On the Number of Popular Differences

Vsevolod Lev

U Haifa — GA Tech

Toronto, April 8, 2008

(joint work with Sergei Konyagin)

Translation Invariance of Integer Sets

A finite set of elements in a group with torsion can be invariant under non-zero translates; a set of elements in a torsion-free group cannot.

The Problem

To what degree a finite set of *integers* can be translation-invariant?

Also, what are the most translation-invariant sets?
("Sure, arithmetic progressions"?)

The degree of invariance of a set $A \subseteq \mathbb{Z}$ is measured by the function

$$\Delta_A(d) := |(A+d) \setminus A|; \quad d \in \mathbb{Z}$$

showing by how much A "moves out itself" when gets translated by d; considered, say, by Olson in 1968 and by Erdős and Heilbronn in 1964.

Translation Invariance of Integer Sets

A finite set of elements in a group with torsion can be invariant under non-zero translates; a set of elements in a torsion-free group cannot.

The Problem

To what degree a finite set of *integers* can be translation-invariant?

Also, what are the most translation-invariant sets?
("Sure, arithmetic progressions"?)

The degree of invariance of a set $A \subseteq \mathbb{Z}$ is measured by the function

$$\Delta_A(d) := |(A+d) \setminus A|; \quad d \in \mathbb{Z}$$

showing by how much A "moves out itself" when gets translated by d; considered, say, by Olson in 1968 and by Erdős and Heilbronn in 1964.

Translation Invariance of Integer Sets

A finite set of elements in a group with torsion can be invariant under non-zero translates; a set of elements in a torsion-free group cannot.

The Problem

To what degree a finite set of *integers* can be translation-invariant?

Also, what are the most translation-invariant sets?
("Sure, arithmetic progressions"?)

The degree of invariance of a set $A \subseteq \mathbb{Z}$ is measured by the function

$$\Delta_A(d) := |(A+d) \setminus A|; \quad d \in \mathbb{Z}$$

showing by how much A "moves out itself" when gets translated by d; considered, say, by Olson in 1968 and by Erdős and Heilbronn in 1964.

The Properties of the Olson-Erdős-Heilbronn function

$$\Delta_A(d) := |(A+d) \setminus A|; \quad d \in \mathbb{Z}$$

Basic properties of the function Δ_A:

- $\Delta_A(0) = 0$;
- $\Delta_A(-d) = \Delta_A(d)$;
- $\Delta_A(d_1 + d_2) \leq \Delta_A(d_1) + \Delta_A(d_2)$, whence $\Delta_A(hd) \leq h\Delta_A(d)$.

Furthermore,

- $\Delta_A(d) = |A| - \nu_A(d)$, where $\nu_A(d)$ is the number of representations of d as a difference of two elements of A;
- $\Delta_A(d)$ is the minimal number of arithmetic progressions with difference d into which A can be partitioned.

The Properties of the Olson-Erdős-Heilbronn function

$$\Delta_A(d) := |(A+d) \setminus A|; \quad d \in \mathbb{Z}$$

Basic properties of the function Δ_A:

- $\Delta_A(0) = 0$;
- $\Delta_A(-d) = \Delta_A(d)$;
- $\Delta_A(d_1 + d_2) \le \Delta_A(d_1) + \Delta_A(d_2)$, whence $\Delta_A(hd) \le h\Delta_A(d)$.

Furthermore,

- $\Delta_A(d) = |A| - \nu_A(d)$, where $\nu_A(d)$ is the number of representations of d as a difference of two elements of A;
- $\Delta_A(d)$ is the minimal number of arithmetic progressions with difference d into which A can be partitioned.

How Many Small Values can Δ_A Attain?

We seek to show that Δ_A does not assume too many small values: the "enemy" gives us a set D, we try to select $d \in D$ with $\Delta_A(d)$ large.

As $\Delta_A(-d) = \Delta_A(d)$, we assume $d > 0$ whenever convenient. Easy:

- there is at most one $d \in \mathbb{N}$ with $\Delta_A(d) \leq 1$; moreover, for such d to exist, A must be an arithmetic progression;
- there are at most two $d \in \mathbb{N}$ with $\Delta_A(d) \leq 2$; moreover, for *two* such d to exist, A must be an arithmetic progression or a progression with the second smallest / largest element deleted.

(Thus, given $D \subseteq \mathbb{N}$ with $|D| \geq 2$, we can find $d \in D$ with $\Delta_A(d) \geq 2$; if $|D| \geq 3$, we can find $d \in D$ with $\Delta_A(d) \geq 3$ — provided $|A| \geq 3$.)

Messy:

How many $d \in \mathbb{N}$ can there be with $\Delta_A(d) \leq 4$? With $\Delta_A(d) \leq 5$?

How Many Small Values can Δ_A Attain?

We seek to show that Δ_A does not assume too many small values: the "enemy" gives us a set D, we try to select $d \in D$ with $\Delta_A(d)$ large.

As $\Delta_A(-d) = \Delta_A(d)$, we assume $d > 0$ whenever convenient. Easy:

- there is at most one $d \in \mathbb{N}$ with $\Delta_A(d) \le 1$;
  moreover, for such d to exist, A must be an arithmetic progression;
- there are at most two $d \in \mathbb{N}$ with $\Delta_A(d) \le 2$; moreover, for *two* such d to exist, A must be an arithmetic progression or a progression with the second smallest / largest element deleted.

(Thus, given $D \subseteq \mathbb{N}$ with $|D| \ge 2$, we can find $d \in D$ with $\Delta_A(d) \ge 2$; if $|D| \ge 3$, we can find $d \in D$ with $\Delta_A(d) \ge 3$ — provided $|A| \ge 3$.)

Messy:

How many $d \in \mathbb{N}$ can there be with $\Delta_A(d) \le 4$? With $\Delta_A(d) \le 5$?

How Many Small Values can Δ_A Attain?

We seek to show that Δ_A does not assume too many small values: the "enemy" gives us a set D, we try to select $d \in D$ with $\Delta_A(d)$ large.

As $\Delta_A(-d) = \Delta_A(d)$, we assume $d > 0$ whenever convenient. Easy:

- there is at most one $d \in \mathbb{N}$ with $\Delta_A(d) \le 1$;
  moreover, for such d to exist, A must be an arithmetic progression;
- there are at most two $d \in \mathbb{N}$ with $\Delta_A(d) \le 2$; moreover, for *two* such d to exist, A must be an arithmetic progression or a progression with the second smallest / largest element deleted.

(Thus, given $D \subseteq \mathbb{N}$ with $|D| \ge 2$, we can find $d \in D$ with $\Delta_A(d) \ge 2$; if $|D| \ge 3$, we can find $d \in D$ with $\Delta_A(d) \ge 3$ — provided $|A| \ge 3$.)

Messy:

How many $d \in \mathbb{N}$ can there be with $\Delta_A(d) \le 4$? With $\Delta_A(d) \le 5$?

How Many Small Values can Δ_A Attain?

We seek to show that Δ_A does not assume too many small values: the "enemy" gives us a set D, we try to select $d \in D$ with $\Delta_A(d)$ large.

As $\Delta_A(-d) = \Delta_A(d)$, we assume $d > 0$ whenever convenient. Easy:

- there is at most one $d \in \mathbb{N}$ with $\Delta_A(d) \le 1$; moreover, for such d to exist, A must be an arithmetic progression;
- there are at most two $d \in \mathbb{N}$ with $\Delta_A(d) \le 2$; moreover, for *two* such d to exist, A must be an arithmetic progression or a progression with the second smallest / largest element deleted.

(Thus, given $D \subseteq \mathbb{N}$ with $|D| \ge 2$, we can find $d \in D$ with $\Delta_A(d) \ge 2$; if $|D| \ge 3$, we can find $d \in D$ with $\Delta_A(d) \ge 3$ — provided $|A| \ge 3$.)

Messy:

How many $d \in \mathbb{N}$ can there be with $\Delta_A(d) \le 4$? With $\Delta_A(d) \le 5$?

The Behavior in Average

If A is a block of consecutive integers, then for every $1 \le m < |A|$ there is exactly one $d \in \mathbb{N}$ with $\Delta_A(d) = m$; thus, there are exactly m positive integers d with $\Delta_A(d) \le m$.

This turns out to be the "worst case in average":

Theorem (Gabriel 1932, extending Hardy-Littlewood 1928)

For any finite sets $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ we have

$$\frac{1}{|D|}\sum_{d=1}^{|D|}\Delta_{[1,|A|]}(d) \le \frac{1}{|D|}\sum_{d\in D}\Delta_A(d).$$

That is, for $|A|$ and $|D|$ prescribed, the sum $\sum_{d\in D}\Delta_A(d)$ gets minimized when $A = [1,|A|]$ and $D = [1,|D|]$.

The Behavior in Average

If A is a block of consecutive integers, then for every $1 \leq m < |A|$ there is exactly one $d \in \mathbb{N}$ with $\Delta_A(d) = m$; thus, there are exactly m positive integers d with $\Delta_A(d) \leq m$.

This turns out to be the "worst case in average":

Theorem (Gabriel 1932, extending Hardy-Littlewood 1928)

For any finite sets $A \subseteq \mathbb{Z},\ D \subseteq \mathbb{N}$ we have

$$\frac{1}{|D|} \sum_{d=1}^{|D|} \Delta_{[1,|A|]}(d) \leq \frac{1}{|D|} \sum_{d \in D} \Delta_A(d).$$

That is, for $|A|$ and $|D|$ prescribed, the sum $\sum_{d \in D} \Delta_A(d)$ gets minimized when $A = [1, |A|]$ and $D = [1, |D|]$.

From Average to Pointwise

In other words: for every $m \geq 1$, the average of the m smallest values of Δ_A is minimized when A is a block of consecutive integers; more generally, when A an arithmetic progression.

Are arithmetic progressions optimal pointwise?

Let

$$\mu_A(D) := \max_{d \in D} \Delta_A(d); \quad A, D \subseteq \mathbb{Z}.$$

By Gabriel,

$$\mu_A(D) \geq \frac{1}{|D|} \sum_{d=1}^{|D|} \Delta_{[1,|A|]}(d) = \frac{1}{|D|} \sum_{d=1}^{|D|} d = \frac{1}{2}(|D|+1),$$

provided that $|D| \leq |A|$. (If $d > |A|$, then $\Delta_A(d) = |A| \neq d$.)

From Average to Pointwise

In other words: for every $m \geq 1$, the average of the m smallest values of Δ_A is minimized when A is a block of consecutive integers; more generally, when A an arithmetic progression.

Are arithmetic progressions optimal pointwise?

Let

$$\mu_A(D) := \max_{d \in D} \Delta_A(d); \quad A, D \subseteq \mathbb{Z}.$$

By Gabriel,

$$\mu_A(D) \geq \frac{1}{|D|} \sum_{d=1}^{|D|} \Delta_{[1,|A|]}(d) = \frac{1}{|D|} \sum_{d=1}^{|D|} d = \frac{1}{2}\left(|D| + 1\right),$$

provided that $|D| \leq |A|$. (If $d > |A|$, then $\Delta_A(d) = |A| \neq d$.)

Beating Arithmetic Progressions

$$\Delta_A(d) = |(A+d) \setminus A|, \quad \mu_A(D) = \max_{d \in D} \Delta_A(d); \; A, D \subseteq \mathbb{Z}$$

If A is an AP, then $\mu_A(D) \geq |D|$ for any $D \subseteq \mathbb{N}$ with $|D| \leq |A|$.

Is it true that $\mu_A(D) \geq |D|$ for any $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ (with $|D| \leq |A|$)?

For an integer $m > 2$, let

$$A := \bigcup_{0 \leq k < \log_2 m} [km, (k+1)m - 2^k).$$

Then $\Delta_A(d) \leq m-1$ for every $d \in [1, m]$; that is, for $D = [1, m]$ we have $\mu_A(D) < |D|$ — whereas $|D| = m \sim |A|/\log|A|$!

Beating Arithmetic Progressions

$$\Delta_A(d) = |(A+d) \setminus A|, \quad \mu_A(D) = \max_{d \in D} \Delta_A(d); \; A, D \subseteq \mathbb{Z}$$

If A is an AP, then $\mu_A(D) \geq |D|$ for any $D \subseteq \mathbb{N}$ with $|D| \leq |A|$.

Is it true that $\mu_A(D) \geq |D|$ for any $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ (with $|D| \leq |A|$)?

No!

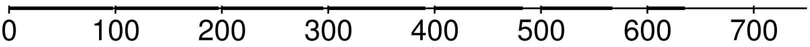
For an integer $m > 2$, let

$$A := \bigcup_{0 \leq k < \log_2 m} [km, (k+1)m - 2^k).$$

Then $\Delta_A(d) \leq m - 1$ for every $d \in [1, m]$; that is, for $D = [1, m]$ we have $\mu_A(D) < |D|$ — whereas $|D| = m \sim |A| / \log |A|$!

The Interpretation

$m = 100$:



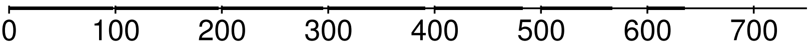
For long time we believed that the answer is "ALMOST "YES":

There is an absolute constant $c > 0$ such that $\mu_A(D) \geq |D|$ holds for all finite sets $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ with $|D| < c|A|$.

The right interpretation of the example above: $|D| \leq c|A|$ is *insufficient* for $\mu_A(D) \geq |D|$ to hold, a stronger assumption is needed!

The Interpretation

$m = 100$:



For long time we believed that the answer is "ALMOST "YES":

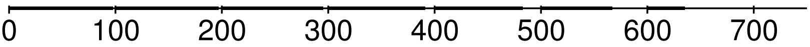
A Wrong Theorem

There is an absolute constant $c > 0$ such that $\mu_A(D) \geq |D|$ holds for all finite sets $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ with $|D| < c|A|$.

The right interpretation of the example above: $|D| \leq c|A|$ is *insufficient* for $\mu_A(D) \geq |D|$ to hold, a stronger assumption is needed!

The Interpretation

$m = 100$:



For long time we believed that the answer is “ALMOST “YES”:

A Wrong Theorem

There is an absolute constant $c > 0$ such that $\mu_A(D) \geq |D|$ holds for all finite sets $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ with $|D| < c|A|$.

The right interpretation of the example above: $|D| \leq c|A|$ is *insufficient* for $\mu_A(D) \geq |D|$ to hold, a stronger assumption is needed!

The Main Result

Turns out that $|D| < c|A| / \log |A|$ is sufficient:

The *True* Theorem (Konyagin, Lev)

There is an absolute constant $c > 0$ such that $\mu_A(D) \geq |D|$ holds for all finite sets $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ with $|D| < c|A| / \log |A|$.

- Both $\mu_A(D) \geq |D|$ and $|D| < c|A| / \log |A|$ are best possible, as shown by the AP example and the "logarithmic example".

A simple proof can be given if the assumption is strengthened:

The $\sqrt{|A|}$-Theorem

We have $\mu_A(D) \geq |D|$ for all finite sets $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ with $|D| \leq \sqrt{|A|}$.

The Main Result

Turns out that $|D| < c|A|/\log|A|$ is sufficient:

The *True* Theorem (Konyagin, Lev)

There is an absolute constant $c > 0$ such that $\mu_A(D) \geq |D|$ holds for all finite sets $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ with $|D| < c|A|/\log|A|$.

- Both $\mu_A(D) \geq |D|$ and $|D| < c|A|/\log|A|$ are best possible, as shown by the AP example and the "logarithmic example".

A simple proof can be given if the assumption is strengthened:

The $\sqrt{|A|}$-Theorem

We have $\mu_A(D) \geq |D|$ for all finite sets $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ with $|D| \leq \sqrt{|A|}$.

The $\sqrt{|A|}$-Theorem

We have $\mu_A(D) \ge |D|$ for all finite sets $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ with $|D| \le \sqrt{|A|}$.

Proof of the $\sqrt{|A|}$-Theorem.

$$d_1, \dots, d_m \in \mathbb{N},\ m \le \sqrt{|A|} \quad \overset{?}{\Rightarrow} \quad \Delta_A(d_i) \ge m \text{ for some } i \in [1, m]$$

For a contradiction, suppose that $\Delta_A(d_i) \le m - 1$ for $i = 1, \dots, m$; thus, A is a union of at most $m - 1$ AP with difference d_i, for each i.

At least one of these AP has m or more terms (as $(m-1)^2 < |A|$); say, $a + kd_i \in A$ for $k = 1, \dots, m$. But A is also a union of at most $m - 1$ AP with difference d_j! Hence, $a + k_1 d_i \equiv a + k_2 d_i \pmod{d_j}$ for some $k_1, k_2 \in [1, m]$, $k_1 \ne k_2$.

This yields $d_j \mid (k_2 - k_1) d_i$, implying $d_j / \gcd(d_i, d_j) \mid k_2 - k_1$ and, consequently, $d_j / \gcd(d_i, d_j) \le m - 1$, contradicting "Graham's g.c.d. conjecture"! $\square$

The $\sqrt{|A|}$-Theorem

We have $\mu_A(D) \geq |D|$ for all finite sets $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ with $|D| \leq \sqrt{|A|}$.

Proof of the $\sqrt{|A|}$-Theorem.

$$d_1, \dots, d_m \in \mathbb{N},\ m \leq \sqrt{|A|} \quad \stackrel{?}{\Rightarrow} \quad \Delta_A(d_i) \geq m \text{ for some } i \in [1, m]$$

For a contradiction, suppose that $\Delta_A(d_i) \leq m - 1$ for $i = 1, \dots, m$; thus, A is a union of at most $m - 1$ AP with difference d_i, for each i.

At least one of these AP has m or more terms (as $(m-1)^2 < |A|$); say, $a + kd_i \in A$ for $k = 1, \dots, m$. But A is also a union of at most $m - 1$ AP with difference d_j! Hence, $a + k_1 d_i \equiv a + k_2 d_i \pmod{d_j}$ for some $k_1, k_2 \in [1, m],\ k_1 \neq k_2$.

This yields $d_j \mid (k_2 - k_1) d_i$, implying $d_j / \gcd(d_i, d_j) \mid k_2 - k_1$ and, consequently, $d_j / \gcd(d_i, d_j) \leq m - 1$, contradicting “Graham's g.c.d. conjecture”! □

The $\sqrt{|A|}$-Theorem

We have $\mu_A(D) \ge |D|$ for all finite sets $A \subseteq \mathbb{Z}$, $D \subseteq \mathbb{N}$ with $|D| \le \sqrt{|A|}$.

Proof of the $\sqrt{|A|}$-Theorem.

$$d_1, \dots, d_m \in \mathbb{N},\ m \le \sqrt{|A|} \quad \overset{?}{\Rightarrow} \quad \Delta_A(d_i) \ge m \text{ for some } i \in [1, m]$$

For a contradiction, suppose that $\Delta_A(d_i) \le m - 1$ for $i = 1, \dots, m$; thus, A is a union of at most $m - 1$ AP with difference d_i, for each i.

At least one of these AP has m or more terms (as $(m-1)^2 < |A|$); say, $a + kd_i \in A$ for $k = 1, \dots, m$. But A is also a union of at most $m - 1$ AP with difference d_j! Hence, $a + k_1 d_i \equiv a + k_2 d_i \pmod{d_j}$ for some $k_1, k_2 \in [1, m],\ k_1 \ne k_2$.

This yields $d_j \mid (k_2 - k_1) d_i$, implying $d_j / \gcd(d_i, d_j) \mid k_2 - k_1$ and, consequently, $d_j / \gcd(d_i, d_j) \le m - 1$, contradicting “Graham's g.c.d. conjecture”! $\square$

The Main Lemma

An important particular case of the Main Theorem, from which the general result is derived, is the case $D = [1, m]$.

The Main Lemma

There is an absolute constant $C > 0$ such that $\mu_A([1, m]) \geq m$ holds for every finite set $A \subseteq \mathbb{Z}$ with $|A| > Cm \log m$.

Plain-terms restatement, avoiding non-standard notation:

if $|A| > Cm \log m$, then there exists $d \in [1, m]$ with $|(A + d) \setminus A| \geq m$.

The "Deduction Toolbox":

- $\mu_A(hD) \leq h\mu_A(D)$ (recall $\Delta_A(d_1 + d_2) \leq \Delta_A(d_1) + \Delta_A(d_2)$!);
- $\mu_A(D) \geq (|D| + 1)/2$ for $|D| \leq |A|$;
- monotonicity: if $D \subseteq C$, then $\mu_A(D) \leq \mu_A(C)$;
- estimates of $|hA|$ and results on the structure of hA.

The Main Lemma

An important particular case of the Main Theorem, from which the general result is derived, is the case $D = [1, m]$.

The Main Lemma

There is an absolute constant $C > 0$ such that $\mu_A([1, m]) \geq m$ holds for every finite set $A \subseteq \mathbb{Z}$ with $|A| > Cm \log m$.

Plain-terms restatement, avoiding non-standard notation:

if $|A| > Cm \log m$, then there exists $d \in [1, m]$ with $|(A + d) \setminus A| \geq m$.

The "Deduction Toolbox":

- $\mu_A(hD) \leq h\mu_A(D)$ (recall $\Delta_A(d_1 + d_2) \leq \Delta_A(d_1) + \Delta_A(d_2)$!);
- $\mu_A(D) \geq (|D| + 1)/2$ for $|D| \leq |A|$;
- monotonicity: if $D \subseteq C$, then $\mu_A(D) \leq \mu_A(C)$;
- estimates of $|hA|$ and results on the structure of hA.

Deduction of the Main Theorem from the Main Lemma

Let $D \subseteq \mathbb{N}$ and suppose that $A \subseteq \mathbb{Z}$ is "large", while $\mu_A(D) < |D|$.
The idea: if D is unstructured, then the sumsets hD grow fast;
hence $\mu_A(hD)$ are large, and so is $\mu_A(D) \geq h^{-1}\mu_A(hD)$:

$$\frac{1}{2}\,|hD| < \mu_A(hD) \leq h\mu_A(D) < h|D|,$$

whence

$$|hD| < 2h|D|.$$

It does not follows that D is "close" to $[1, m]$, and even not that D is dense; however, it follows that hD is dense and consequently, $hD - hD \supseteq [1, |hD| - 1]$ (provided $\gcd(D) = 1$, as we assume).
Now we use monotonicity and the Main Lemma:

$$\mu_A(hD - hD) \geq \mu_A([1, |hD| - 1]) \geq |hD| - 1$$

while, on the other hand,

$$\mu_A(hD - hD) \leq 2h\mu_A(D) < 2h|D|.$$

Deduction of the Main Theorem from the Main Lemma

Let $D \subseteq \mathbb{N}$ and suppose that $A \subseteq \mathbb{Z}$ is "large", while $\mu_A(D) < |D|$.
The idea: if D is unstructured, then the sumsets hD grow fast;
hence $\mu_A(hD)$ are large, and so is $\mu_A(D) \geq h^{-1}\mu_A(hD)$:

$$\frac{1}{2}|hD| < \mu_A(hD) \leq h\mu_A(D) < h|D|,$$

whence

$$|hD| < 2h|D|.$$

It does not follows that D is "close" to $[1, m]$, and even not that D is dense; however, it follows that hD is dense and consequently, $hD - hD \supseteq [1, |hD| - 1]$ (provided $\gcd(D) = 1$, as we assume).
Now we use monotonicity and the Main Lemma:

$$\mu_A(hD - hD) \geq \mu_A([1, |hD| - 1]) \geq |hD| - 1$$

while, on the other hand,

$$\mu_A(hD - hD) \leq 2h\mu_A(D) < 2h|D|.$$

Deduction of the Main Theorem from the Main Lemma

Let $D \subseteq \mathbb{N}$ and suppose that $A \subseteq \mathbb{Z}$ is "large", while $\mu_A(D) < |D|$.
The idea: if D is unstructured, then the sumsets hD grow fast;
hence $\mu_A(hD)$ are large, and so is $\mu_A(D) \geq h^{-1}\mu_A(hD)$:

$$\frac{1}{2}|hD| < \mu_A(hD) \leq h\mu_A(D) < h|D|,$$

whence

$$|hD| < 2h|D|.$$

It does not follows that D is "close" to $[1, m]$, and even not that D is dense; however, it follows that hD is dense and consequently, $hD - hD \supseteq [1, |hD| - 1]$ (provided $\gcd(D) = 1$, as we assume).
Now we use monotonicity and the Main Lemma:

$$\mu_A(hD - hD) \geq \mu_A([1, |hD| - 1]) \geq |hD| - 1$$

while, on the other hand,

$$\mu_A(hD - hD) \leq 2h\mu_A(D) < 2h|D|.$$

Deduction of the Main Theorem from the Main Lemma

Let $D \subseteq \mathbb{N}$ and suppose that $A \subseteq \mathbb{Z}$ is “large”, while $\mu_A(D) < |D|$.
The idea: if D is unstructured, then the sumsets hD grow fast; hence $\mu_A(hD)$ are large, and so is $\mu_A(D) \geq h^{-1}\mu_A(hD)$:

$$\frac{1}{2}\,|hD| < \mu_A(hD) \leq h\mu_A(D) < h|D|,$$

whence

$$|hD| < 2h|D|.$$

It does not follows that D is “close” to $[1, m]$, and even not that D is dense; however, it follows that hD is dense and consequently, $hD - hD \supseteq [1, |hD| - 1]$ (provided $\gcd(D) = 1$, as we assume).
Now we use monotonicity and the Main Lemma:

$$\mu_A(hD - hD) \geq \mu_A([1, |hD| - 1]) \geq |hD| - 1$$

while, on the other hand,

$$\mu_A(hD - hD) \leq 2h\mu_A(D) < 2h|D|.$$

Deduction of the Main Theorem from the Main Lemma

Let $D \subseteq \mathbb{N}$ and suppose that $A \subseteq \mathbb{Z}$ is "large", while $\mu_A(D) < |D|$.
The idea: if D is unstructured, then the sumsets hD grow fast; hence $\mu_A(hD)$ are large, and so is $\mu_A(D) \geq h^{-1}\mu_A(hD)$:

$$\frac{1}{2}|hD| < \mu_A(hD) \leq h\mu_A(D) < h|D|,$$

whence

$$|hD| < 2h|D|.$$

It does not follows that D is "close" to $[1, m]$, and even not that D is dense; however, it follows that hD is dense and consequently, $hD - hD \supseteq [1, |hD| - 1]$ (provided $\gcd(D) = 1$, as we assume).
Now we use monotonicity and the Main Lemma:

$$\mu_A(hD - hD) \geq \mu_A([1, |hD| - 1]) \geq |hD| - 1$$

while, on the other hand,

$$\mu_A(hD - hD) \leq 2h\mu_A(D) < 2h|D|.$$

Deduction of the Main Theorem from the Main Lemma

Let $D \subseteq \mathbb{N}$ and suppose that $A \subseteq \mathbb{Z}$ is "large", while $\mu_A(D) < |D|$.
The idea: if D is unstructured, then the sumsets hD grow fast; hence $\mu_A(hD)$ are large, and so is $\mu_A(D) \geq h^{-1}\mu_A(hD)$:

$$\tfrac{1}{2}\,|hD| < \mu_A(hD) \leq h\mu_A(D) < h|D|,$$

whence

$$|hD| < 2h|D|.$$

It does not follows that D is "close" to $[1, m]$, and even not that D is dense; however, it follows that hD is dense and consequently, $hD - hD \supseteq [1, |hD| - 1]$ (provided $\gcd(D) = 1$, as we assume).
Now we use monotonicity and the Main Lemma:

$$\mu_A(hD - hD) \geq \mu_A([1, |hD| - 1]) \geq |hD| - 1$$

while, on the other hand,

$$\mu_A(hD - hD) \leq 2h\mu_A(D) < 2h|D|.$$

Deduction of the Main Theorem from the Main Lemma

Let $D \subseteq \mathbb{N}$ and suppose that $A \subseteq \mathbb{Z}$ is "large", while $\mu_A(D) < |D|$.
The idea: if D is unstructured, then the sumsets hD grow fast; hence $\mu_A(hD)$ are large, and so is $\mu_A(D) \geq h^{-1}\mu_A(hD)$:

$$\frac{1}{2}|hD| < \mu_A(hD) \leq h\mu_A(D) < h|D|,$$

whence

$$|hD| < 2h|D|.$$

It does not follows that D is "close" to $[1, m]$, and even not that D is dense; however, it follows that hD is dense and consequently, $hD - hD \supseteq [1, |hD| - 1]$ (provided $\gcd(D) = 1$, as we assume).
Now we use monotonicity and the Main Lemma:

$$\mu_A(hD - hD) \geq \mu_A([1, |hD| - 1]) \geq |hD| - 1$$

while, on the other hand,

$$\mu_A(hD - hD) \leq 2h\mu_A(D) < 2h|D|.$$

The Real Deduction, I

To make this approach work, we consider the set

$$D^{\pm} := (-D) \cup \{0\} \cup D$$

instead of D: it grows faster, while $\mu_A(D^{\pm}) = \mu_A(D)$.

If $\mu_A(D) < |D|$, then (as above) we get

$$|hD^{\pm}| < 2h|D^{\pm}|$$

implying

$$2hD^{\pm} = hD^{\pm} - hD^{\pm} \supseteq [1, |hD^{\pm}| - 1].$$

By monotonicity and the Main Lemma,

$$\mu_A(2hD^{\pm}) \geq \mu_A([1, |hD^{\pm}| - 1]) \geq |hD^{\pm}| - 1.$$

The Real Deduction, I

To make this approach work, we consider the set

$$D^{\pm} := (-D) \cup \{0\} \cup D$$

instead of D: it grows faster, while $\mu_A(D^{\pm}) = \mu_A(D)$.

If $\mu_A(D) < |D|$, then (as above) we get

$$|hD^{\pm}| < 2h|D^{\pm}|$$

implying

$$2hD^{\pm} = hD^{\pm} - hD^{\pm} \supseteq [1, |hD^{\pm}| - 1].$$

By monotonicity and the Main Lemma,

$$\mu_A(2hD^{\pm}) \geq \mu_A([1, |hD^{\pm}| - 1]) \geq |hD^{\pm}| - 1.$$

The Real Deduction, II

Comparing

$$\mu_A(2hD^{\pm}) \geq |hD^{\pm}| - 1$$

(from the last slide) to

$$\mu_A(2hD^{\pm}) \leq 2h\mu_A(D^{\pm}) = 2h\mu_A(D) < 2h|D|$$

we get

$$|hD^{\pm}| - 1 < 2h|D| = h(|D^{\pm}| - 1),$$
$$|hD^{\pm}| \leq h|D^{\pm}| - h,$$

which is impossible. □

In fact, this approach works already for $h = 3$.

The Real Deduction, II

Comparing

$$\mu_A(2hD^\pm) \ge |hD^\pm| - 1$$

(from the last slide) to

$$\mu_A(2hD^\pm) \le 2h\mu_A(D^\pm) = 2h\mu_A(D) < 2h|D|$$

we get

$$|hD^\pm| - 1 < 2h|D| = h(|D^\pm| - 1),$$
$$|hD^\pm| \le h|D^\pm| - h,$$

which is impossible. □

In fact, this approach works already for $h = 3$.

The Real Deduction, II

Comparing

$$\mu_A(2hD^{\pm}) \geq |hD^{\pm}| - 1$$

(from the last slide) to

$$\mu_A(2hD^{\pm}) \leq 2h\mu_A(D^{\pm}) = 2h\mu_A(D) < 2h|D|$$

we get

$$|hD^{\pm}| - 1 < 2h|D| = h(|D^{\pm}| - 1),$$
$$|hD^{\pm}| \leq h|D^{\pm}| - h,$$

which is impossible. $\square$

In fact, this approach works already for $h = 3$.

m-Coverable Sets

Remainder of the talk: sketch of the proof of the Main Lemma.

The Main Lemma

There is an absolute constant $C > 0$ such that $\mu_A([1, m]) \geq m$ holds for every finite set $A \subseteq \mathbb{Z}$ with $|A| > Cm \log m$.

The Main Lemma, Restated

There is an absolute constant $C > 0$ such that if the set $A \subseteq \mathbb{Z}$ is m-coverable, then $|A| < Cm \log m$.

A (finite) set $A \subseteq \mathbb{Z}$ is m-coverable if

- $\mu_A([1, m]) < m$; that is, if
- $\Delta_A(d) \leq m - 1$ for every $d \in [1, m]$; in other words, if
- for every $d \in [1, m]$, the set A is a union of at most $m - 1$ arithmetic progressions with difference d.

m-Coverable Sets

Remainder of the talk: sketch of the proof of the Main Lemma.

The Main Lemma

There is an absolute constant $C > 0$ such that $\mu_A([1, m]) \geq m$ holds for every finite set $A \subseteq \mathbb{Z}$ with $|A| > Cm \log m$.

The Main Lemma, Restated

There is an absolute constant $C > 0$ such that if the set $A \subseteq \mathbb{Z}$ is m-coverable, then $|A| < Cm \log m$.

A (finite) set $A \subseteq \mathbb{Z}$ is *m-coverable* if

- $\mu_A([1, m]) < m$; that is, if
- $\Delta_A(d) \leq m - 1$ for every $d \in [1, m]$; in other words, if
- for every $d \in [1, m]$, the set A is a union of at most $m - 1$ arithmetic progressions with difference d.

m-Coverable Sets

Remainder of the talk: sketch of the proof of the Main Lemma.

The Main Lemma

There is an absolute constant $C > 0$ such that $\mu_A([1, m]) \geq m$ holds for every finite set $A \subseteq \mathbb{Z}$ with $|A| > Cm \log m$.

The Main Lemma, Restated

There is an absolute constant $C > 0$ such that if the set $A \subseteq \mathbb{Z}$ is m-coverable, then $|A| < Cm \log m$.

A (finite) set $A \subseteq \mathbb{Z}$ is *m-coverable* if

- $\mu_A([1, m]) < m$; that is, if
- $\Delta_A(d) \leq m - 1$ for every $d \in [1, m]$; in other words, if
- for every $d \in [1, m]$, the set A is a union of at most $m - 1$ arithmetic progressions with difference d.

m-Coverable Sets

Remainder of the talk: sketch of the proof of the Main Lemma.

The Main Lemma

There is an absolute constant $C > 0$ such that $\mu_A([1,m]) \geq m$ holds for every finite set $A \subseteq \mathbb{Z}$ with $|A| > Cm \log m$.

The Main Lemma, Restated

There is an absolute constant $C > 0$ such that if the set $A \subseteq \mathbb{Z}$ is m-coverable, then $|A| < Cm \log m$.

A (finite) set $A \subseteq \mathbb{Z}$ is *m-coverable* if

- $\mu_A([1,m]) < m$; that is, if
- $\Delta_A(d) \leq m - 1$ for every $d \in [1,m]$; in other words, if
- for every $d \in [1,m]$, the set A is a union of at most $m-1$ arithmetic progressions with difference d.

m-Coverable Sets

Remainder of the talk: sketch of the proof of the Main Lemma.

The Main Lemma

There is an absolute constant $C > 0$ such that $\mu_A([1, m]) \geq m$ holds for every finite set $A \subseteq \mathbb{Z}$ with $|A| > Cm \log m$.

The Main Lemma, Restated

There is an absolute constant $C > 0$ such that if the set $A \subseteq \mathbb{Z}$ is m-coverable, then $|A| < Cm \log m$.

A (finite) set $A \subseteq \mathbb{Z}$ is *m-coverable* if

- $\mu_A([1, m]) < m$; that is, if
- $\Delta_A(d) \leq m - 1$ for every $d \in [1, m]$; in other words, if
- for every $d \in [1, m]$, the set A is a union of at most $m - 1$ arithmetic progressions with difference d.

Gaps and Problems

A set $A \subseteq \mathbb{Z}$ is m-coverable if for every $d \in [1, m]$ it is a union of at most $m - 1$ progressions with difference d.

The Main Lemma: if $A \subseteq \mathbb{Z}$ is m-coverable, then $|A| < Cm \log m$.

Notice, that for any $l \in \mathbb{N}$ (and even very large), the interval $A = [1, l]$ is "almost" m-coverable: for each $d \in [1, m - 1]$, it is a union of at most $m - 1$ progressions with difference d. The only trouble is with $d = m$!

Two central notions in the proof of the Main Lemma are gaps and problems.

- A gap in a set S is an element of S which is not in A. We write $\mathfrak{g}_A(S) := |S \setminus A|$; this is the number of gaps in S.
- A problem is a pair $(a, a + d)$ with $a \in A$, $a + d \notin A$, and $d \in [1, m]$. To every $d \in [1, m]$ there correspond at most $m - 1$ problems.

Gaps and Problems

A set $A \subseteq \mathbb{Z}$ is m-coverable if for every $d \in [1, m]$ it is a union of at most $m - 1$ progressions with difference d.

The Main Lemma: if $A \subseteq \mathbb{Z}$ is m-coverable, then $|A| < Cm \log m$.

Notice, that for any $l \in \mathbb{N}$ (and even very large), the interval $A = [1, l]$ is "almost" m-coverable: for each $d \in [1, m - 1]$, it is a union of at most $m - 1$ progressions with difference d. The only trouble is with $d = m$!

Two central notions in the proof of the Main Lemma are gaps and problems.

- A gap in a set S is an element of S which is not in A. We write $\mathfrak{g}_A(S) := |S \setminus A|$; this is the number of gaps in S.
- A problem is a pair $(a, a + d)$ with $a \in A$, $a + d \notin A$, and $d \in [1, m]$. To every $d \in [1, m]$ there correspond at most $m - 1$ problems.

The Three Pillars

Lemma 1

Suppose that A is m-coverable. If $\varepsilon > 0$ and $L \geq m$ have the property that for every $u \in \mathbb{Z}$ there exists $w \in \mathbb{Z}$ with $|w - u| \leq L$ such that $\mathfrak{g}_A([w+1, w+m]) \geq \varepsilon m$, then $|A| < 30\varepsilon^{-1}L$.

Lemma 2

There is an absolute constant $K \geq 2$ with the following property: if A is m-coverable, then for every $u \in \mathbb{Z}$ with $K \leq \mathfrak{g}_A([u+1, u+m]) \leq m/K$ there exists $w \in \mathbb{Z}$ such that $|w - u| \leq Km$ and

$$\mathfrak{g}_A([w+1, w+m]) > 2\mathfrak{g}_A([u+1, u+m]).$$

Lemma 3

If A is m-coverable, then for every $u \in \mathbb{Z}$ and $1 \leq K \leq m/2$ there exists $w \in \mathbb{Z}$ with $|w - u| < Km$ such that $\mathfrak{g}_A([w+1, w+m]) \geq K$.

How it works

Combining Lemmas 1–3, we prove the Main Lemma as follows.

Suppose that A is m-coverable, and let $u \in \mathbb{Z}$.

- Applying Lemma 3, find $w_0 \in \mathbb{Z}$ with $|w_0 - u| < Km$ and $\mathfrak{g}_A([w_0 + 1, w_0 + m]) \geq K$ (where K is a sufficiently large constant).
- Applying Lemma 2 iteratively about $\log_K m$ times, find $w \in \mathbb{Z}$ with $|w - w_0| < Km \ln m$ and $\mathfrak{g}_A([w + 1, w + m]) > m/K$.
- Thus, for every $u \in \mathbb{Z}$ there is $w \in \mathbb{Z}$ with $|w - u| < 2Km \ln m$ and $\mathfrak{g}_A([w + 1, w + m]) > m/K$. That is, the assumptions of Lemma 1 are satisfied with $L = 2Km \ln m$ and $\varepsilon = 1/K$. Hence, if A is m-coverable, then $|A| < 60K^2 m \ln m$, proving the Main Lemma.

How it works

Combining Lemmas 1–3, we prove the Main Lemma as follows.

Suppose that A is m-coverable, and let $u \in \mathbb{Z}$.

- Applying Lemma 3, find $w_0 \in \mathbb{Z}$ with $|w_0 - u| < Km$ and $\mathfrak{g}_A([w_0+1, w_0+m]) \geq K$ (where K is a sufficiently large constant).
- Applying Lemma 2 iteratively about $\log_K m$ times, find $w \in \mathbb{Z}$ with $|w - w_0| < Km \ln m$ and $\mathfrak{g}_A([w+1, w+m]) > m/K$.
- Thus, for every $u \in \mathbb{Z}$ there is $w \in \mathbb{Z}$ with $|w-u| < 2Km \ln m$ and $\mathfrak{g}_A([w+1, w+m]) > m/K$. That is, the assumptions of Lemma 1 are satisfied with $L = 2Km \ln m$ and $\varepsilon = 1/K$. Hence, if A is m-coverable, then $|A| < 60K^2 m \ln m$, proving the Main Lemma.

How it works

Combining Lemmas 1–3, we prove the Main Lemma as follows.

Suppose that A is m-coverable, and let $u \in \mathbb{Z}$.

- Applying Lemma 3, find $w_0 \in \mathbb{Z}$ with $|w_0 - u| < Km$ and $\mathfrak{g}_A([w_0 + 1, w_0 + m]) \geq K$ (where K is a sufficiently large constant).
- Applying Lemma 2 iteratively about $\log_K m$ times, find $w \in \mathbb{Z}$ with $|w - w_0| < Km \ln m$ and $\mathfrak{g}_A([w + 1, w + m]) > m/K$.
- Thus, for every $u \in \mathbb{Z}$ there is $w \in \mathbb{Z}$ with $|w - u| < 2Km \ln m$ and $\mathfrak{g}_A([w + 1, w + m]) > m/K$. That is, the assumptions of Lemma 1 are satisfied with $L = 2Km \ln m$ and $\varepsilon = 1/K$. Hence, if A is m-coverable, then $|A| < 60K^2 m \ln m$, proving the Main Lemma.

How it works

Combining Lemmas 1–3, we prove the Main Lemma as follows.

Suppose that A is m-coverable, and let $u \in \mathbb{Z}$.

- Applying Lemma 3, find $w_0 \in \mathbb{Z}$ with $|w_0 - u| < Km$ and $\mathfrak{g}_A([w_0+1, w_0+m]) \ge K$ (where K is a sufficiently large constant).
- Applying Lemma 2 iteratively about $\log_K m$ times, find $w \in \mathbb{Z}$ with $|w - w_0| < Km \ln m$ and $\mathfrak{g}_A([w+1, w+m]) > m/K$.
- Thus, for every $u \in \mathbb{Z}$ there is $w \in \mathbb{Z}$ with $|w - u| < 2Km \ln m$ and $\mathfrak{g}_A([w+1, w+m]) > m/K$. That is, the assumptions of Lemma 1 are satisfied with $L = 2Km \ln m$ and $\varepsilon = 1/K$. Hence, if A is m-coverable, then $|A| < 60K^2 m \ln m$, proving the Main Lemma.

Why it works?

Suppose that A is m-coverable.

Since A is a union of at most $m-1$ progressions with difference m, some residue class $\pmod{m}$ is not represented in A. Hence every interval of length m contains a gap.

This gap is a terminating point of m progressions with differences $1, 2, \dots, m$. This potentially creates m problems as an element of A, followed by an element not in A at distance $d \in [1, m]$, results in terminating a progression in A with difference d; however the total supply of such progressions is limited (at most $m-1$).

To avoid having too many problems, a typical gap must have many other gaps in its neighborhood. (If $g \notin A$, but $g-d \in A$ for $d \in [1, m]$, we have a problem.) Thus, gaps “breed”!

When “critical mass” of gaps is reached, there is no room for elements of A around: mixing elements of A with gaps creates a *lot* of problems.

Why it works?

Suppose that A is m-coverable.

Since A is a union of at most $m-1$ progressions with difference m, some residue class $\pmod m$ is not represented in A. Hence every interval of length m contains a gap.

This gap is a terminating point of m progressions with differences $1, 2, \dots, m$. This potentially creates m problems as an element of A, followed by an element not in A at distance $d \in [1, m]$, results in terminating a progression in A with difference d; however the total supply of such progressions is limited (at most $m-1$).

To avoid having too many problems, a typical gap must have many other gaps in its neighborhood. (If $g \notin A$, but $g-d \in A$ for $d \in [1, m]$, we have a problem.) Thus, gaps "breed"!

When "critical mass" of gaps is reached, there is no room for elements of A around: mixing elements of A with gaps creates a *lot* of problems.

Why it works?

Suppose that A is m-coverable.

Since A is a union of at most $m-1$ progressions with difference m, some residue class $\pmod{m}$ is not represented in A. Hence every interval of length m contains a gap.

This gap is a terminating point of m progressions with differences $1, 2, \ldots, m$. This potentially creates m problems as an element of A, followed by an element not in A at distance $d \in [1, m]$, results in terminating a progression in A with difference d; however the total supply of such progressions is limited (at most $m-1$).

To avoid having too many problems, a typical gap must have many other gaps in its neighborhood. (If $g \notin A$, but $g-d \in A$ for $d \in [1, m]$, we have a problem.) Thus, gaps "breed"!

When "critical mass" of gaps is reached, there is no room for elements of A around: mixing elements of A with gaps creates a *lot* of problems.

Why it works?

Suppose that A is m-coverable.

Since A is a union of at most $m-1$ progressions with difference m, some residue class $\pmod{m}$ is not represented in A. Hence every interval of length m contains a gap.

This gap is a terminating point of m progressions with differences $1, 2, \ldots, m$. This potentially creates m problems as an element of A, followed by an element not in A at distance $d \in [1, m]$, results in terminating a progression in A with difference d; however the total supply of such progressions is limited (at most $m-1$).

To avoid having too many problems, a typical gap must have many other gaps in its neighborhood. (If $g \notin A$, but $g-d \in A$ for $d \in [1, m]$, we have a problem.) Thus, gaps “breed”!

When “critical mass” of gaps is reached, there is no room for elements of A around: mixing elements of A with gaps creates a *lot* of problems.

Why it works?

Suppose that A is m-coverable.

Since A is a union of at most $m-1$ progressions with difference m, some residue class $\pmod{m}$ is not represented in A. Hence every interval of length m contains a gap.

This gap is a terminating point of m progressions with differences $1, 2, \dots, m$. This potentially creates m problems as an element of A, followed by an element not in A at distance $d \in [1, m]$, results in terminating a progression in A with difference d; however the total supply of such progressions is limited (at most $m-1$).

To avoid having too many problems, a typical gap must have many other gaps in its neighborhood. (If $g \notin A$, but $g-d \in A$ for $d \in [1, m]$, we have a problem.) Thus, gaps "breed"!

When "critical mass" of gaps is reached, there is no room for elements of A around: mixing elements of A with gaps creates a *lot* of problems.

Open Problems

Problem 1: $\mathbb{Z}/p\mathbb{Z}$

How about abelian groups, other than $\mathbb{Z}$? Is it true that for any $A, D \subseteq \mathbb{Z}/p\mathbb{Z}$ with $|D| < c|A|/\ln|A|$ there exists $d \in D$ with $|(A+d) \setminus A| \geq (|D|-1)/2$?

Problem 2: Popular Sums

How about popular *sums*? Is it true that for any finite sets $A, D \subseteq \mathbb{Z}$ with $|D| < c|A|/\ln|A|$ there exists $d \in D$ with $|(d-A) \setminus A| \geq (|D|-1)/2$?

Problem 3: Relaxing the Assumptions

Is it true that for any finite $A \subseteq \mathbb{Z}$ and $D \subseteq \mathbb{N}$ with $|D| < c|A|$ there exists $d \in D$ with $|(A+d) \setminus A| \geq |D| - O(1)$? That is, does $|D| < c|A|$ imply $\mu_A(D) \geq |D| - O(1)$?

Open Problems

Problem 1: $\mathbb{Z}/p\mathbb{Z}$

How about abelian groups, other than $\mathbb{Z}$? Is it true that for any $A, D \subseteq \mathbb{Z}/p\mathbb{Z}$ with $|D| < c|A|/\ln|A|$ there exists $d \in D$ with $|(A+d) \setminus A| \geq (|D|-1)/2$?

Problem 2: Popular Sums

How about popular *sums*? Is it true that for any finite sets $A, D \subseteq \mathbb{Z}$ with $|D| < c|A|/\ln|A|$ there exists $d \in D$ with $|(d-A) \setminus A| \geq (|D|-1)/2$?

Problem 3: Relaxing the Assumptions

Is it true that for any finite $A \subseteq \mathbb{Z}$ and $D \subseteq \mathbb{N}$ with $|D| < c|A|$ there exists $d \in D$ with $|(A+d) \setminus A| \geq |D| - O(1)$? That is, does $|D| < c|A|$ imply $\mu_A(D) \geq |D| - O(1)$?