On Density of Integers and the Sumset

Li Anqi

under the direction of
Hong Wang
Department of Mathematics
Massachusetts Institute of Technology

Research Science Institute
August 2, 2017

Abstract

The Schnirelmann density of a set X of non-negative integers containing 0 is defined as $d(X) = \inf_{n \geq 1} \frac{X(n)}{n}$. Mann's theorem states that for sets A, B of non-negative integers, we have $d(A+B) \geq d(A) + d(B)$. We consider a modified density that accounts for the global average density $d_{\lim}(X) = \lim_{n \to \infty} \inf_{m \geq n} \frac{X(m)}{m}$ and establish an analogue to Mann's Theorem which holds for this modified density: $d_{\lim}(A+B) \geq \max\{d_{\lim}(A), d_{\lim}(B)\} + \frac{\min\{d_{\lim}(A), d_{\lim}(B)\}}{2}$. We also show that this bound is sharp.

Summary

The focus of this paper is on properties of infinite sets of integers and in particular on how dense they are: the proportion of natural numbers that belong in the set. Schnirelmann offered a formula that gives a quantitative measure of density, much like how temperature is a measure of how hot or cold something is. A remarkable property is that when we sum multiple sets together, the density of the sumset is basically given by the sum of the density of each constituent set. However, as it turns out, the Schnirelmann density was susceptible to local fluctuations and may not capture the "big picture" density of an infinite sequence. Thus, we adjusted his definition to capture the global picture. We may then ask: can we find tight bounds on the density of sumsets for the asymptotic density? This is the focus of the paper. We solved this problem for sets having positive asymptotic density. A second question we answered is: what happens when one of our constituent sets is extremely sparse? It turns out that should these sets satisfy certain properties, the sumset constructed from it results in a set of large density – an almost magical result!

1 Introduction

Historically, the birth of additive combinatorics was from examination of classical number theoretic problems through a more combinatorial lens. In the last fifty years, the field of additive combinatorics has blossomed. This is because it has become increasingly clear that it is extremely effective to use combinatorial methods to attack often deep number theory and asymptotic group theory results. In classical number theory it is often the case that one begins with a set of integers – very often the set of primes – and tries to understand how other integers can be written as a sum of elements of the aforementioned set. However, we do the opposite in additive combinatorics: we provide an assumption on the additive properties of a set and then attempt to understand the structure of such sets.

The turning point in the history of additive combinatorics is Schnirelmann's [1] approach to Goldbach's conjecture. Goldbach conjectured that any integer greater than 3 can be expressed as the sum of at most three primes. Schnirelmann managed to show a weaker result that every integer greater than 1 is a sum of a bounded finite number of primes. In more technical terms, he essentially demonstrated that the set of primes form an additive basis for the natural numbers. To this end he broke the problem into two steps: first he demonstrated that sets of integers with positive density form a basis, and then he showed that the integers which can be written as a sum of two primes have positive density. Schnirelmann's work on the weaker Goldbach problem rekindled the interest of the community in the additive properties of sumsets of integers.

We define the sumset of $A, B \subseteq \mathbb{N}$ by $A + B = \{n \in \mathbb{N} \mid n = a + b, a \in A, b \in b\}$. Define $X(n) = |X \cap [1, n]|$ for a set of non-negative integers X. Schnirelmann's lemma states that for $d(X) = \inf\limits_{n \geq 1} \frac{X(n)}{n}$,

$$d(A + B) \geq d(A) + d(B) - d(A)d(B) \tag{1}$$

which gives a relationship between the density of the sumset and the densities of its components.

The following year, L.G. Schnirelmann and L.D. Landau [5] conjectured that the inequality can be strengthened to a much more elegant form

$$d(A+B) \geq d(A) + d(B), \tag{2}$$

provided that $d(A) + d(B) \leq 1$. Notice that by induction, Inequality (2) can be generalized to an arbitrary number of summands: if $\sum_{j=1}^{n} d(X_j) \leq 1$, then $d\left(\sum_{j=1}^{n} X_j\right) \geq \sum_{j=1}^{n} d(X_j)$.

The simplicity and elegance of this problem attracted the attention of many scholars. The biggest breakthrough came through A.Y. Khinchin [2] who established the inequality in the case $d(A) = d(B)$. Many fruitless attempts to generalise this special case followed. Eventually in 1942, this problem was finally fully resolved by H.B. Mann [3], whose proof builds upon the work of A.Y. Khinchin to prove Inequality (2), which then became known as *Mann's Theorem*. The following year in 1943, Artin and Scherk [4] found a different proof of Mann's Theorem that highlighted the structural properties of densities.

The Schnirelmann density has a strange property. Let $X = \mathbb{N}\backslash\{2, \cdots, n\}$ for some finite value n. Intuitively, X contains almost all the natural numbers and should therefore be dense. However, by definition we compute $d(X) = \frac{1}{n}$. We therefore define a new density that is not susceptible to such local fluctuations. In this paper we consider density from a global viewpoint by taking the limit to obtain a new density function $d_{\lim}(X) = \lim_{n\to\infty} \inf_{m \geq n} \frac{X(m)}{m}$. The focus of this paper is to identify properties of $d_{\lim}$ in the hopes of stimulating further research into similarly defined densities.

Unless otherwise mentioned, all sets consists infinitely many non-negative integers and the element 0. In Section 2, we begin by establishing the neccessary conditions for certain special sequences to satisfy the equality case in Mann's Theorem. In Section 3.1, we work on understanding how Mann's Theorem should be modified through a series of special sequences.

In section 3.2 we prove a sharp bound for our modified density and explore the surprising claim that for arbitrary sets A with $d_{\text{lim}}(A) > 0$ and certain sets X such that $d_{\text{lim}}(X) = 0$, $d_{\text{lim}}(A + X) > d_{\text{lim}}(A)$.

2 Equality in Mann's Theorem

We begin by defining our notation. Notice that for an infinite set of integers A, $\frac{A(n)}{n}$ can be thought of as an average gauge of the local density of the sequence in the interval $[1, n]$. For the whole sequence, recall that the *Schnirelmann density* is defined as

$$d(A) = \inf_{n \geq 1} \frac{A(n)}{n}.$$

We consider sets of non-negative integers X such that $d(X) > 0$. In particular, by definition of the Schnirelmann density, this assumption mandates that $1 \in X$. Szeméredi's Theorem states that in any sequence with positive upper density – defined for a set of non-negative integers X as $d_{\text{upper}}(X) = \limsup \frac{X(m)}{m}$ – there exists an infinite arithmetic progression. In the case of d_{lim}, observe that if $A = B = \{0, 1\} \cup \{iq\}_{i=0}^{\infty}$ then $d(A + B) = d(A) + d(B)$. Indeed, it is immediate to see that $d(A) = d(B) = \frac{1}{q}$ while $A + B$ consists of elements $\equiv 0, 1 \pmod{q}$ and as such $d(A + B) = \frac{2}{q} = d(A) + d(B)$. Therefore, we explore if translations on arithmetic progressions would still satisfy the equality case in Mann's Theorem.

Proposition 1. *Suppose $A = \{0, 1, a+x, 2a+x, 3a+x, \ldots\}$ and $B = \{0, 1, b+y, 2b+y, 3b+y, \ldots\}$ are such that $d(A + B) = d(A) + d(B)$ where $x, y \in \mathbb{Z}^+$. Then $a = b$ and $x = y$.*

Proof. The proof is fairly elementary, and due to space constraints can be found in Appendix A. □

However, if we instead consider arithmetic progressions given by consecutive integers, then we restrict our sets to be a union of intervals.

Definition 1. *A draco sequence of size α is of the form $A = \bigcup_{i=0}^{\infty} \left[i^i + 1, i^i + x_{i+1} - x_i\right]$ where $x_i = \lceil \alpha i^i \rceil$.*

Claim 1. *If A is a draco sequence of size α and B is a draco sequence of size β, then $d(A+B) = d(A) + d(B)$.*

Proof. Let $A + B = C$. We prove Claim 1 in two steps. First, we establish that $d(A+B) = \alpha + \beta$. Second, we show that $d(A) = \alpha$ and $d(B) = \beta$.

1. For an element $c \in C \cap [1, n^n]$, decompose it as $c = a + b$ for some $a \in A$ and $b \in B$. Because A is composed of intervals of the form $[i^i + 1, i^i + x_{i+1} - x_i]$, then we have $a \leq (n-1)^{n-1} + x_n - x_{n-1}$. Similarly, $b \leq (n-1)^{n-1} + y_n - y_{n-1}$. Thus $c \leq 2(n-1)^{n-1} + (x_n + y_n) - (x_{n-1} + y_{n-1}) \leq 2(n-1)^{n-1} + (x_n + y_n)$. This gives

   $$d(C) \leq \frac{C(n^n)}{n^n} = \frac{2(n-1)^{n-1} + x_n + y_n}{n^n}.$$

   As $n \to \infty$, $\frac{2(n-1)^{n-1} + x_n + y_n}{n^n} \to \alpha + \beta$. Thus,

   $$d(C) \leq \alpha + \beta. \tag{3}$$

2. Next, let us verify that the densities of A, B as constructed are respectively α, β. By assumption $x_1 = 1$ so we have

   $$d(A) \leq \frac{A(n^n)}{n^n} = \frac{\sum_{i=0}^{n-1} (x_{i+1} - x_i)}{n^n} = \frac{x_n}{n^n}.$$

   As $n \to \infty$, we have that

   $$d(A) \leq \alpha.$$

By the consecutive interval construction, for any n we derive a lower bound as follows:

$$\frac{A(n)}{n} \geq \frac{A(i^i)}{i^i} \geq \alpha, \tag{4}$$

where i^i is the smallest number of its form larger than n. By taking n to be sufficiently large, it follows that $d(A) = \alpha$. Analogously, we derive $d(B) = \beta$. Combining Equation (4) with Equation (3) gives us the desired claim. $\square$

We also notice an interesting relation to $d_{\lim}$ in the draco sequence construction.

Remark 1. *If A and B are draco sequences, $d_{\lim}(A+B) = d(A+B)$, $d_{\lim}(A) = d(A)$ and $d_{\lim}(B) = d(B)$. In other words, we have $d_{\lim}(A+B) = d_{\lim}(A) + d_{\lim}(B)$ which draws a nice parallel to $d(A+B) = d(A) + d(B)$.*

3 On Asymptotic Schnirelmann Density

3.1 Preliminaries

We notice that by Remark 1, because the draco construction gives us sets A, B such that $d_{\lim}(A+B) = d_{\lim}(A) + d_{\lim}(B)$, it is natural to ask whether Mann's Theorem is still valid for $d_{\lim}$. It turns out that the answer to the previous statement is negative.

Definition 2. *A n-**bly AP(a)** sequence is given by $\{0,1\} \cup \left\{ \bigcup_{i=0}^{n-1} \{ka+i\}_{k=1}^{\infty} \right\}$.*

Claim 2. *There exists sets A, B such that $d_{\lim}(A+B) < d_{\lim}(A) + d_{\lim}(B)$.*

Proof. We construct $A = B$ to be two identical 2-bly AP(a) sequences. Then $A + B = \{0,1,2\} \cup \{ka\}_{k=1}^{\infty} \cup \{ka+1\}_{k=1}^{\infty} \cup \{ka+2\}_{k=1}^{\infty}$. It follows that $d_{\lim}(A) = d_{\lim}(B) = \frac{2}{a}$ and $d_{\lim}(A+B) = \frac{3}{a}$ and $\frac{2}{a} + \frac{2}{a} > \frac{3}{a}$. $\square$

As such, it is natural to guess that perhaps the weaker form of Mann's Theorem in the form of Schnirelmann's Lemma given by Inequality (1) is true. However, we demonstrate that this is not the case; Schnirelmann's Lemma does not hold for d_{lim}.

Claim 3. *There exists sets A, B such that $d_{\text{lim}}(A+B) < d_{\text{lim}}(A) + d_{\text{lim}}(B) - d_{\text{lim}}(A)d_{\text{lim}}(B)$.*

Proof. We construct $A = B$ to be two identical 2-bly AP(a). Notice that as $a \to \infty$, $\frac{2}{a} \times \frac{2}{a} \to 0$. Thus, by taking sufficiently large a,

$$\frac{3}{a} < \frac{2}{a} + \frac{2}{a} - \frac{2}{a} \times \frac{2}{a}.$$

$\square$

Thus, we explore the question: to what extent is the asymptotic density of a sumset dependent on its summands? In particular, can we find the sharpest bound for this correlation?

3.2 Sharp Mann-like bounds for d_{lim}

We find an analogue for Mann's Theorem in the case of d_{lim}. One unique property of d_{lim} is that it is *invariant under finite translations*. As such, we can utilize translations to transform our sets A, B to some sets whose Schnirelmann densities approximate respectively $d_{\text{lim}}(A),\ d_{\text{lim}}(B)$.

Without loss of generality, we assume that $d_{\text{lim}}(A) \geq d_{\text{lim}}(B)$. In addition, we also assume that $1 \in A, B$ and $d_{\text{lim}}(A),\ d_{\text{lim}}(B) > 0$.

Theorem 1. *For any sets, $A, B \subset \mathbb{N}$,*

$$d_{\text{lim}}(A+B) \begin{cases} \geq d_{\text{lim}}(A) + \frac{d_{\text{lim}}(B)}{2} & \text{if } 0 < d_{\text{lim}}(A) + d_{\text{lim}}(B) \leq 1 \\ = 1, & \text{otherwise.} \end{cases}$$

The following three lemmas will facilitate our proof of Theorem 2.

Lemma 1. *If $A(n)+B(n)>n-1$, then $n\in A+B$.*

Proof. This lemma holds by the pigeonhole principle. A full proof is provided in Appendix B. □

Lemma 2. *For any $X\subset\mathbb{N}$, $d_{\lim}(X)\geq d(X)$.*

Lemma 3. *For any positive integer c, $d_{\lim}(X+c)=d_{\lim}(X)$.*

Proof. The proofs of Lemmas 2 and 3 are fairly elementary, and due to space constraints can also be found in Appendix B. □

For sake of clarity, we recall the statement of Mann's Theorem

Theorem (Mann's Theorem)**.** *For any sets $A,B\subseteq\mathbb{N}$,*

$$d(A+B)\geq\min\{1,d(A)+d(B)\}.$$

Now we prove Theorem 1.

Proof of Theorem 1. Let $d_{\lim}(A)=\alpha$, $d_{\lim}(B)=\beta$ and $A+B=C$.

Case 1: By Lemma 1 we can show that $d_{\lim}(A+B)=1$ when $d_{\lim}(A)+d_{\lim}(B)>1$. Suppose $\alpha+\beta=1+\eta$ where η is some positive constant. Write $\eta=\eta_1+\eta_2$ with $\eta_1,\eta_2\in\mathbb{R}^+$. Note that there exists a constant c_A such that for all $x\geq c_A$, $\frac{A(x)}{x}\geq\alpha-\eta_1$. Analogously, there exists c_B such that for all $y\geq c_B$, $\frac{B(y)}{y}\geq\beta-\eta_2$. Let $c=\max\{c_A,c_B\}$. Then for any $z>c$, $A(z)+B(z)\geq(\alpha+\beta-\eta_1-\eta_2)z=z>z-1$ so by Lemma 1, $z\in A+B$. That means $A+B$ contains all sufficiently large integers. This implies that $d_{\lim}(A+B)=1$.

Case 2: Now we assume $d_{\lim}(A)+d_{\lim}(B)\leq 1$. In particular, $\alpha,\beta<1$. For any sufficiently small positive constant ε, note the following properties of the set $\mathcal{A}$ of x such that $A(x)\leq x(\alpha-\varepsilon)$:

- $\mathcal{A}$ is of *positive cardinality*. Indeed, $0 \in \mathcal{A}$ as $A(0) \leq 0(\alpha - \varepsilon)$.
- $\mathcal{A}$ is of *finite cardinality*. Notice that there exists a finite constant $K(\varepsilon)$ such that for all $x > K(\varepsilon)$, $\frac{A(x)}{x} > \alpha - \varepsilon$. As such, it follows that all the elements in $\mathcal{A}$ can only be integers from the interval $[1, K(\varepsilon)]$ which demonstrates that it is a finite set.

In particular, this allows us to take k to be the largest element of $\mathcal{A}$. By the maximality of k, it follows that $k+1 \in A$; otherwise, $\frac{A(k+1)}{k+1} = \frac{A(k)}{k+1} < \frac{A(k)}{k} \leq \alpha - \varepsilon$ giving $k+1 \in \mathcal{A}$, contradicting the maximality of k.

Define the set $A' = \{\{A - k\} \cap \mathbb{N}\} \sqcup \{0\}$. In particular, since $k+1 \in A$, $1 \in A'$. When computing the Schnirelmann density of A', notice that $d(A') = \inf_n \frac{|A \cap [k+1, n+k]|}{n} = \inf_n \frac{A(n+k) - A(k)}{n} \geq \alpha - \varepsilon$. Intuitively we constructed the set A' to approximate $d_{\lim}(A)$ and since k might not be an element of A, we append 0 if neccessary to the set A to apply Mann's Theorem. Analogously, there exists a positive integer ℓ such that by doing a similar operation on B we get a set B' such that $d(B') > \alpha - \varepsilon$. Since $\alpha + \beta \leq 1$ by assumption, we can apply Mann's Theorem to A' and B' to obtain $d(A' + B') \geq d(A') + d(B') \geq (\alpha + \beta) - 2\varepsilon$.

By Lemma 2, $d_{\lim}(A' + B') \geq d(A' + B')$. We also have by Lemma 3 that $C' = A' + B' + k + \ell + 1$ is such that $d_{\lim}(C') = d_{\lim}(A' + B')$. Thus, $d_{\lim}(C') \geq d(A' + B') \geq (\alpha + \beta) - 2\varepsilon$.

However, we can observe that $A' + B'$ consists elements of the form $\{a_i - k\}_{a_i > k} \cup \{b_j - \ell\}_{b_j > \ell} \cup \{a_i + b_j - k - \ell\}_{a_i > k, b_j > \ell} \cup \{0\}$. As such, $A' + B' + (k + \ell + 1)$ consists elements of the form $\{a_i + \ell + 1\}_{a_i > k} \cup \{b_j + k + 1\}_{b_j > \ell} \cup \{a_i + b_j + 1\}_{a_i > k, b_j > \ell} \sqcup \{k + \ell + 1\}$.

Since $k+1 \in A$ and $\ell + 1 \in B$ by prior arguments, it follows that $a_i + (\ell + 1) \in C$ and $b_j + k + 1 \in C$. This demonstrates that

$$\{A' + B' + (k + \ell + 1)\} \backslash \{k + \ell + 1\} \subset \{C \cup (C + 1)\}.$$

Denote by $P = \{a_i + b_j + 1 \mid a_i + b_j + 1 \notin C\} = C \backslash C'$. Fix x to be a sufficiently large integer. We can derive the lower bound of $\frac{C(x)}{x}$ from C' – since $d_{\lim}(C')$ can be lower bounded – in the following two ways.

In order to better visualise the counting technique, we represent the sets in a Venn Diagram. In the diagram, we are only looking at the elements of the sets that fall in the interval $[1, n]$ for sufficiently large n. This is a summary and proof sketch in tandem with the visual representation in Figure 1:

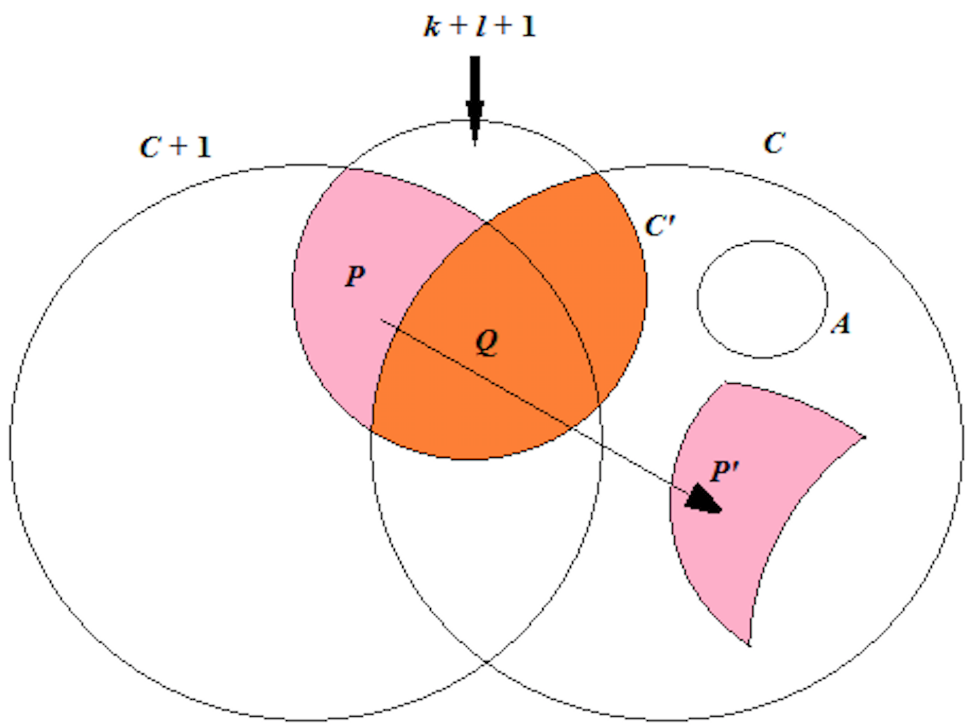

Figure 1: Venn Diagram representing the relationship between the sets in consideration. Here $Q = C' \cap C$ and P' is the image of a map (described later) on P.

1. We construct sets A', B' to be such that $d(A'), d(B')$ are good approximations of $d_{\lim}(A), d_{\lim}(B)$ up to some small factor ϵ. From which, we construct the set $C' = \{A' + B' + (k + \ell + 1)\}$ which we can find a lower bound $d(C') = \alpha + \beta - 2\varepsilon$.

We would like to use this bound on $d(C')$ to estimate $d_{\lim}(C)$. Now decompose C' into P and Q as shown in Figure 1.

2. On one hand, we consider the set Q that is $C \cap C'$ as a lower bound for C.

3. On the other hand, consider the mapping from $C+1$ to C which send P to P' (the bijection is represented by the arrow). P' turns out to be disjoint from A, and thus we can bound C from below by $P' \sqcup A$.

We also make the following remark that $k+\ell+1$ is the only potential element of C' that is disjoint from A' and B'.

– As seen in Figure 1, on one hand, we can bound C by $C \cap C' = C' \backslash \{\{C \backslash C'\} \sqcup \{k+\ell+1\}\} = C' \backslash \{P \sqcup \{k+\ell+1\}\}$. Thus,

$$\frac{C(x)}{x} \geq \frac{C'(x) - P(x) - 1}{x}. \tag{5}$$

– There exists a bijection from $C+1$ to C. Let P map to P' in this bijection. Notice that since all elements a_i+b_j+1 of P are not in C, this means $a_i+b_j \notin A$ as $1 \in B$. Thus, P' is disjoint from A. This allows us to estimate $C(x)$ as $P'(x)+A(x)$. Thus,

$$\frac{C(x)}{x} \geq \frac{A(x) + P'(x)}{x} \geq \frac{A(x) + P(x) - 1}{x}. \tag{6}$$

Summing Equations (5) and (6), we get

$$2\frac{C(x)}{x} \geq \frac{C'(x)}{x} + \frac{A(x)}{x} - \frac{2}{x}. \tag{7}$$

Taking the $\liminf$ in Equation 7 gives

$$d_{\lim}(C) \geq \alpha + \frac{\beta}{2} - \varepsilon$$

which proves the theorem as we can take ε to be arbitrarily small.

$\square$

Remark 2. *It is fruitful to remark that this bound is sharp. Indeed, let A be a 3-bly $AP(a)$ sequence and B be a 2-bly $AP(a)$ sequence. So C is a 4-bly $AP(a)$ sequence. Now, $d_{\lim}(A) = \frac{3}{a}$, $d_{\lim}(B) = \frac{2}{a}$ and $d_{\lim}(A + B) = \frac{4}{a} = \frac{3}{a} + \frac{1}{2} \cdot \frac{2}{a}$.*

Remark 3. *One might be tempted to believe that Artin and Scherk's proof generalizes. But as we have seen in Section 3.1, in fact we cannot claim that an exact equivalent Mann-like bound to d holds in the case of $d_{\lim}$. In their proof, Artin and Scherk repeatedly used the fact that for any y, $\frac{A(y)}{y} \geq d(A)$. Thus, the reason why their proof cannot be replicated for $d_{\lim}$ is because at a particular finite instance, $\inf\limits_{1 \leq x \leq k} \frac{X(k)}{k} > \lim\limits_{n\to\infty} \inf\limits_{m \geq n} \frac{X(m)}{m}$ which renders most discrete arguments fruitless.*

On the other hand, Theorem 1 does not tell us anything about the density of sumsets $A + B$ when the density of one of the summands is 0. In this case, consider the set Q of square integers. Indeed, $d_{\lim}(Q) = \lim\limits_{n\to\infty} \inf\limits_{m \geq n} \frac{1}{\sqrt{m}} = 0$. Since $0 \in Q$, for any set A, it is evident that $A + Q \supset A$ which gives $d_{\lim}(A + Q) \geq d_{\lim}(A)$. However, by utilizing Lagrange's Four Squares Theorem, we are able to obtain a more non-trivial bound.

Proposition 2. *For $A \subset \mathbb{N}$ and $Q = \{n^2\}_{n=0}^{\infty}$*

$$d_{\lim}(A + Q) \geq \left(1 + \frac{d_{\lim}(A)(1 - d_{\lim}(A))}{8}\right) d_{\lim}(A) \tag{8}$$

In what follows, let $C = A + Q$ and $d_{\lim}(A) = \alpha$.

Let $q \in Q$. A method for estimating the density of C is to study how the set $\{A + q\} \backslash A$ behaves. Intuitively, because C contains $A \cup \{A + q\} = A \sqcup \{\{A + q\} \backslash A\}$, we are able to form a lower bound on the density of C. Define a sequence of *holes* $\{h_i\}$ where h_i is the ith smallest positive integer not in A. Let the *set of holes* be denoted H. This gives us the notation to formalise our intuition of maximizing $|\{A+q\}\backslash A|$. Let a hole h_i be *covered by a translate of* q if there exists some $a \in A$ such that $a+q=h_i$.

As a precursor to the proof, we study some properties of such translations. We begin with an overview of what we establish about covering holes.

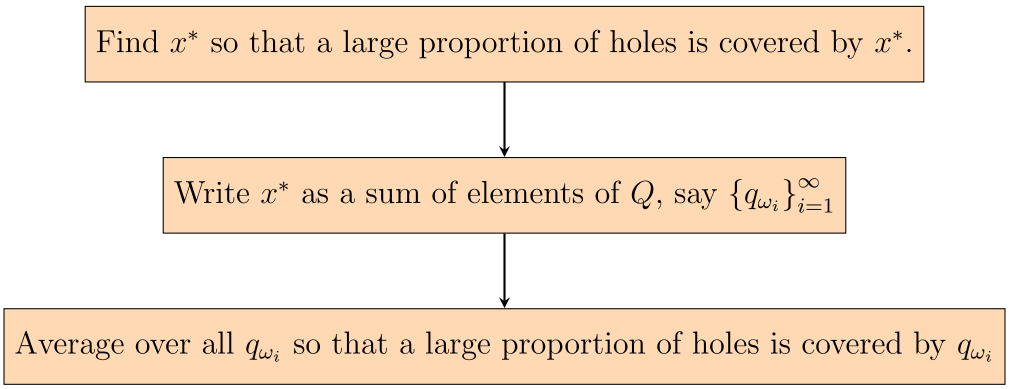

Figure 2: A general roadmap for demonstrating that $|\{A+q\}\backslash A|$ is large.

To get a good measure for the proportion of holes that can be covered by a translate, we need one additional variable: $\delta_c = \sum\limits_{h_i \in \{H \cap [1,c]\}} (h_i - i)$.

Lemma 4. *There exists some $x \in [1,c]$ such that at least $\frac{\delta_c}{c}$ holes in the interval $[1,c]$ are covered by a translate of x.*

Proof. When we examine elements in A of the form $h_j - x$, notice that when we take $x=j$, there are $h_j - j$ elements of A in the interval $[1,h_j]$ as there are exactly j holes in this interval. For an example, refer to Figure 3.

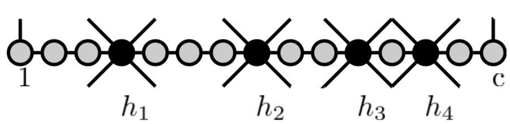

Figure 3: The set represented in this case is given by the black dots; only these natural numbers belong to the set. The grey dots on the number-line represents the holes in the interval $[1,c]=[1,15]$. In this case, $h_1=4, h_2=8, h_3=11, h_4=13$. The only elements of A which are coloured in black that can satisfy $a+x=h_i$ for $i=1,2,3,4$ are those in the interval $[1,13]$ and there are $9=13-4=h_4-4$ of them.

In particular, this allows us to count the number of solutions W to $a = h_j - x$ in two different ways: first, we count by a. Suppose that for each x there are θ_x values of $a \in A$ such that $h_j - x = a$. On one hand, W is given by $\sum_{i=1}^{c} \theta_i$. On the other hand, we can count by $x = h_j - a$, which is equivalent to counting the number of $a \in A$ that are less than h_j. By our previous observation, this is given by $\sum_{h_i \in \{H \cap [1,c]\}} (h_i - i)$. Thus $\sum_{i=1}^{c} \theta_i = W = \sum_{h_i \in \{H \cap [1,c]\}} (h_i - i) = \delta_c$. By the average principle, there exists some x such that $\theta_x \geq \frac{\delta_c}{c}$ holes in the interval $[1, c]$ are covered by a translate of x, as desired.

□

We fix c. Let x^* be the value of x achieving the bound in Lemma 4. We relate this x^* to elements of Q. For this, we need the following well-known result.

Theorem 2 (Lagrange's Theorem)**.** *Every integer can be written as the sum of at most* 4 *squares.*

By Theorem 2, x^* can be written as a sum of finitely many elements of Q, say $\sum_{i=1}^{4} q_{\omega_i} = x^*$. Let the number of holes covered by q_{ω_i} be denoted κ_i.

Lemma 5. *The number of holes covered by $q_{\omega_1} + q_{\omega_2}$ is at most $\kappa_1 + \kappa_2$.*

Proof. Firstly, by definition q_{ω_1} covers κ_1 holes. We can decompose the set $A + q_{\omega_1}$ as follows:

$$A + q_{\omega_1} = \begin{cases} a_i \in A & \\ h_j, & \text{with exactly } \kappa_1 \text{ of this} \end{cases}$$

It follows therefore that:

$$A + q_{\omega_1} + q_{\omega_2} = \begin{cases} a_i + q_{\omega_2} & \text{which generates at most } \kappa_2 \text{ holes} \\ h_j + q_{\omega_2}, & \text{with at most } \kappa_1 \text{ holes} \end{cases}$$

Summing gives the desired.

$\square$

Now we show that we can average the number of holes covered by x^* over each κ.

Lemma 6. *There exists an index s_c such that $\kappa_{s_c} \geq \frac{\delta_c}{4c}$.*

Proof. As a corollary of Lemma 5, it is not difficult to see that the number of holes covered by $x^* = q_{\omega_1} + q_{\omega_2} + q_{\omega_3} + q_{\omega_4}$ is at most $\kappa_1 + \kappa_2 + \kappa_3 + \kappa_4$. By the definition of x^*, it follows that:

$$\kappa_1 + \kappa_2 + \kappa_3 + \kappa_4 \geq \frac{\delta_c}{c}. \tag{9}$$

Apply the average principle in Equation (9) to show that there exists an index s_c such that $\kappa_{s_c} \geq \frac{\delta_c}{4c}$. $\square$

Now we can proceed back to our main proof of Proposition 2.

Proof of Proposition 2. We begin by bounding δ_c. Observe that because the density of A is known, we intuitively are able to describe how the density of $\mathbb{N} \backslash A$ – that is, the set of holes – behaves. This allows us to form a crude bound on how $h_i - i$ behaves. Note that there exists a postive finite constant λ such that $A(x) \geq x\alpha - x\varepsilon$ for all $x > \lambda$. Then for any $h_j \geq \lambda$, $h_j - j = A(h_j) \geq h_j(\alpha - \varepsilon)$. This rearranges to

$$h_j \geq \frac{j}{1 - \alpha + \varepsilon}. \tag{10}$$

Another quick remark is that the set of indices J such that $h_j < \lambda(\varepsilon)$ is finite. This implies that if we bound all the h_i using Inequality 10, we incur at most a finite error term. Additionally, write $\delta_c = \sum_{i=1}^{f(c)} (h_i - i)$ where $f(c) = c - A(c)$. Now, we can bound δ_c as

$$
\begin{aligned}
\delta_c = \sum_{i=1}^{f(c)} (h_i - i) &\geq \left(\frac{1}{1-\alpha+\varepsilon} - 1 \right) \sum_{i=1}^{f(c)} i - T \\
&= \frac{a-\varepsilon}{1-\alpha+\varepsilon} \frac{(c-A(c))(c-A(c)+1)}{2} - T \\
&\geq K(c-A(c))^2 - T
\end{aligned}
$$

where we absorb all the terms independent of the variable c into the constant term K and T is some finite value from the overestimation of $h_j - j$ for indices $j \in J$. Utilizing this, for a sufficiently large fixed n,

$$
\frac{C(n)}{n} \geq \frac{A(n) + (\{A + q_{\omega_c}\} \backslash A)\,(n)}{n} = \frac{A(n) + \frac{\delta_n}{4n}}{n} = \frac{nA(n) + \frac{K(n-A(n))^2}{n}}{4n}. \tag{11}
$$

Now we find a lower bound for this gargantuan expression. Recall that $A(n) \geq n(\alpha - \varepsilon)$. If we treat $A(n) + \frac{K(n-A(n))^2}{n}$ as a quadratic function g of $A(n)$ on the interval $[n(\alpha-\varepsilon), \infty)$, notice that

$$
g'(x) = 1 - \frac{2K(n-x)}{n} \geq 1 - \frac{\alpha - \varepsilon 1 - \alpha + \varepsilon(n - n(\alpha - \varepsilon))}{n} = 1 + \varepsilon - \alpha > 0.
$$

So the function g is increasing on the interval $[n(\alpha-\varepsilon), \infty)$ meaning that $g(x)$ takes its minimum value when $x = n(\alpha - \varepsilon)$. Substituting this into Equation (11), with n sufficiently large and taking ε to be sufficiently small, we can reduce our equation to

$$
d_{\lim}(C) \geq \left(1 + \frac{\alpha(1-\alpha)}{8} \right) \alpha
$$

as desired.

$\square$

4 Conclusion

We study an asymptotic density by adding a limit to the Schnirelmann density and derive sharp bounds of this asymptotic density on the sumset. Additionally, we remark that one case of equality in our theorem is when we take a union of several arithmetic progressions. This motivates us to conjecture the following.

Conjecture 1. *Suppose sets A and B are equality cases in Theorem 1. Then there exists arithmetic progressions X_A and X_B as well as positive constants c_a depending on A and c_b depending on B such that*

$$\lim_{n\to\infty} \inf_{m\geq n} \frac{(X_A \cap A)(m)}{m} \geq c_a d_{\lim}(A), \quad \textit{and} \tag{12}$$

$$\lim_{n\to\infty} \inf_{m\geq n} \frac{(X_B \cap B)(m)}{m} \geq c_b d_{\lim}(A). \tag{13}$$

The core of the conjecture is the problem of given a good estimate about the sumset, what structural properties of the constituent sets can we identify? Arithmetic progressions are fundamentally indestructible structures; not only are these sequences preserved by translation operations on sets, we also usually find their existence in sets for simple congruence reasons.

Another point is that in the proof for Theorem 1 on $d_{\lim}$, we drew parallels to and utilized Mann's Theorem. It would be interesting to further investigate the properties of $d_{\lim}$ to devise a proof independent of Mann's Theorem; while the proof of Mann's Theorem is precise and constructive, the properties of $d_{\lim}$ – particularly that given by Lemma 3 – affords us a lot more flexibility to work with.

5 Acknowledgments

I would like to thank my mentor Ms. Hong Wang of the Massachusetts Institute of Technology Department of Mathematics for her guidance, teaching and inspiration. I would also like to thank Prof. Larry Guth for his support and encouragement as well as Prof. David Jerison, Prof. Ankur Moitra and Dr. Slava Gerovitch for their supervision and coordination. I would also like to express my gratitude to Dr. Tanya Khovanova of the Massachussets Institute of Technology Mathematics Department and my tutor Dr. John Rickert of the Rose-Hulman Institute of Technology for their invaluable insights and input with regards to ideas for my project, writing and presentation. Additionally, I would like to thank Dr. Emma Zbarsky of the Wentworth Institute of Technology Department of Applied Mathematics, Mr. Peter Tian, Mr. Stanislav Atanasov and Ms. Jessie Oehrlein for their indispensable feedback and comments on my paper. Lastly, I am extremely grateful to the Center of Excellence in Education, Massachusetts Institute of Technology, Research Science Institute and the Singapore Ministry of Education for sponsoring me to participate in this wonderful program.

References

[1] L. Schnirelmann. Ob additiwnich swoistwach tschisel. *Izvestiya Donskowo Polytechnitscheskowo Instituta (Nowotscherkask) vol*, 14:3–28, 1930.

[2] E. Landau. *Über einige neuere Fortschritte der additiven Zahlentheorie*. Cambridge University Press Cambridge, 1937.

[3] H. B. Mann. A proof of the fundamental theorem on the density of sums of set of positive integers. *Annals of Mathematics*, pages 523–527, 1942.

[4] E. Artin and P. Scherk. On the sum of two sets of integers. *Annals of Mathematics*, pages 138–142, 1943.

[5] A. Y. Khinchin. Three pearls of number theory. *Nieuw Arch. Wisk*, 15:212–216, 1952.

Appendix A Equality Case in Mann's Theorem

The following is the proof of Proposition 1.

Proof. Without loss of generality, $b+y-1 > a+x-1$. Note that $d(A) = \frac{1}{a+x-1}, d(B) = \frac{1}{b+y-1}$. The first few elements of C are (in some order):

$$0, 1, 2, a+x, a+x+1, 2a+x, 2a+x+1, b+y, b+y+1.$$

Notice that $d(C) = \min\left\{\frac{C(a+x-1)}{a+x-1}, \frac{C(b+y-1)}{b+y-1}\right\}$. Consider the following two cases :

1. $\frac{4}{b+y-1} < \frac{2}{a+x-1}$ which rearranges to $4a+4x-2b-2y < 2$. Thus, $4a+4x-2b-2y$ being a non-negative even integer, must be 0. That is, $b+y = 2(a+x)$. But then the fact that $a+x > 2$ means that

$$\frac{3}{b+y-1} = \frac{3}{2(a+x)-1} > \frac{1}{a+x-1}.$$

   Thus, $d(C) = \frac{4}{b+y-1} = \frac{1}{b+y-1} + \frac{3}{b+y-1} > \frac{1}{a+x-1} + \frac{1}{b+y-1} = d(A) + d(B)$.

2. When $\frac{2}{a+x-1} > \frac{4}{b+y-1}$, notice that $d(C) = \frac{2}{a+x-1} \geq \frac{1}{a+x-1} + \frac{1}{b+y-1} = d(A) + d(B)$ Equality holds when $a+x = b+y$.

$\square$

Appendix B Lemmatas in Preliminaries

The following is the proof of Lemma 1.

Proof. Consider the elements of A in the interval $[1, n-1]$. Suppose these are $a_1, a_2, \ldots, a_k$. Similarly define $b_1, b_2, \ldots, b_\ell$. Consider the at least n numbers $a_1, a_2, \ldots a_k, n-b_1, n-b_2, \ldots, n-b_\ell$. They all belong in the interval $[1, n-1]$ so by the pigeonhole principle there exists two elements that are equal. That is, since all the a_i (respective b_j) are distinct from each other, there exists indices s, t such that $a_s = n - b_t \Rightarrow a_s + b_t = n$. $\square$

The following is the proof of Lemma 2.

Proof. Let the local infimum density of X in $[n, \infty)$ be $d_n(X) = \inf\limits_{m \geq n} \frac{X(m)}{m}$. The increasing property of $d_j(x)$ implies that

$$d(X) = d_1(X) \leq d_2(X) \leq \cdots \leq d_i(X) \leq \cdots \leq \lim_{n \to \infty} d_n(X) = d_{\lim}(X) \tag{14}$$

as desired.

$\square$

The following is the proof of Lemma 3.

Proof. Let $X + c = X'$. Then for any positive integer n, it follows that $X'(n + c) = X(n)$. As such $\frac{X'(n+c)}{n+c} = \frac{X(n)}{n+c}$. Notice that as $n \to \infty$, we have $|\frac{X'(n+c)}{n+c} - \frac{X(n)}{n}| = |\frac{X(n)}{n+c} - \frac{X(n)}{n}| = |\frac{cX(n)}{n(n+c)}| \leq |\frac{cn}{n^2}| = |\frac{c}{n}| \to 0$, which proves the lemma. $\square$