.

Then the first set becomes:

$$\{2h(ak+b) \mid -h \le a \le h,\ -1 \le b \le \min\{h_1-1, h-|a|-1\}, \text{and } a \equiv b (\text{mod } 2).\}$$

The other set is:

$$\{i(2hk+2)+(h_1-i)(-2hk+2)-j(2h-2)-l(2hk-2)-(h_2-j-l)(-2hk-2)$$

$$0 \le i, j, l,\ i \le h_1,\ j + l \le h_2\}$$

$$= \{2h(2ki - hk + 2kh_2 - 2kl - kj - j + 1)\}$$

$$= \{2h(ak + b) \mid a = 2i + 2h_2 - 2l - h - j,\ b = -j + 1\}.$$

Again, we have $-h \le a \le h$ and b even if and only if a is. Since $j = 2i + 2h_2 - 2l - h - a$ and $l \le h_2 - j$, we find $j \le h - |a|$ and the set is

$$= \{2h(ak + b) \mid\ -h \le a \le h,\ \max\{-h + |a|, -h_2 + 1 \le b \le 1, \text{and } a \equiv b (\text{mod } 2)\}.$$

So, let $h_2 = 0$. We find that the elements of $f_1(A)$ which are multiples of $2h$ are described by:

$$= \{2h(ak + b) \mid -h \le a \le h,\ a \equiv b (\text{mod } 2), \text{and} (b = 1 \text{ or } -1 \le b \le h - |a| - 1)\}.$$

For f_2, where $h_2 = 1$, the set is

$$= \{2h(ak + b) \mid -h \le a \le h,\ a \equiv b (\text{mod } 2), \text{and}$$
$$(0 \le b \le 1 \text{ or } -1 \le b \le \min\{h - |a| - 1, h - 2\})\}.$$

And for f_3, where $h_2 = 2$, it is

$$= \{2h(ak + b) \mid -h \le a \le h,\ a \equiv b (\text{mod } 2), \text{and}$$
$$(-1 \le b \le 1 \text{ or } -1 \le b \le \min\{h - |a| - 1, h - 3\})\}.$$

Then $b = h - 1$, $a = 0$ is possible only when $h_2 = 0$, so that $f_1(A)/f_2(A) = \{2h^2 - 2h\}$.

Similarly, $b = h - 2$, $a = \pm 1$ is not possible when $h_2 = 2$, so $f_1(A)/f_3(A) = \{2h^2 - 2h, \pm 2hk + 2h^2 - 4h\}$.

Additional Observation. I have also noticed that sometimes, adding additional elements to the set A which are congruent to $-2 (\text{mod } 2h)$ allows me to deal with other cases.

For example, let $h = 11$, so that $f_1 = 11A$ and $f_6 = 6A - 5A$, and let $k = 50$. Let $C = B \cup \{-24, -2, 20, 42\}$. Then $f_1(C) = 23,498$ and $f_6(C) = 6C - 5C = 23,437$.

Thus far, I have not been able to systematize such examples, because the computations become too unwieldy.

3.3 A large family

Let $f(A) = A + A + ... + A - A - A - ... - A$. Let $h_1 =$ the number of positive terms and $h_2 =$ number of negative terms. Assume without loss of generality that $h_1 \geq h_2$, and let $h_1 + h_2 = h$.

Definition: A set $A \subset [1, k]$, $1, k, \in A$, is P_n^h if any linear form with coefficients in $\{-1, 1\}$ and h terms contains all but the first n and last n possible elements. That is, $f(A) \supset [h_1(1) - h_2(k) + n, h_1(k) - h_2(1) - n]$. If a set is P_n^2, say that it is P_n.

The idea of P_n sets, as well as the structure of this theorem and proof, are owed to Miller and Scheinerman's proof for the case $h = 2$. [16]

Theorem 3.2. *For any linear forms f, g, with h coefficients in $\{-1, 1\}$, if there exists a finite set of integers A which is P_n^h, with $A \subset [1, 2n]$ and $1, 2n \in A$, and $|f(A)| > |g(A)|$, then there exists an infinite family of such sets.*

Lemma 3.5. *For any linear forms f, g, with h coefficients in $\{0, 1\}$, let $A = L \cup R$ be a P_n^h set, where $L \subset [1, n], R \subset [n + 1, 2n]$. Form $A' = L \cup M \cup R'$, where $M \subset [n + 1, n + m]$*

and $R' = R + m$.

If A' *is a* P_n^h *set, then* $|f(A')| - |f(A)| = |g(A')| - |g(A)|$. *Thus, if* $|f(A)| > |g(A)|$, *the same is true for* A'.

Proof. Since $A \subset [1, 2n]$ and A_n^h, we know that

$$f(A) \subset [h_1 - 2nh_2, 2nh_1 - h_2]$$

and

$$f(A) \supset [h_1 - 2nh_2 + n, 2nh_1 - h_2 - n].$$

Any elements in the right fringe of $f(A)$, $f(A) \cap [h_1 - 2nh_2, h_1 - 2nh_2 + n - 1]$ come only from $L + L + L + ... + L - R - R - R - ... - R$.

Since $A' \subset [1, 2n + m]$, we know

$$f(A') \subset [h_1 - (2n + m)h_2, (2n + m)h_1 - h_2]$$

and

$$[h_1 - (2n + m)h_2 + n, (2n + m)h_1 - h_2 - n] \subset f(A)].$$

Any elements in the left fringe of $f(A')$, that is, $f(A') \cap [h_1 - (2n + m)h_2, h_1 - (2n + m)h_2 + n - 1]$, must come from $L + L + L + ... + L - R' - R' - R' - ... - R'$, which is simply a translation of $L + L + L + ... + L - R - R - R - ... - R$.

A similar argument works for the right fringe of $f(A')$. Then $|f(A')| = |f(A)| + hm$. Since the computation depends only on h, it holds for any pair of forms with h coefficients. □

From Miller and Scheinerman's work, [16] we have:

Lemma 3.6. *Let $A = L \cup R$ be a P_n-set where $L \subset [1, n]$, $R \subset [n+1, 2n]$, and $1, 2n \in A$. Fix a $k \geq n$ and let m be arbitrary. Choose any $M \subset [n+k+1, n+k+m]$ with the property that M does not have a run of more than k missing elements, and form $A(M;k) = L \cup O_1 \cup M \cup O_2 \cup R'$ where $O_1 = [n+1, n+k]$, $O_2 = [n+k+m+1, n+2k+m]$, and $R' = R + 2k + m$. Then $A(M;k)$ is a P_n-set.*

Lemma 3.7. *If A is P_n, and $2n \leq |A|$, then A is also P_n^h.*

Proof. Let A be a P_n set, where $A \subset [1, k]$ and $1, k \in A$. Assume $k \geq 2n$. Thus $A + A \cap [n+2, 2k-n] = [n+2, 2k-n]$.

Let f be a form with h terms. Let $h \geq 3$, else the lemma is trivial, then $h_1 \geq 2$, since,by assumption, $h_1 \geq h_2$. Define $f'(A)$ to be a linear form with $h-2$ coefficients, $h_1 - 2$ positive, and h_2 negative, so that $f'(A) + A + A = f(A)$.

I will show that $f'(\{1, 2n\}) + A + A$ contains all the necessary elements, $[h_1 - kh_2 + n, h_1k - h_2 - n]$. By $f'(\{1, 2n\})$, I mean all numbers of the form $a_1 + a_2 + ... + a_{h_1-2} - a_{h_1-1} - ... - a_{h_1+h_2-2}$, where $a_i \in \{1, 2n\}$.

$$f'(\{1, 2n\}) = \{a(k) + (j_1 - 2 - a)(1) - b(1) - (j_2 - b)(k) \mid 0 \leq a \leq j_1 - 2,\ 0 \leq b \leq j_2\}$$

$$= \{(a+b)k - j_2k + j_1 - 2 - (a+b)\}$$

Setting $i = a + b$, we have

$$f'(\{1, 2n\}) = \{j_1 - 2 - i + k(i - j_2) \mid 0 \leq i \leq j - 2\}.$$

Since A is P_n, $[n+2, 2k-n] \subset A+A$. Then

$$\bigcup_{i=0}^{h-2} [L_i, U_i] \subset f'(\{1, 2n\}) + A + A,$$

where

$$L_i = j_1 - 2 - i + k(i - h_2) + n + 2$$

$$U_i = h_1 - 2 - i + k(i - h_2) + 2k - n.$$

Then we see that $L_0 = h_1 - kh_2 + n$ and $U_{h-2} = h_1 k - h_2 - n$, which are the endpoints of the range we need to cover. It suffices to show the intervals $[L_i, U_i]$ do not have gaps between them.

Since $2n \leq k$, we have:

$$\begin{aligned} L_i - 1 &= h_1 - i + k(i - h_2) + n - 1 \\ &= (h_1 - i + ki - h_2 k - 1) + n \\ &\leq (h_1 - i + ki - h_2 k - 1) + k - n \\ &= h_1 - 2 - (i-1) + k((i-1) - h_2) + 2k - n \\ &\leq U_{i-1}. \end{aligned}$$

Then there are no gaps between the intervals $[L_{i-1}, U_{i-1}]$, $[L_i, U_i]$ and they cover the necessary range. Since the argument holds for any form with h terms, the lemma is proven. $\square$

Note that Lemma 3.7 is not true if the size of n is unrestricted. To take an extreme example, let $A = \{1, 10\}$ and let $n = 9$. Then A is P_n, because $11 \in A + A$, and $0 \in A - A$,

but A is not P_n^3.

Lemma 3.7 shows that the sets described in Lemma 3.6 are also P_n^h. Then, by Lemma 3.5, the sets of Lemma 3.6 also have $|f(A)| > |g(A)|$, and the theorem is proven.

However, the conditions of the theorem call for an initial example to start the process. Currently such examples are known for $h = 2, 3$ only. As shown in [16], let

$$A = \{1, 2, 3, 5, 8, 9, 13, 15, 16\}$$

and

$$B = \{1, 2, 5, 6, 16, 19, 22, 26, 32, 34, 35, 39, 43, 48, 49, 50\}.$$

Then it's easy to check that $|A+A| > |A-A|$ and A is P_n, and that $|B+B+B| > |B+B-B|$, and B is P_n^3.

The nearly symmetric sets described earlier in this chapter are not P_n^h, and hence do not satisfy the conditions of the theorem. The $h = 3$ case required several days of computer search time, checking random sets with density about 1/3. A similar search for the $h = 4$ case ran for a long time without results.

Chapter 4

Phi functions and asymptotic estimates

The Euler phi function, $\phi(n)$ is the number of natural numbers less than n and relatively prime to n. In the case of a prime, for example, $\phi(p) = p - 1$.

For $A \in \mathbb{N}$, let $\gcd(A) = 1$ be the largest number that divides all elements of A. We say that A is relatively prime if $\gcd(A) = 1$. In very recent years, mathematicians have considered the possibility of a phi function for sets, a way of counting sets which are relatively prime. In [11], Nathanson first described several such functions. He defined $f(n)$ to be the number of subsets of $\{1, 2, ..., n\}$ which are relatively prime, and $f_k(n)$ as the number of relatively prime subsets of cardinality k of $\{1, 2, ..., n\}$.

Using the Mobius inversion theorem, he determined that

$$f(n) = \sum_{d=1}^{n} \mu(d)(2^{\lfloor n/d \rfloor} - 1)$$

and

$$f_k(n) = \sum_{d=1}^{n} \mu(d) \binom{\lfloor n/d \rfloor}{k},$$

where $\mu(d)$ is the Mobius function. From those formulas, he found the asymptotic estimates

$$2^n - 2^{\lfloor n/2 \rfloor} - n2^{\lfloor n/3 \rfloor} \leq f(n) \leq 2^n - 2^{\lfloor n/2 \rfloor}$$

and

$$\binom{n}{k} - \binom{\lfloor n/2 \rfloor}{k} - n\binom{\lfloor n/3 \rfloor}{k} \leq f_k(n) \leq \binom{n}{k} - \binom{\lfloor n/2 \rfloor}{k}.$$

Additionally, Nathanson defined $\Phi(n)$ to be the number of nonempty subsets $A \subset \{1, 2, ..., n\}$ such that $\gcd(A)$ is relatively prime to n, and $\Phi_k(n)$ as the number of subsets of $A \subset \{1, 2, ..., n\}$ of cardinality k such that $\gcd(A)$ is relatively prime to n. He found that

$$\Phi(n) = \sum_{d|n} \mu(d) 2^{n/d}$$

and

$$\Phi_k(n) = \sum_{d|n} \mu(d) \binom{n/d}{k}.$$

He found asymptotic estimates for these formulas as well. If n is odd, then

$$\Phi(n) = 2^n + O(n2^{n/3})$$

and

$$\Phi_k(n) = \binom{n}{k} + O\left(n\binom{\lfloor n/3 \rfloor}{k}\right).$$

If n is even,

$$\Phi(n) = 2^n - 2^{n/2} + O(n2^{n/3})$$

and

$$\Phi_k(n) = \binom{n}{k} - \binom{n/2}{k} + O\left(n\binom{\lfloor n/3 \rfloor}{k}\right).$$

Hence, it seems that almost all subsets of the interval $\{1, 2, ..., n\}$ are relatively prime.

In [3], El Bachraoui expanded these functions to count relatively prime subsets of an interval $[m, n]$. He defined $f(m, n)$ to be the number of subsets of $[m, n]$ which are relatively prime, and $f_k(m, n)$ to be the number of such subset which have cardinality k. He also defined $\Phi(m, n)$ as the number of $A \subset [m, n]$ such that $\gcd(A)$ is relatively prime to n, and $\Phi_k(m, n)$ as the number of such sets which have cardinality k. Using an extension of the Mobius inversion formula to several variables, he found:

$$f(m,n) = \sum_{d=1}^{n} \mu(d)(2^{\lfloor n/d \rfloor} - 1) - \sum_{i=1}^{m-1} \sum_{d|i} \mu(d) 2^{\lfloor n/d \rfloor - i/d}, \tag{4.1}$$

$$\Phi(m,n) = \sum_{d|n} \mu(d) 2^{n/d} - \sum_{i=1}^{m-1} \sum_{d|(i,n)} \mu(d) 2^{\frac{n-i}{d}}, \tag{4.2}$$

where $d|(i, n)$ means that d varies over all common divisors of i and n,

$$f_k(m,n) = \sum_{d=1}^{n} \mu(d)\binom{\lfloor n/d \rfloor}{k} - \sum_{i=1}^{m-1} \sum_{d|i} \mu(d)\binom{\lfloor \frac{n}{d} \rfloor - \frac{i}{d}}{k-1}, \tag{4.3}$$

$$\Phi_k(m,n) = \sum_{d|n} \mu(d)\binom{n/d}{k} - \sum_{i=1}^{m-1} \sum_{d|(i,n)} \mu(d)\binom{\frac{n-1}{d}}{k-1}. \tag{4.4}$$

I will simplify his formulas and make asymptotic approximations of them, showing that

almost all subsets of such an interval are relatively prime.

4.1 Simplifying the formulas

We will begin with (4.1). Interchanging the double summation yields:

$$f(m,n) = \sum_{d=1}^{n} \mu(d)(2^{\lfloor n/d \rfloor} - 1) - \sum_{d=1}^{m-1} \sum_{i=kd,\, i<m} \mu(d) 2^{\lfloor n/d \rfloor - i/d}$$

$$= \sum_{d=1}^{n} \mu(d) 2^{\lfloor n/d \rfloor} + \mu(d) - \sum_{d=1}^{m-1} \mu(d) \left[2^{\lfloor n/d \rfloor - d/d} + 2^{\lfloor n/d \rfloor - 2d/d} + ... + 2^{\lfloor n/d \rfloor - \lfloor \frac{m-1}{d} \rfloor} \right].$$

Then, we combine the summations:

$$f(m,n) = \sum_{d=1}^{m-1} \mu(d) \left[2^{\lfloor n/d \rfloor} - \left(2^{\lfloor n/d \rfloor - 1} + 2^{\lfloor n/d \rfloor - 2} + ... + 2^{\lfloor n/d \rfloor - \lfloor m - i/d \rfloor} \right) - 1 \right].$$

$$+ \sum_{d=m}^{n} \mu(d) \left(2^{\lfloor n/d \rfloor} - 1 \right)$$

By repeatedly using the fact that $2^n - 2^{n-1} = 2^{n-1}$, we get

$$f(m,n) = \sum_{d=1}^{n} \mu(d)(2^{\lfloor n/d \rfloor - \lfloor \frac{m-1}{d} \rfloor} - 1) + \sum_{d=m}^{n} \mu(d)(2^{\lfloor n/d \rfloor} - 1).$$

Since $\lfloor \frac{m-1}{i} \rfloor = 0$ for $i > m-1$,

$$f(m,n) = \sum_{i=1}^{n} \mu(i)(2^{\lfloor n/i \rfloor - \lfloor \frac{m-1}{i} \rfloor} - 1). \tag{4.5}$$

Since the structure of (4.3) is identical to that of (4.1), by the combinatorial identity $\binom{n}{k}$ - $\binom{n-1}{k-1} = \binom{n-1}{k}$, the same cancellations give:

$$f_k(m,n) = \sum_{i=1}^{m-1} \mu(i) \binom{\lfloor n/i \rfloor - \lfloor \frac{m-1}{i} \rfloor}{k} + \sum_{i=m}^{n} \mu(i) \binom{\lfloor n/i \rfloor}{k}$$

$$f_k(m,n) = \sum_{i=1}^{n} \mu(i) \binom{\lfloor n/i \rfloor - \lfloor \frac{m-1}{i} \rfloor}{k}.$$

We will perform similar computations on the formulas for Φ and Φ_k.

Starting from (4.2):

$$\Phi(m,n) = \sum_{d|n} \mu(d) 2^{n/d} - \sum_{d=1}^{n} \sum_{i:d|(i,n)}^{m-1} \mu(d) 2^{\frac{n-i}{d}}$$

$$\Phi(m,n) = \sum_{d|n} \mu(d) 2^{n/d} - \sum_{d|n,\, d<m} \mu(d)[2^{\frac{n-d}{d}} + 2^{\frac{n-2d}{d}} + ... + 2^{\frac{n}{d} - \lfloor \frac{m-1}{d} \rfloor}]$$

$$\Phi(m,n) = \sum_{d|n} \mu(d)[2^{n/d} - (2^{\frac{n-d}{d}} + 2^{\frac{n-2d}{d}} + ... + 2^{\frac{n}{d} - \lfloor \frac{m-1}{d} \rfloor})]$$

$$\Phi(m,n) = \sum_{d|n} \mu(d) 2^{\frac{n}{d} - \lfloor \frac{m-1}{d} \rfloor}.$$

Lastly,

$$\Phi_k(m,n) = \sum_{d|n} \mu(i) \binom{\frac{n}{d} - \lfloor \frac{m-1}{i} \rfloor}{k}.$$

4.2 Asymptotics

We will now make some asymptotic estimates of $f(m,n)$.

First, we compute the first few terms of the sequence. Then, I observe that, in equation (4.5) the terms in the summation may not be monotone decreasing. So, I replace it with a

sequence that is. Since $\lfloor \frac{n}{i} \rfloor - \lfloor \frac{m-1}{i} \rfloor \leq \lfloor \frac{n-m+1}{i} \rfloor + 1$,

$$\begin{aligned} f(m,n) &= \sum_{i=1}^{n} \mu(i)(2^{\lfloor \frac{n}{i} \rfloor - \lfloor \frac{m-1}{i} \rfloor} - 1) \\ &= 2^{n-m+1} - 2^{\lfloor \frac{n}{2} \rfloor - \lfloor \frac{m-1}{2} \rfloor} + \sum_{i=3}^{n} \mu(i) 2^{\lfloor \frac{n}{i} \rfloor - \lfloor \frac{m-1}{i} \rfloor} + (n-2) \end{aligned}$$

Replacing the summation with one whose terms are non-increasing:

$$\geq 2^{n-m+1} - 2^{\lfloor \frac{n}{2} \rfloor - \lfloor \frac{m-1}{2} \rfloor} + \sum_{i=3}^{n} \mu(i) 2^{\lfloor \frac{n-m+1}{i} \rfloor + 1}$$

$$\geq 2^{n-m+1} - 2^{\lfloor \frac{n}{2} \rfloor - \lfloor \frac{m-1}{2} \rfloor} - n 2^{\lfloor \frac{n-m+1}{3} \rfloor + 1}$$

Additionally, and with less perspiration, we have:

$$f(m,n) \leq 2^{n-m+1} - 2^{\lfloor \frac{n}{2} \rfloor - \lfloor \frac{m-1}{2} \rfloor} + (n-2)$$

The same type of asymptotic bounds apply to $f_k(m,n)$:

$$\begin{aligned} \binom{n-m+1}{k} - \binom{\lfloor \frac{n}{2} \rfloor - \lfloor \frac{m-1}{2} \rfloor}{k} - n \binom{\lfloor \frac{n-m+1}{3} \rfloor + 1}{k} \\ \leq f_k(m,n) \leq \binom{n-m+1}{k} - \binom{\lfloor \frac{n}{2} \rfloor - \lfloor \frac{m-1}{2} \rfloor}{k}. \end{aligned}$$

For Φ and Φ_k, similar results can be found. Once again, I use $\lfloor \frac{n}{i} \rfloor - \lfloor \frac{m-1}{i} \rfloor \leq \lfloor \frac{n-m+1}{i} \rfloor + 1$ to produce summations with monotone decreasing terms.

If n is even,

$$2^{n-m+1} - 2^{\frac{n}{2} - \lfloor \frac{m-1}{2} \rfloor} - n 2^{\lfloor \frac{n-m+1}{3} \rfloor + 1} \leq \Phi(m,n) \leq 2^{n-m+1} - 2^{\frac{n}{2} - \lfloor \frac{m-1}{2} \rfloor}.$$

Also,

$$\binom{n-m+1}{k} - \binom{\lfloor \frac{n}{2} \rfloor - \lfloor \frac{m-1}{2} \rfloor}{k} - n\binom{\lfloor \frac{n-m+1}{3} \rfloor + 1}{k}$$
$$\leq \Phi_k(m,n) \leq \binom{n-m+1}{k} - \binom{\lfloor \frac{n}{2}] - \lfloor \frac{m-1}{2} \rfloor}{k}.$$

If n is odd,

$$2^{n-m+1} - n2^{\lfloor \frac{n-m+1}{3} \rfloor + 1} \leq \Phi(m,n) \leq 2^{n-m+1} - 2^{\frac{n}{3} - \lfloor \frac{m-1}{3} \rfloor}$$

$$\binom{n-m+1}{k} - n\binom{\lfloor \frac{n-m+1}{3} \rfloor + 1}{k} \leq \Phi_k(m,n) \leq \binom{n-m+1}{k} - \binom{\lfloor \frac{n}{3} \rfloor - \lfloor \frac{m-1}{3} \rfloor}{k}$$

These results are similar to those initially found by Nathanson.

4.3 Latest results

Since my work was published in 2007, [14] additional progress has been made. El Bachraoui has since published a formula [4] for counting subsets A of an interval $[1,m]$ such that $\gcd(A)$ is relatively prime to n, where $n \geq m$. His formulas are:

$$\Phi([1,m],n) = \sum_{d|n} \mu(d) 2^{\lfloor m/d \rfloor},$$
$$\Phi_k([1,m],n) = \sum_{d|n} \mu(d) \binom{\lfloor m/d \rfloor}{k}.$$

He also found asymptotic estimates: If p is the smallest prime divisor of n contained in $[1,m]$, then:

$$0 \leq 2^m - 2^{\lfloor m/p \rfloor} - \Phi([1,m],n) \leq m2^{\lfloor m/p \rfloor},$$

$$0 \leq \binom{m}{k} - \binom{\lfloor m/p]}{k} - \Phi_k([1,m],n) \leq m\binom{\lfloor m/p \rfloor}{k}.$$

Bibliography

[1] Valentin Blomer, *Thin bases of order h*, J. Number Theory **98** (2003), no. 1, 34–46.

[2] J. W. S. Cassels, *Über basen der natürlichen zahlenreihe*, Abh. Math. Sem. Univ. Hamburg **21** (1957), 247–257.

[3] Mohamed El Bachraoui, *The number of relatively prime subsets and phi functions for* $\{m, m+1, \ldots, n\}$, Integers **7** (2007), A43, 8.

[4] ———, *On the number of subsets of* $[1, M]$ *relatively prime to* N *and asymptotic estimates*, Integers **8** (2008), A41, 5.

[5] P. Hegarty and S. Miller, *When almost all sets are difference dominated*, (2008).

[6] Peter V. Hegarty, *Some explicit constructions of sets with more sums than differences*, Acta Arith. **130** (2007), no. 1, 61–77.

[7] Gerd Hofmeister, *Thin bases of order two*, J. Number Theory **86** (2001), no. 1, 118–132.

[8] Xing De Jia and Melvyn B. Nathanson, *A simple construction of minimal asymptotic bases*, Acta Arith. **52** (1989), no. 2, 95–101.

[9] Greg Martin and Kevin O'Bryant, *Many sets have more sums than differences*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 287–305.

[10] M. B. Nathanson, *Supersequences, rearrangements of sequences, and the spectrum of bases in additive number theory*, (2008).

[11] Melvyn B. Nathanson, *Affine invariants, relatively prime sets, and a phi function for subsets of* $\{1, 2, \ldots, n\}$, Integers **7** (2007), A1, 7 pp. (electronic).

[12] ———, *Sets with more sums than differences*, Integers **7** (2007), A5, 24 pp. (electronic).

[13] Melvyn B. Nathanson, Kevin O'Bryant, Brooke Orosz, Imre Ruzsa, and Manuel Silva, *Binary linear forms over finite sets of integers*, Acta Arith. **129** (2007), no. 4, 341–361.

[14] Melvyn B. Nathanson and Brooke Orosz, *Asymptotic estimates for phi functions for subsets of* $\{m + 1, m + 2, \ldots, n\}$, Integers **7** (2007), A54, 5.

[15] D. Raikov, *Über basen der natürlichen zahlenreihe*, Mat. Sbornik N. S. 2 **44** (1937), 595597.

[16] D. Scheinerman S. Miller, B. Orosz, *Explicit constructions of families of mstd sets*, (2008).

[17] A. Stohr, *Eine basis h-ordnung für die menge aller natürlichen zahlen*, Math. Zeit. **42** (1937), 739743.