Subset Sums

Jacob Fox

Stanford University

Harvard Mathematical Picture Seminar

February 21, 2024

Joint work with David Conlon and Huy Tuan Pham

Goldbach's conjecture

Every even integer at least 4 is the sum of two primes.

Gauss' Eureka theorem

Every positive integer is the sum of three triangular numbers.

Lagrange's four square theorem

Every positive integer is the sum of four perfect squares.

Complete sequences

Let A be a sequence of positive integers. Let $\Sigma(A)$ be the set of integers representable as a sum of distinct terms of A.

Complete sequences

Let A be a sequence of positive integers. Let $\Sigma(A)$ be the set of integers representable as a sum of distinct terms of A.

A is *complete* if every sufficiently large integer is in $\Sigma(A)$, Let A be a sequence of positive integers. Let $\Sigma(A)$ be the set of integers representable as a sum of distinct terms of A.

A is *complete* if every sufficiently large integer is in $\Sigma(A)$, and *entirely complete* if every positive integer is in $\Sigma(A)$.

Examples:

- $\{2^i : i \geq 0\}$ is entirely complete.

Complete sequences

Let A be a sequence of positive integers. Let $\Sigma(A)$ be the set of integers representable as a sum of distinct terms of A.

A is *complete* if every sufficiently large integer is in $\Sigma(A)$, and *entirely complete* if every positive integer is in $\Sigma(A)$.

Examples:

- $\{2^i : i \geq 0\}$ is entirely complete.
- $\{2^i : i \geq 1\}$ is not complete.

Complete sequences

Let A be a sequence of positive integers. Let $\Sigma(A)$ be the set of integers representable as a sum of distinct terms of A.

A is *complete* if every sufficiently large integer is in $\Sigma(A)$, and *entirely complete* if every positive integer is in $\Sigma(A)$.

Examples:

- $\{2^i : i \geq 0\}$ is entirely complete.
- $\{2^i : i \geq 1\}$ is not complete.
- For all $k \in \mathbb{N}$, $\{i^k : i \geq 1\}$ is complete (Sprague 1947).

Complete sequences

Let A be a sequence of positive integers. Let $\Sigma(A)$ be the set of integers representable as a sum of distinct terms of A.

A is *complete* if every sufficiently large integer is in $\Sigma(A)$, and *entirely complete* if every positive integer is in $\Sigma(A)$.

Examples:

- $\{2^i : i \geq 0\}$ is entirely complete.
- $\{2^i : i \geq 1\}$ is not complete.
- For all $k \in \mathbb{N}$, $\{i^k : i \geq 1\}$ is complete (Sprague 1947).
- $p, q \geq 2$ coprime $\Rightarrow \{p^i q^j : i, j \geq 0\}$ is complete (Birch 1959).

Complete sequences

Let A be a sequence of positive integers. Let $\Sigma(A)$ be the set of integers representable as a sum of distinct terms of A.

A is *complete* if every sufficiently large integer is in $\Sigma(A)$, and *entirely complete* if every positive integer is in $\Sigma(A)$.

Examples:

- $\{2^i : i \geq 0\}$ is entirely complete.
- $\{2^i : i \geq 1\}$ is not complete.
- For all $k \in \mathbb{N}$, $\{i^k : i \geq 1\}$ is complete (Sprague 1947).
- $p, q \geq 2$ coprime $\Rightarrow \{p^i q^j : i, j \geq 0\}$ is complete (Birch 1959).
- The set of even numbers is not complete.

Entirely complete sequences

Entirely complete sequences

Proposition: (Graham)

$A = \{a_1 \leq a_2 \leq \ldots\}$ is entirely complete iff
$a_1 = 1$ and $a_k - 1 \leq \sum_{j<k} a_k$ for all $k > 1$.

Entirely complete sequences

Proposition: (Graham)

$A = \{a_1 \le a_2 \le \dots\}$ is entirely complete iff
$a_1 = 1$ and $a_k - 1 \le \sum_{j<k} a_k$ for all $k > 1$.

Proof: $\Rightarrow$ If $a_k - 1 > \sum_{j<k} a_j$, then $a_k - 1$ is not in $\Sigma(A)$.

Entirely complete sequences

Proposition: (Graham)

$A = \{a_1 \leq a_2 \leq \ldots\}$ is entirely complete iff
$a_1 = 1$ and $a_k - 1 \leq \sum_{j<k} a_k$ for all $k > 1$.

Proof: $\Rightarrow$ If $a_k - 1 > \sum_{j<k} a_j$, then $a_k - 1$ is not in $\Sigma(A)$.

$\Leftarrow$ By induction on k, we get $\Sigma(\{a_j\}_{j=1}^{k}) = [\sum_{j=1}^{k} a_j]$.

Entirely complete sequences

Proposition: (Graham)

$A = \{a_1 \leq a_2 \leq \ldots\}$ is entirely complete iff
$a_1 = 1$ and $a_k - 1 \leq \sum_{j<k} a_k$ for all $k > 1$.

Proof: $\Rightarrow$ If $a_k - 1 > \sum_{j<k} a_j$, then $a_k - 1$ is not in $\Sigma(A)$.

$\Leftarrow$ By induction on k, we get $\Sigma(\{a_j\}_{j=1}^{k}) = [\sum_{j=1}^{k} a_j]$.

Lemma: (Graham)

Suppose $\Sigma(A)$ contains all integers in the interval $[x, x+y)$.

1. If a is a positive integer with $a \leq y$ and $a \notin A$, then $\Sigma(A \cup \{a\})$ contains all integers in the interval $[x, x+y+a)$.
2. If $a_1, \ldots, a_s$ are positive integers such that $a_i \leq y + \sum_{j<i} a_j$ and $a_i \notin A$ for $i = 1, \ldots, s$, then $\Sigma(A \cup \{a_1, a_2, \ldots, a_s\})$ contains all integers in the interval $[x, x+y+\sum_{i=1}^{s} a_i)$.

Complete polynomial sequences

Complete polynomial sequences

Let P be a polynomial in one variable and $A = \{P(n)\}_{n\geq 1}$.

Complete polynomial sequences

Let P be a polynomial in one variable and $A = \{P(n)\}_{n \geq 1}$.

If A is complete, then

- the leading coefficient of P is positive, and

Complete polynomial sequences

Let P be a polynomial in one variable and $A = \{P(n)\}_{n \geq 1}$.

If A is complete, then

- the leading coefficient of P is positive, and
- for every prime p there is n such that p does not divide $P(n)$.

Complete polynomial sequences

Let P be a polynomial in one variable and $A = \{P(n)\}_{n \geq 1}$.

If A is complete, then

- the leading coefficient of P is positive, and
- for every prime p there is n such that p does not divide $P(n)$.

Necessary conditions are sufficient (Roth and Szekeres 1959)

Complete polynomial sequences

Let P be a polynomial in one variable and $A = \{P(n)\}_{n \geq 1}$.

If A is complete, then

- the leading coefficient of P is positive, and
- for every prime p there is n such that p does not divide $P(n)$.

Necessary conditions are sufficient (Roth and Szekeres 1959)

Another proof can be deduced from (Cassels 1962)

Let P be a polynomial in one variable and $A = \{P(n)\}_{n \geq 1}$.

If A is complete, then

- the leading coefficient of P is positive, and
- for every prime p there is n such that p does not divide $P(n)$.

Necessary conditions are sufficient (Roth and Szekeres 1959)

Another proof can be deduced from (Cassels 1962)

Another characterization by (Graham 1964):

Complete polynomial sequences

Let P be a polynomial in one variable and $A = \{P(n)\}_{n \geq 1}$.

If A is complete, then

- the leading coefficient of P is positive, and
- for every prime p there is n such that p does not divide $P(n)$.

Necessary conditions are sufficient (Roth and Szekeres 1959)

Another proof can be deduced from (Cassels 1962)

Another characterization by (Graham 1964):

For $P(x) = \alpha_k \binom{x}{k} + \alpha_{k-1} \binom{x}{k-1} + \cdots + \alpha_0 \binom{x}{0} \in \mathbb{R}[x]$,

A is complete iff

$\alpha_k > 0$ and $\alpha_i = p_i/q_i$ rational $\forall i$ with $\gcd(p_0, p_1, \ldots, p_k) = 1$.

Ramsey complete sequences (Burr and Erdős 1985)

Let $A(n) = |A \cap [n]|$.

Ramsey complete sequences (Burr and Erdős 1985)

Let $A(n) = |A \cap [n]|$.

A is *r-Ramsey complete* if for every partition $A = A_1 \cup \ldots \cup A_r$, every sufficiently large integer is in $\bigcup_{i=1}^{r} \Sigma(A_i)$.

Ramsey complete sequences (Burr and Erdős 1985)

Let $A(n) = |A \cap [n]|$.

A is *r-Ramsey complete* if for every partition $A = A_1 \cup \ldots \cup A_r$, every sufficiently large integer is in $\bigcup_{i=1}^{r} \Sigma(A_i)$.

$\exists$ a 2-Ramsey complete set A with $A(n) = O((\log n)^3)$.

Ramsey complete sequences (Burr and Erdős 1985)

Let $A(n) = |A \cap [n]|$.

A is *r-Ramsey complete* if for every partition $A = A_1 \cup \ldots \cup A_r$, every sufficiently large integer is in $\bigcup_{i=1}^{r} \Sigma(A_i)$.

$\exists$ a 2-Ramsey complete set A with $A(n) = O((\log n)^3)$.

If A is 2-Ramsey complete, then $A(n) = \Omega((\log n)^2)$.

Problems

1. Improve these bounds.

Ramsey complete sequences (Burr and Erdős 1985)

Let $A(n) = |A \cap [n]|$.

A is *r-Ramsey complete* if for every partition $A = A_1 \cup \ldots \cup A_r$, every sufficiently large integer is in $\bigcup_{i=1}^{r} \Sigma(A_i)$.

$\exists$ a 2-Ramsey complete set A with $A(n) = O((\log n)^3)$.

If A is 2-Ramsey complete, then $A(n) = \Omega((\log n)^2)$.

Problems

1. Improve these bounds.
2. Prove there is a sparse r-Ramsey complete sequence for $r > 2$.

Ramsey complete sequences (Burr and Erdős 1985)

Let $A(n) = |A \cap [n]|$.

A is *r-Ramsey complete* if for every partition $A = A_1 \cup \ldots \cup A_r$, every sufficiently large integer is in $\bigcup_{i=1}^r \Sigma(A_i)$.

$\exists$ a 2-Ramsey complete set A with $A(n) = O((\log n)^3)$.

If A is 2-Ramsey complete, then $A(n) = \Omega((\log n)^2)$.

Problems

1. Improve these bounds.
2. Prove there is a sparse r-Ramsey complete sequence for $r > 2$.
3. Determine the r-Ramsey complete polynomial sequences.

Ramsey complete sequences (Burr and Erdős 1985)

A is *r-Ramsey complete* if for every partition $A = A_1 \cup \ldots \cup A_r$, every sufficiently large integer is in $\bigcup_{i=1}^{r} \Sigma(A_i)$.

$\exists$ a 2-Ramsey complete set A with $A(n) = O((\log n)^3)$.

If A is 2-Ramsey complete, then $A(n) = \Omega((\log n)^2)$.

Problems

1. Improve these bounds. (Erdős \$100)
2. Prove there is a sparse r-Ramsey complete set for $r > 2$. (Erdős \$250)
3. Determine the r-Ramsey complete polynomial sequences.

Ramsey complete sequences (Burr and Erdős 1985)

A is r-*Ramsey complete* if for every partition $A = A_1 \cup \ldots \cup A_r$, every sufficiently large integer is in $\bigcup_{i=1}^{r} \Sigma(A_i)$.

$\exists$ a 2-Ramsey complete set A with $A(n) = O((\log n)^3)$.

If A is 2-Ramsey complete, then $A(n) = \Omega((\log n)^2)$.

Problems

1. Improve these bounds. (Erdős $100)
2. Prove there is a sparse r-Ramsey complete set for $r > 2$. (Erdős $250)
3. Determine the r-Ramsey complete polynomial sequences.

We prove a result which solves all of these problems.

Ramsey complete sequences

Theorem 1

Let $r \geq 2$. There exists an r-Ramsey complete sequence A with $A(n) \leq Cr(\log n)^2$ for all n.

If A is r-Ramsey complete, then $A(n) \geq cr(\log n)^2$ for all large n.

Ramsey complete sequences

Theorem 1

Let $r \geq 2$. There exists an r-Ramsey complete sequence A with $A(n) \leq Cr(\log n)^2$ for all n.
If A is r-Ramsey complete, then $A(n) \geq cr(\log n)^2$ for all large n.

Theorem 2

If degree d polynomial P satisfies $\{P(n)\}_{n\geq 1}$ is complete, then there is $A \subset \{P(n)\}_{n\geq 1}$ with $A(n) \leq C_d r(\log n)^2$ for all n such that A is r-Ramsey complete.

Ramsey complete sequences

Theorem 1

Let $r \geq 2$. There exists an r-Ramsey complete sequence A with $A(n) \leq Cr(\log n)^2$ for all n.
If A is r-Ramsey complete, then $A(n) \geq cr(\log n)^2$ for all large n.

Theorem 2

If degree d polynomial P satisfies $\{P(n)\}_{n\geq 1}$ is complete, then there is $A \subset \{P(n)\}_{n\geq 1}$ with $A(n) \leq C_d r(\log n)^2$ for all n such that A is r-Ramsey complete.

Corollary

If $A = \{P(n)\}_{n\geq 1}$ is complete, then A is r-Ramsey complete $\forall r$.

Ramsey complete sequences

Theorem 1

For $r \geq 2$, there is an r-Ramsey complete sequence A with $A(n) \leq Cr(\log n)^2$ for all n.

Ramsey complete sequences

Theorem 1

For $r \geq 2$, there is an r-Ramsey complete sequence A with $A(n) \leq Cr(\log n)^2$ for all n.

Key Lemma

For $0 < \varepsilon \leq 1/2$ and n sufficiently large, there is $S_n \subset [n, 2n)$ with $|S_n| \leq 4000\varepsilon^{-1} \log n$ such that for every $A' \subset S_n$ with $|A'| \geq \varepsilon |S_n|$, we have $[y_n, 3y_n] \subset \Sigma(A')$ with $y_n = 1000n \log n$.

Ramsey complete sequences

Theorem 1

For $r \geq 2$, there is an r-Ramsey complete sequence A with $A(n) \leq Cr(\log n)^2$ for all n.

Key Lemma

For $0 < \varepsilon \leq 1/2$ and n sufficiently large, there is $S_n \subset [n, 2n)$ with $|S_n| \leq 4000\varepsilon^{-1} \log n$ such that for every $A' \subset S_n$ with $|A'| \geq \varepsilon |S_n|$, we have $[y_n, 3y_n] \subset \Sigma(A')$ with $y_n = 1000n \log n$.

Proof of Theorem 1: Apply Key Lemma with $\varepsilon = 1/r$ and $n = 2^i$ for each $i \geq i_0$, and let $A = \bigcup_{i \geq i_0} S_{2^i}$.

Theorem 1

For $r \geq 2$, there is an r-Ramsey complete sequence A with $A(n) \leq Cr(\log n)^2$ for all n.

Key Lemma

For $0 < \varepsilon \leq 1/2$ and n sufficiently large, there is $S_n \subset [n, 2n)$ with $|S_n| \leq 4000\varepsilon^{-1} \log n$ such that for every $A' \subset S_n$ with $|A'| \geq \varepsilon |S_n|$, we have $[y_n, 3y_n] \subset \Sigma(A')$ with $y_n = 1000 n \log n$.

Proof of Theorem 1: Apply Key Lemma with $\varepsilon = 1/r$ and $n = 2^i$ for each $i \geq i_0$, and let $A = \bigcup_{i \geq i_0} S_{2^i}$. Consider an r-coloring of A.

Theorem 1

For $r \geq 2$, there is an r-Ramsey complete sequence A with $A(n) \leq Cr(\log n)^2$ for all n.

Key Lemma

For $0 < \varepsilon \leq 1/2$ and n sufficiently large, there is $S_n \subset [n, 2n)$ with $|S_n| \leq 4000\varepsilon^{-1} \log n$ such that for every $A' \subset S_n$ with $|A'| \geq \varepsilon |S_n|$, we have $[y_n, 3y_n] \subset \Sigma(A')$ with $y_n = 1000 n \log n$.

Proof of Theorem 1: Apply Key Lemma with $\varepsilon = 1/r$ and $n = 2^i$ for each $i \geq i_0$, and let $A = \bigcup_{i \geq i_0} S_{2^i}$. Consider an r-coloring of A. Let $A_i \subset S_{2^i}$ consist of the elements of the most common color.

Theorem 1

For $r \geq 2$, there is an r-Ramsey complete sequence A with $A(n) \leq Cr(\log n)^2$ for all n.

Key Lemma

For $0 < \varepsilon \leq 1/2$ and n sufficiently large, there is $S_n \subset [n, 2n)$ with $|S_n| \leq 4000\varepsilon^{-1} \log n$ such that for every $A' \subset S_n$ with $|A'| \geq \varepsilon |S_n|$, we have $[y_n, 3y_n] \subset \Sigma(A')$ with $y_n = 1000n \log n$.

Proof of Theorem 1: Apply Key Lemma with $\varepsilon = 1/r$ and $n = 2^i$ for each $i \geq i_0$, and let $A = \bigcup_{i \geq i_0} S_{2^i}$. Consider an r-coloring of A. Let $A_i \subset S_{2^i}$ consist of the elements of the most common color. Intervals $I_i = [y_{2^i}, 3y_{2^i}]$ cover all large integers and $I_i \subset \Sigma(A_i)$.

Key Lemma

For $0 < \varepsilon \leq 1/2$ and n sufficiently large, there is $S_n \subset [n, 2n)$ with $|S_n| \leq 4000\varepsilon^{-1} \log n$ such that for every $A' \subset S_n$ with $|A'| \geq \varepsilon |S_n|$, we have $[y_n, 3y_n] \subset \Sigma(A')$ with $y_n = 1000 n \log n$.

Key Lemma

For $0 < \varepsilon \leq 1/2$ and n sufficiently large, there is $S_n \subset [n, 2n)$ with $|S_n| \leq 4000\varepsilon^{-1} \log n$ such that for every $A' \subset S_n$ with $|A'| \geq \varepsilon|S_n|$, we have $[y_n, 3y_n] \subset \Sigma(A')$ with $y_n = 1000n \log n$.

How do we build the set S_n?

Key Lemma

For $0 < \varepsilon \leq 1/2$ and n sufficiently large, there is $S_n \subset [n, 2n)$ with $|S_n| \leq 4000\varepsilon^{-1} \log n$ such that for every $A' \subset S_n$ with $|A'| \geq \varepsilon |S_n|$, we have $[y_n, 3y_n] \subset \Sigma(A')$ with $y_n = 1000 n \log n$.

How do we build the set S_n?

At random!

Key Lemma

For $0 < \varepsilon \leq 1/2$ and n sufficiently large, there is $S_n \subset [n, 2n)$ with $|S_n| \leq 4000\varepsilon^{-1} \log n$ such that for every $A' \subset S_n$ with $|A'| \geq \varepsilon |S_n|$, we have $[y_n, 3y_n] \subset \Sigma(A')$ with $y_n = 1000 n \log n$.

How do we build the set S_n?

At random! (but not in the most obvious way)

Key Lemma

For $0 < \varepsilon \leq 1/2$ and n sufficiently large, there is $S_n \subset [n, 2n)$ with $|S_n| \leq 4000\varepsilon^{-1} \log n$ such that for every $A' \subset S_n$ with $|A'| \geq \varepsilon |S_n|$, we have $[y_n, 3y_n] \subset \Sigma(A')$ with $y_n = 1000 n \log n$.

How do we build the set S_n?

At random! (but not in the most obvious way)

Pick $S_n \subset [n, 2n)$ uniformly at random of size $4000\varepsilon^{-1} \log n$ so that no element of S_n has a prime factor less than $(\log n)/2$.

Key Lemma

For $0 < \varepsilon \leq 1/2$ and n sufficiently large, there is $S_n \subset [n, 2n)$ with $|S_n| \leq 4000\varepsilon^{-1} \log n$ such that for every $A' \subset S_n$ with $|A'| \geq \varepsilon |S_n|$, we have $[y_n, 3y_n] \subset \Sigma(A')$ with $y_n = 1000 n \log n$.

How do we build the set S_n?

At random! (but not in the most obvious way)

Pick $S_n \subset [n, 2n)$ uniformly at random of size $4000\varepsilon^{-1} \log n$ so that no element of S_n has a prime factor less than $(\log n)/2$.

With very high probability, a random $A' \subset [n, 2n)$ with $|A'| = 4000 \log n$ and each element has no prime factor less than $(\log n)/2$ satisfies $[y_n, 3y_n] \subset \Sigma(A')$.

Key Lemma

For $0 < \varepsilon \leq 1/2$ and n sufficiently large, there is $S_n \subset [n, 2n)$ with $|S_n| \leq 4000\varepsilon^{-1} \log n$ such that for every $A' \subset S_n$ with $|A'| \geq \varepsilon |S_n|$, we have $[y_n, 3y_n] \subset \Sigma(A')$ with $y_n = 1000 n \log n$.

How do we build the set S_n?

At random! (but not in the most obvious way)

Pick $S_n \subset [n, 2n)$ uniformly at random of size $4000\varepsilon^{-1} \log n$ so that no element of S_n has a prime factor less than $(\log n)/2$.

With very high probability, a random $A' \subset [n, 2n)$ with $|A'| = 4000 \log n$ and each element has no prime factor less than $(\log n)/2$ satisfies $[y_n, 3y_n] \subset \Sigma(A')$.

We then union bound over all $A' \subset S_n$.

A recipe for finding intervals in subset sums

Let A be a set of integers.

1. Partition A into ℓ sets $A_1, \ldots, A_\ell$.
2. Main step: Partition $A_i = B_i \cup C_i$ so that the set of subset sums of B_i is large modulo each $c \in C_i$.
3. Using the previous step, obtain $\Sigma(A_i) = \Sigma(B_i \cup C_i)$ is large.
4. Using that each $\Sigma(A_i)$ is large, we get their sumset and hence $\Sigma(A)$ contains a long interval.

A recipe for finding intervals in subset sums

Let A be a set of integers.

1. Partition A into ℓ sets $A_1, \ldots, A_\ell$.
2. Main step: Partition $A_i = B_i \cup C_i$ so that the set of subset sums of B_i is large modulo each $c \in C_i$.
3. Using the previous step, obtain $\Sigma(A_i) = \Sigma(B_i \cup C_i)$ is large.
4. Using that each $\Sigma(A_i)$ is large, we get their sumset and hence $\Sigma(A)$ contains a long interval.

Claim, helpful for step 3

Let $c \in \mathbb{N}$, $B \subset \mathbb{Z}$ with $c \notin B$ and the size of $\Sigma(A)$ considered modulo c is at least h, then $|\Sigma(A \cup \{c\})| \geq |\Sigma(A)| + h$.

A recipe for finding intervals in subset sums

Let A be a set of integers.

1. Partition A into ℓ sets $A_1, \ldots, A_\ell$.
2. Main step: Partition $A_i = B_i \cup C_i$ so that the set of subset sums of B_i is large modulo each $c \in C_i$.
3. Using the previous step, obtain $\Sigma(A_i) = \Sigma(B_i \cup C_i)$ is large.
4. Using that each $\Sigma(A_i)$ is large, we get their sumset and hence $\Sigma(A)$ contains a long interval.

Lemma (Lev), helpful for step 4

Let $\ell, q \geq 1$ and $n \geq 3$ are integers with $\ell \geq 2\lceil (q-1)/(n-2) \rceil$. If $A_1, \ldots, A_\ell \subset \mathbb{Z}$ with each $|A_i| \geq n$, each A_i a subset of an interval of at most $q+1$ integers and none of which is a subset of an arithmetic progression of common difference greater than one, then $A_1 + \cdots + A_\ell$ contains an interval of length at least $\ell(n-1)+1$.

Density complete sequences

A set A is ε-*complete* if every $A' \subset A$ with $A'(n) \geq \varepsilon A(n)$ for n sufficiently large is complete.

Question

How sparse can an ε-complete sequence be?

Density complete sequences

A set A is ε-*complete* if every $A' \subset A$ with $A'(n) \geq \varepsilon A(n)$ for n sufficiently large is complete.

Question

How sparse can an ε-complete sequence be?

An ε-complete A must satisfy modularity and growth conditions:

1. For each prime p, the multiples of p in A have density $\leq \varepsilon$.

Density complete sequences

A set A is ε-*complete* if every $A' \subset A$ with $A'(n) \geq \varepsilon A(n)$ for n sufficiently large is complete.

Question

How sparse can an ε-complete sequence be?

An ε-complete A must satisfy modularity and growth conditions:
1. For each prime p, the multiples of p in A have density $\leq \varepsilon$.
2. $\exists$ C such that $a_k \leq \sum_{i<\varepsilon k+C} a_i$ for all k.

Roughly, a random sequence satisfying the modularity and growth conditions is almost surely ε-complete. In particular, we have:

Theorem

Let $f_1, \ldots, f_t \in \mathbb{N}$ for $t \geq 1/\varepsilon$ and $f_m = \sum_{i\leq\varepsilon m} f_i$ for $m > t$.

Density complete sequences

A set A is ε-*complete* if every $A' \subset A$ with $A'(n) \geq \varepsilon A(n)$ for n sufficiently large is complete.

Question

How sparse can an ε-complete sequence be?

An ε-complete A must satisfy modularity and growth conditions:

1. For each prime p, the multiples of p in A have density $\leq \varepsilon$.
2. $\exists$ C such that $a_k \leq \sum_{i<\varepsilon k+C} a_i$ for all k.

Roughly, a random sequence satisfying the modularity and growth conditions is almost surely ε-complete. In particular, we have:

Theorem

Let $f_1, \ldots, f_t \in \mathbb{N}$ for $t \geq 1/\varepsilon$ and $f_m = \sum_{i \leq \varepsilon m} f_i$ for $m > t$. If A is ε-complete, then $a_k = O(f_k)$.

Density complete sequences

A set A is ε-*complete* if every $A' \subset A$ with $A'(n) \geq \varepsilon A(n)$ for n sufficiently large is complete.

Question

How sparse can an ε-complete sequence be?

An ε-complete A must satisfy modularity and growth conditions:

1. For each prime p, the multiples of p in A have density $\leq \varepsilon$.
2. $\exists$ C such that $a_k \leq \sum_{i<\varepsilon k+C} a_i$ for all k.

Roughly, a random sequence satisfying the modularity and growth conditions is almost surely ε-complete. In particular, we have:

Theorem

Let $f_1, \ldots, f_t \in \mathbb{N}$ for $t \geq 1/\varepsilon$ and $f_m = \sum_{i \leq \varepsilon m} f_i$ for $m > t$.
If A is ε-complete, then $a_k = O(f_k)$.
There exists an ε-complete sequence A with $a_k = \Theta(f_k)$.

Conjecture

Conjecture

Let $1 < p_1 < \ldots < p_{r+1}$ be pairwise relatively prime.
The sequence $\{p_1^{i_1} p_2^{i_2} \cdots p_{r+1}^{i_{r+1}}\}_{i_1,\ldots,i_{r+1} \geq 0}$ is Ramsey r-complete.

Conjecture

Conjecture

Let $1 < p_1 < \ldots < p_{r+1}$ be pairwise relatively prime.
The sequence $\{p_1^{i_1} p_2^{i_2} \cdots p_{r+1}^{i_{r+1}}\}_{i_1,\ldots,i_{r+1} \geq 0}$ is Ramsey r-complete.

Remark: The sequence is not $(r+1)$-Ramsey complete:

Conjecture

Conjecture

Let $1 < p_1 < \ldots < p_{r+1}$ be pairwise relatively prime.
The sequence $\{p_1^{i_1} p_2^{i_2} \cdots p_{r+1}^{i_{r+1}}\}_{i_1,\ldots,i_{r+1} \geq 0}$ is Ramsey r-complete.

Remark: The sequence is not $(r+1)$-Ramsey complete:
Assign $p_1^{i_1} p_2^{i_2} \cdots p_{r+1}^{i_{r+1}}$ a color j for which i_j is nonzero and $j \leq r$, and color $r+1$ otherwise.

Sure monochromatic subset sums

Definition

Let $f(n)$ be the minimum r such that there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Sure monochromatic subset sums

Definition

Let $f(n)$ be the minimum r such that there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Example: $f(23) \leq 3$ given by the coloring

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22

Sure monochromatic subset sums

Definition

Let $f(n)$ be the minimum r such that there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Example: $f(23) \leq 3$ given by the coloring

$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22$

More generally, the greedy coloring uses $\sim \sqrt{n/2}$ colors.

Sure monochromatic subset sums

Definition

Let $f(n)$ be the minimum r such that there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Example: $f(23) \leq 3$ given by the coloring

$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22$

More generally, the greedy coloring uses $\sim \sqrt{n/2}$ colors.

However, this is not close to best possible.

Sure monochromatic subset sums

Definition

Let $f(n)$ be the minimum r such that there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Example: $f(23) \leq 3$ given by the coloring

$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22$

More generally, the greedy coloring uses $\sim \sqrt{n/2}$ colors.

However, this is not close to best possible.

Erdős first proved $f(n) = o(n^{1/3})$ and asked if $f(n) = n^{1/3 - o(1)}$.

Sure monochromatic subset sums

Definition

Let $f(n)$ be the minimum r such that there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Example: $f(23) \leq 3$ given by the coloring

$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22$

More generally, the greedy coloring uses $\sim \sqrt{n/2}$ colors.

However, this is not close to best possible.

Erdős first proved $f(n) = o(n^{1/3})$ and asked if $f(n) = n^{1/3 - o(1)}$.

Sure monochromatic subset sums

Definition

Let $f(n)$ be the minimum r such that there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Theorem: (Alon and Erdős 1996)

There are positive constants c, C such that

$$\frac{cn^{1/3}}{\log^{4/3} n} \leq f(n) \leq \frac{Cn^{1/3}(\log\log n)^{1/3}}{(\log n)^{1/3}}.$$

Sure monochromatic subset sums

Definition

Let $f(n)$ be the minimum r such that there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Theorem: (Alon and Erdős 1996)

There are positive constants c, C such that

$$\frac{cn^{1/3}}{\log^{4/3} n} \leq f(n) \leq \frac{Cn^{1/3}(\log\log n)^{1/3}}{(\log n)^{1/3}}.$$

They conjectured that $f(n)$ grows more like the upper bound.

Sure monochromatic subset sums

Definition

Let $f(n)$ be the minimum r such that there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Theorem: (Alon and Erdős 1996)

There are positive constants c, C such that

$$\frac{cn^{1/3}}{\log^{4/3} n} \leq f(n) \leq \frac{Cn^{1/3}(\log\log n)^{1/3}}{(\log n)^{1/3}}.$$

They conjectured that $f(n)$ grows more like the upper bound.
Vu improved the lower bound to $f(n) \geq cn^{1/3}/\log n$.

Sure monochromatic subset sums

Definition

Let $f(n)$ be the minimum r such that there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Theorem: (Alon and Erdős 1996)

There are positive constants c, C such that

$$\frac{cn^{1/3}}{\log^{4/3} n} \leq f(n) \leq \frac{Cn^{1/3}(\log \log n)^{1/3}}{(\log n)^{1/3}}.$$

They conjectured that $f(n)$ grows more like the upper bound.
Vu improved the lower bound to $f(n) \geq cn^{1/3}/\log n$.

Theorem: (Conlon-F.-Pham)

$$f(n) = \Theta\left(\frac{n^{1/3}(n/\phi(n))}{(\log n)^{1/3}(\log \log n)^{2/3}}\right)$$

The Alon-Erdős coloring

Theorem: (Alon and Erdős 1996)

For $r := Cn^{1/3}(\log n)^{-1/3}(\log\log n)^{1/3}$, there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

The Alon-Erdős coloring

Theorem: (Alon and Erdős 1996)

For $r := Cn^{1/3}(\log n)^{-1/3}(\log\log n)^{1/3}$, there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Type 1: For $j \in [r/2]$, color integers in $[n/(j+1), n/j)$ color j.

The Alon-Erdős coloring

Theorem: (Alon and Erdős 1996)

For $r := Cn^{1/3}(\log n)^{-1/3}(\log\log n)^{1/3}$, there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Type 1: For $j \in [r/2]$, color integers in $[n/(j+1), n/j)$ color j.

Type 2: For each of the first $r/4$ primes p that do not divide n, color the multiples of p using one color.

The Alon-Erdős coloring

Theorem: (Alon and Erdős 1996)

For $r := Cn^{1/3}(\log n)^{-1/3}(\log\log n)^{1/3}$, there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Type 1: For $j \in [r/2]$, color integers in $[n/(j+1), n/j)$ color j.

Type 2: For each of the first $r/4$ primes p that do not divide n, color the multiples of p using one color.

Type 3: We can group the remaining uncolored elements in $[n-1]$ into $r/4$ color classes each with sum less than n.

The Alon-Erdős coloring

Theorem: (Alon and Erdős 1996)

For $r := Cn^{1/3}(\log n)^{-1/3}(\log\log n)^{1/3}$, there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Type 1: For $j \in [r/2]$, color integers in $[n/(j+1), n/j)$ color j.

Type 2: For each of the first $r/4$ primes p that do not divide n, color the multiples of p using one color.

Type 3: We can group the remaining uncolored elements in $[n-1]$ into $r/4$ color classes each with sum less than n.

Example: $f(39) \geq 4$. Four color classes: $[20, 38]$ and $[13, 19]$ are type 1, $\{2, 4, 6, 8, 10, 12\}$ is type 2, and $\{1, 3, 5, 7, 9, 11\}$ is type 3.

A new coloring

Theorem: (Conlon-F.-Pham)

For $r := Cn^{1/3}(\log n)^{-1/3}(\log\log n)^{-2/3}(n/\phi(n))$, there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Type 1: For $j \in [r/2]$, color integers in $[n/(j+1), n/j)$ color j.

Type 2: For each of the first $r/4$ primes p that do not divide n, color the multiples of p using one color.

A new coloring

Theorem: (Conlon-F.-Pham)

For $r := Cn^{1/3}(\log n)^{-1/3}(\log\log n)^{-2/3}(n/\phi(n))$, there is an r-coloring of $[n-1]$ such that no monochromatic subset sums to n.

Type 1: For $j \in [r/2]$, color integers in $[n/(j+1), n/j)$ color j.

Type 2: For each of the first $r/4$ primes p that do not divide n, color the multiples of p using one color.

Let d be maximum such that $(d,n)=1$ and $\phi(d) < r/16$.
For each $t \in (\mathbb{Z}/d\mathbb{Z})^{\times}$, let $x_t \in [d]$ with $x_t \equiv nt^{-1} \pmod{d}$. .

If $\sum_{i=1}^{s} a_i = n$ and each $a_i \equiv t \pmod{d}$, then $s \equiv x_t \pmod{d}$.

One color class consists of those $a \equiv t \pmod{d}$ with $a \geq n/x_t$, and one for those $a \equiv t \pmod{d}$ with $a \in [n/(x_t+d), n/x_t)$.

If a is uncolored, then $a < n/d$. Group into size d color classes.

Theorem: (Szemerédi-Vu 2006)

If $A \subset [n]$ and $|A| \geq C\sqrt{n}$, then $\Sigma(A)$ contains an n-term AP.

Long arithmetic progressions in subset sums

Theorem: (Szemerédi-Vu 2006)

If $A \subset [n]$ and $|A| \geq C\sqrt{n}$, then $\Sigma(A)$ contains an n-term AP.

An AP $a, a+d, \ldots, a+(k-1)d$ is *homogeneous* if $d|a$.

Theorem: (Freiman 1993, Sárközy 1994)

If $A \subset [n]$ and $|A| \geq C\sqrt{n \log n}$, then $\Sigma(A)$ contains an n-term HAP

Long arithmetic progressions in subset sums

Theorem: (Szemerédi-Vu 2006)

If $A \subset [n]$ and $|A| \geq C\sqrt{n}$, then $\Sigma(A)$ contains an n-term AP.

An AP $a, a+d, \ldots, a+(k-1)d$ is *homogeneous* if $d|a$.

Theorem: (Freiman 1993, Sárközy 1994)

If $A \subset [n]$ and $|A| \geq C\sqrt{n \log n}$, then $\Sigma(A)$ contains an n-term HAP

Conjecture: (Sárközy and Tran-Vu-Wood)

If $A \subset [n]$ and $|A| \geq C\sqrt{n}$, then $\Sigma(A)$ contains an n-term HAP.

Long arithmetic progressions in subset sums

Theorem: (Szemerédi-Vu 2006)

If $A \subset [n]$ and $|A| \geq C\sqrt{n}$, then $\Sigma(A)$ contains an n-term AP.

An AP $a, a+d, \ldots, a+(k-1)d$ is *homogeneous* if $d|a$.

Theorem: (Freiman 1993, Sárközy 1994)

If $A \subset [n]$ and $|A| \geq C\sqrt{n \log n}$, then $\Sigma(A)$ contains an n-term HAP

Theorem: (Conlon-F.-Pham)

If $A \subset [n]$ and $|A| \geq C\sqrt{n}$, then $\Sigma(A)$ contains an n-term HAP.

Non-averaging subsets

Definition

A set A of numbers is non-averaging if no element is the average of some of the other elements of the set.

It was known that
every non-averaging subset of $[n]$ has size $O(n^{1/2} \log n)$,
and there is a non-averaging subset of $[n]$ of size $\Omega(n^{1/4})$.

Non-averaging subsets

Definition

A set A of numbers is non-averaging if no element is the average of some of the other elements of the set.

It was known that
every non-averaging subset of $[n]$ has size $O(n^{1/2} \log n)$,
and there is a non-averaging subset of $[n]$ of size $\Omega(n^{1/4})$.

Theorem: (Conlon-F.-Pham)

If $A \subset [n]$ is non-averaging, then $|A| \leq n^{\sqrt{2}-1+o(1)}$.

Non-averaging subsets

Definition

A set A of numbers is non-averaging if no element is the average of some of the other elements of the set.

It was known that
every non-averaging subset of $[n]$ has size $O(n^{1/2}\log n)$,
and there is a non-averaging subset of $[n]$ of size $\Omega(n^{1/4})$.

Theorem: (Conlon-F.-Pham)

If $A \subset [n]$ is non-averaging, then $|A| \le n^{\sqrt{2}-1+o(1)}$.

Theorem: (Conlon-F.-Pham)

If $A \subset [n]$, $k > 1$, and $|A| \ge Cn^{1/k}$, then there is $d < k$ such that $\Sigma(A)$ contains a proper homogeneous generalized arithmetic progression of dimension d of size at least $c|A|^{d+1}$.

Erdős distinct subset sum problem

Erdős Distinct Subset Sum Conjecture

If $a_1 < \ldots < a_k$ has distinct subset sums, then $a_k = \Omega(2^k)$.

Erdős distinct subset sum problem

Erdős Distinct Subset Sum Conjecture

If $a_1 < \ldots < a_k$ has distinct subset sums, then $a_k = \Omega(2^k)$.

The sequence with $a_k = 2^{k-1}$ shows this would be optimal.

Erdős distinct subset sum problem

Erdős Distinct Subset Sum Conjecture

If $a_1 < \ldots < a_k$ has distinct subset sums, then $a_k = \Omega(2^k)$.

The sequence with $a_k = 2^{k-1}$ shows this would be optimal.

The pigeonhole principle implies $2^k \leq k a_k$, so $a_k \geq 2^k/k$.

Erdős distinct subset sum problem

Erdős Distinct Subset Sum Conjecture

If $a_1 < \ldots < a_k$ has distinct subset sums, then $a_k = \Omega(2^k)$.

The sequence with $a_k = 2^{k-1}$ shows this would be optimal.

The pigeonhole principle implies $2^k \leq k a_k$, so $a_k \geq 2^k/k$.

Erdős and Moser: $a_k \geq \frac{1}{4} 2^k/\sqrt{k}$ by the second moment method.

Erdős Distinct Subset Sum Conjecture

If $a_1 < \ldots < a_k$ has distinct subset sums, then $a_k = \Omega(2^k)$.

The sequence with $a_k = 2^{k-1}$ shows this would be optimal.

The pigeonhole principle implies $2^k \leq ka_k$, so $a_k \geq 2^k/k$.

Erdős and Moser: $a_k \geq \frac{1}{4}2^k/\sqrt{k}$ by the second moment method.

Consider random sum $X = \varepsilon_1 a_1 + \cdots + \varepsilon_k a_k$ with each $\varepsilon_i \in \{0, 1\}$.

Erdős distinct subset sum problem

Erdős Distinct Subset Sum Conjecture

If $a_1 < \ldots < a_k$ has distinct subset sums, then $a_k = \Omega(2^k)$.

The sequence with $a_k = 2^{k-1}$ shows this would be optimal.

The pigeonhole principle implies $2^k \leq k a_k$, so $a_k \geq 2^k / k$.

Erdős and Moser: $a_k \geq \frac{1}{4} 2^k / \sqrt{k}$ by the second moment method.

Consider random sum $X = \varepsilon_1 a_1 + \cdots + \varepsilon_k a_k$ with each $\varepsilon_i \in \{0, 1\}$.

$4\mathrm{Var}[X] = a_1^2 + \cdots + a_k^2 \leq k a_k^2$ and use Chebyshev's inequality.

Erdős distinct subset sum problem

Erdős Distinct Subset Sum Conjecture

If $a_1 < \ldots < a_k$ has distinct subset sums, then $a_k = \Omega(2^k)$.

The sequence with $a_k = 2^{k-1}$ shows this would be optimal.

The pigeonhole principle implies $2^k \leq ka_k$, so $a_k \geq 2^k/k$.

Erdős and Moser: $a_k \geq \frac{1}{4}2^k/\sqrt{k}$ by the second moment method.

Consider random sum $X = \varepsilon_1 a_1 + \cdots + \varepsilon_k a_k$ with each $\varepsilon_i \in \{0,1\}$.

$4\mathrm{Var}[X] = a_1^2 + \cdots + a_k^2 \leq ka_k^2$ and use Chebyshev's inequality.

Theorem (F.-Dubroff-Xu)

If $a_1 < \ldots < a_k$ has distinct subset sums, then $a_k \geq \binom{k}{\lfloor k/2 \rfloor}$.

Erdős distinct subset sum problem

Erdős Distinct Subset Sum Conjecture

If $a_1 < \ldots < a_k$ has distinct subset sums, then $a_k = \Omega(2^k)$.

The sequence with $a_k = 2^{k-1}$ shows this would be optimal.

The pigeonhole principle implies $2^k \leq ka_k$, so $a_k \geq 2^k/k$.

Erdős and Moser: $a_k \geq \frac{1}{4}2^k/\sqrt{k}$ by the second moment method.

Consider random sum $X = \varepsilon_1 a_1 + \cdots + \varepsilon_k a_k$ with each $\varepsilon_i \in \{0,1\}$.

$4\mathrm{Var}[X] = a_1^2 + \cdots + a_k^2 \leq ka_k^2$ and use Chebyshev's inequality.

Theorem (F.-Dubroff-Xu)

If $a_1 < \ldots < a_k$ has distinct subset sums, then $a_k \geq \binom{k}{\lfloor k/2 \rfloor}$.

Two proofs: One uses Harper's vertex isoperimetric inequality.

Erdős distinct subset sum problem

Erdős Distinct Subset Sum Conjecture

If $a_1 < \ldots < a_k$ has distinct subset sums, then $a_k = \Omega(2^k)$.

The sequence with $a_k = 2^{k-1}$ shows this would be optimal.

The pigeonhole principle implies $2^k \le k a_k$, so $a_k \ge 2^k / k$.

Erdős and Moser: $a_k \ge \frac{1}{4} 2^k / \sqrt{k}$ by the second moment method.

Consider random sum $X = \varepsilon_1 a_1 + \cdots + \varepsilon_k a_k$ with each $\varepsilon_i \in \{0, 1\}$.

$4\mathrm{Var}[X] = a_1^2 + \cdots + a_k^2 \le k a_k^2$ and use Chebyshev's inequality.

Theorem (F.-Dubroff-Xu)

If $a_1 < \ldots < a_k$ has distinct subset sums, then $a_k \ge \binom{k}{\lfloor k/2 \rfloor}$.

Two proofs: One uses Harper's vertex isoperimetric inequality.
Another shows that the sequence either satisfies Erdős' conjecture or the random sum X is close to a normal distribution.

Thank you!