Structural approach to subset-sum problems

Endre Szemerédi

Rutgers University, New Jersey

Alfréd Rényi Institute, Budapest

(joint work with Van H. Vu)

Notation

AP stands for Arithmetic Progression

$\mathcal{A}$ is a set of integers, $\mathcal{A}(n) = |\mathcal{A} \cap [1, n]|$.

$2\mathcal{A} = \mathcal{A} + \mathcal{A}$

$$\ell\mathcal{A} = \{a_1 + \ldots + a_\ell \mid a_i \in \mathcal{A}\}$$

is the collection of those numbers which can be represented as a sum of ℓ elements of $\mathcal{A}$.

$$\ell^*\mathcal{A} = \{a_1 + \ldots + a_\ell \mid a_i \in \mathcal{A}\}$$

is the collection of those numbers which can be represented as a sum of ℓ *different elements* of $\mathcal{A}$.

Example. (Vinogradov's theorem) If $\mathbf{P}$ is the set of primes, then $3\mathbf{P}$ contains every sufficiently large odd number.

Example. (Waring's conjecture, proved by Hilbert, Hardy, Littlewood, Hua) asserts that for any given r there are numbers $\ell_1(r)$ and $\ell_2(r)$ such that both $\ell_1 \mathbb{N}^r$ and ... contain every sufficiently large positive integer.

For a finite set $\mathcal{A}$, then natural analogue of Vinogradov-Waring results is to show that under appropriate conditions, a finite sum-set $\ell\mathcal{A}$ (resp. $\ell^*\mathcal{A}$) contains a long **AP** .

$\mathcal{A} \subseteq \{1, 2, \ldots, , n\}$

$f(|\mathcal{A}|, \ell, n)$ (resp. $f^*(|\mathcal{A}|, \ell, n)$) denotes the minimum length of the longest arithmetic progression in $\ell\mathcal{A}$, $\ell^*\mathcal{A}$.

Some earlier results:

Bourgain (1990) proved that if $|\mathcal{A}| = \gamma n$ where $\gamma > 0$ is a constant, then $2\mathcal{A}$ contains an arithmetic progression of length $e^{\varepsilon(\gamma)(\log n)^{1/3}}$.

Green improved Bourgain's result by replacing $(\log n)^{1/3}$ with $(\log n)^{1/2}$.

On the other hand I. Ruzsa constructed a set $\mathcal{A}$ of positive density, such that $|2\mathcal{A}| \leq e^{(\log n)^{2/3}}$.

Freiman, Halberstam and Ruzsa (1992) considered sum-sets modulo a prime and proved that

> Let n be a prime and $\mathcal{A}$ a set of residues modulo n. Let $|\mathcal{A}| = \gamma n$, $0 < \gamma < 1$ may depend on n. Then $\ell\mathcal{A}$ contains an arithmetic progression (modulo n) of length
>
> $$n^{\gamma/10}$$
>
> If the density of the sequence is $\leq \frac{1}{\log n}$ then the previous results do not say too much.

When $|\mathcal{A}|$ is "small",

still, something can be said: E. Croot, I. Ruzsa, T. Shoen

$\mathcal{A} \subseteq [1, n]$

$$|\mathcal{A}| \geq N^{1-\frac{1}{k-1}}$$

$\Longrightarrow$ $2\mathcal{A}$ contains an arithmetic progression of length at least k.

There is an $\mathcal{A} \subseteq [1, N]$ such that

$$|\mathcal{A}| \geq N^{1-\frac{1}{k-1}}$$

$$|\mathcal{A} + \mathcal{A}| \leq e^{k^{2/3}}.$$

Many summands

Sárközy (1990) proved that

> There are two positive constants c and C such that the following holds.
>
> If $\mathcal{A}$ is a subset of $[n]$ and ℓ is a positive integer such that $\ell|\mathcal{A}| \geq Cn$, then $\ell\mathcal{A}$ contains an arithmetic progressions of length $c\ell|A|$.

Sárközy's result is sharp up to a constant factor. (If $\mathcal{A}$ is an interval, then $\ell\mathcal{A}$ is also an interval, of length at most $|\ell\mathcal{A}|$. The most interesting case is when $\ell = |\mathcal{A}|$ and $|\mathcal{A}| > c\sqrt{n}$.)

Question: What happens if $\ell\mathcal{A} \ll n$?

(Typical case, when $\ell = n^\alpha$, $|\mathcal{A}| = n^\beta$, where $0 < \alpha, \beta < 1$.)

Question: What happens for $\ell^*\mathcal{A}$?

First focus on $\ell\mathcal{A}$ ($\ell^*\mathcal{A}$ is much harder)

For simplicity, we assume that n and ℓ are fixed and think of $f(|\mathcal{A}|, \ell, n)$ as a function on $|\mathcal{A}|$, say $g(|\mathcal{A}|)$. A. Sárközy's theorem asserts that if

$$|A| > Cn/\ell \qquad g(|A|) = \theta(\ell|\mathcal{A}|)).$$

Taking $\mathcal{A}$ to be an interval implies the upper bound $g(|\mathcal{A}|) = O(\ell|\mathcal{A}|)$.

Crucial observation

When $|\mathcal{A}| < n/\ell$, there are better upper bounds on $g(|\mathcal{A}|)$.

We present a construction with a set $\mathcal{A} \subseteq [n]$ and an ℓ such that $\ell|A| \approx n/4$, while the length of the longest arithmetic progression in $\ell\mathcal{A}$ is only $O(\ell|\mathcal{A}|^{1/2})$, which is much smaller than $\ell|\mathcal{A}|$.

Construction

$$\mathcal{A} = \{p_1x_1 + p_2x_2 \mid 1 \leq x_1, x_2 \leq m\}$$

where $p_1 \approx p_2 \approx \frac{n}{2m}$ are two primes and $p_1, p_2 > m$, and $m < \frac{1}{10}n^{1/2}$.
It is easy to see that $|\mathcal{A}| = m^2$.

Let $\ell = \frac{n}{4|\mathcal{A}|} = \frac{n}{4m^2}$. Then

$$\ell\mathcal{A} = \{p_1x_1 + p_2x_2 \mid 1 \leq x_1, x_2 \leq \ell m.\}$$

If $\mathcal{P}$ is an **AP** in $\ell\mathcal{A}$, then the coordinates of the elements of $\mathcal{P}$ form **AP** of the same length. Thus $|\mathcal{P}|$ is at most $\ell m = \ell|\mathcal{A}|^{1/2}$.

$\mathcal{A}$ is a $d+1$-dimensional cube. The general construction shows that for any fixed d there is a constant $c(d)$ such that if $\ell^d|\mathcal{A}| \leq cn$ then

$$|\ell\mathcal{A}| \leq \ell|\mathcal{A}|^{\frac{1}{d+1}}.$$

This suggests that $g(|\mathcal{A}|)$ is not a continuous function and follows a threshold behaviour, where the threshold points are

$$\frac{n}{\ell}, \ \ldots \frac{n}{\ell^2}, \ \frac{n}{\ell^d}.$$

Theorem (Van Vu-Sz. (2004)). *For any fixed positive integer d there are positive constants C and c (depending on d) such that the following holds: For any positive integers n and ℓ and any set $\mathcal{A} \subseteq [n]$ satisfying $|\mathcal{A}| \geq Cn/\ell^d$ contains an arithmetic progression of length $c\ell|\mathcal{A}|^{1/d}$.*

Corollary 1. *For any fixed positive integer d there are positive constants C_1, C_2, c_1, c_2 depending on d such that whenever*

$$\frac{C_1 n}{\ell^d} \leq |\mathcal{A}| \leq \frac{C_2 n}{\ell^{d-1}}$$

then

$$c_1 \ell |\mathcal{A}|^{\frac{1}{d}} \leq g(|\mathcal{A}|) \leq c_2 \ell |\mathcal{A}|^{1/d}.$$

The corollary confirms our intuition about thresholds. The threshold points are indeed

$$\frac{n}{\ell}, \ \ldots \frac{n}{\ell^2}, \ \frac{n}{\ell^d}.$$

$g(|\mathcal{A}|)$ behaves like $\ell|\mathcal{A}|^{1/d}$; to the left it behaves like $\ell|\mathcal{A}|^{1/(d+1)}$.

Now let us turn to $\ell^*\mathcal{A}$

Recall that

$$\ell^*\mathcal{A} = \{a_1 + \ldots + a_\ell \mid a_i \in \mathcal{A},\ a_i \neq a_j\}$$

The requirement that the summands must be different usually poses a great challenge in additive problems. One of the most well-known examples is the celebrated Erdős-Heilbronn's conjecture. In order to describe this conjecture, let us start with the classical Cauchy–Davenport theorem which asserts that if $\mathcal{A}$ is a set of residues modulo n, where n is a prime, then

$$2|\mathcal{A}| \geq \min\{n, 2|\mathcal{A}| - 1\}$$

For $\mathcal{A}$ being an arithmetic progression, the bound is sharp. Now let us consider $2^*\mathcal{A}$. We wish to bound $|2^*\mathcal{A}|$ from below with something similar to the Cauchy-Davenport bound. Observe that in the special case when $\mathcal{A}$ in an **AP** , $2^*|\mathcal{A}| = \min\{n, 2|\mathcal{A}| - 3\}$ holds for any set.

This is what Erdős and Heilbronn conjectured.

While the Cauchy-Davenport theorem is quite easy to prove, the Erdős-Heilbronn conjecture had been open for about thirty years, until it was proved by de Silva and Hamidounne in 1994.

With a lot of extra work Theorem 1 could be extended to

Theorem (Van Vu-Sz. (2004)). *For any fixed positive integer d there are positive constants C and c depending on d such that the following holds. Fix any positive integer n and ℓ and any set $\mathcal{A} \subseteq [n]$, satisfying $\ell^d|\mathcal{A}| \geq Cn$. Then $\ell^*\mathcal{A}$ contains an* **AP** *of length $c\ell|\mathcal{A}|^{1/d}$.*

While the two theorems look formally the same, Theorem 2 is a much harder result, even if $d = 1$.

Stronger, more structural results

Definition **GAP** . (generalized arithmetic progressions)

$$\mathcal{A} := \left\{ \sum_{i=1}^{d} a_i x_i \mid 0 \leq x_i \leq n_i \right\}.$$

dimension = d

Volume

$$\prod_{i=1}^{d} (n_i + 1).$$

PROPER

all $\sum a_i x_i$ are different

$$\sum_{i=1}^{d} a_i x_i \Longleftrightarrow (x_1, x_2, \ldots, x_d).$$

GAP theorem

Theorem (Van Vu-Sz. (2004)). *For any fixed positive integer d there are positive constants C and c depending on d such that the following holds. Fix any positive integer n and ℓ and any set $\mathcal{A} \subseteq [n]$, satisfying $\ell^d|\mathcal{A}| \geq Cn$. Then $\ell\mathcal{A}$ contains a* PROPER **GAP** *for some dimension $d' \leq d$, volume $c\ell^{d'}|A|$.*

This implies that $\ell\mathcal{A}$ contains an **AP** of length $c\ell|\mathcal{A}|^{1/d}$.

Same holds for $\ell^*\mathcal{A}$.

Another extension of the theorem on $\ell\mathcal{A}$

Let $\mathcal{A}_i$ be sets of integers. Define

$$\mathcal{A}_1 + \ldots + \mathcal{A}_\ell = \{a_1 + \ldots + a_\ell \mid a_i \in \mathcal{A}_i\}$$

Theorem 4 (Van Vu-Sz. (2006)). *For any fixed positive integer d there are positive constants C and c depending on d such that the following holds. For any positive integers n and ℓ and collection $\mathcal{A}_1 \subset [n]$,$\ldots$, $\mathcal{A}_\ell \subset [n]$, where $|\mathcal{A}_i| = |\mathcal{A}_j| = A$, and $\ell^d A > Cn$,*

$$\mathcal{A}_1 + \ldots + \mathcal{A}_\ell$$

contains an **AP** *of length $c\ell A^{1/d}$.*

New results

Theorem (Van Vu-Sz. (2009)). *If $\mathcal{A} \subseteq [1, n]$ and $|\mathcal{A}| > 2\sqrt{n}$ then $\mathcal{S}_{\mathcal{A}}$ contains a homogenous* **AP** *of length n.*

($\mathcal{A}$ is homogenous if $\mathcal{A} = \{d(x + c) \ : \ \ell_1 \le x \le \ell_2\}$)

O. Serra + Y. Hamidounne +A. Lada resently proved that $\bmod n$ the sumset covers all the n residue classes. (This solves an old conjecture of Olson.)

Our result implies their theorem. Our result is tight.
The following example yields the tightness:

$$\mathcal{A} = \{1, 2, [\sqrt{n}],\ n, n-1, n-2, n-[\sqrt{n}].$$

Our constans can be improved.

Applications

An infinite set $\mathcal{A}$ of positive integers is *complete* if every sufficiently large positive integer can be represented as a sum of different elements of $\mathcal{A}$

For instance, Waring's conjecture implies that the set

$$\{1^r, 2^r, 3^r, \dots, \}$$

is complete for any fixed r.

What would be necessary for a sequence to be complete?

Well, density must be the answer: one cannot hope to represent every positive integer with a *very sparse* sequence. But density itself would not be enough. The set of even numbers has very high density but clearly, is not complete. This shows that we also need a condition involving *modularity*.

In the following $\mathcal{A}(n) = |\mathcal{A} \cap [1, n]|$.

Conjecture, Erdős, 1962

There is a constant c such that the following holds. An increasing sequence $\mathcal{A} = \{a_1 < a_2 < a_3 < \ldots\}$ is complete if

(a) $\mathcal{A}(n) > cn^{1/2}$.

(b) $\mathcal{S}_{\mathcal{A}}$ contains an element of every infinite **AP** .

(This says that for any a, b there is an $s \in \mathcal{S}_{\mathcal{A}}$ that equals a modulo b.)

The bound on $\mathcal{A}(n)$ is the best possible, up to the constant factor c, as shown by Cassels, (1960).

Results

Erdős (1962) proved a weaker form of his conjecture:

It holds if one replaces (a) by a stronger condition

$$\mathcal{A}(n) > cn^{\frac{1}{2}\sqrt{5}-1}.$$

Folkman (1962) proved that $\mathcal{A}(n) > cn^{\frac{1}{2}+\varepsilon}$ is sufficient, for any constant $\varepsilon > 0$.

Hegyvári (1994) and Łuczak & Schoen (1994) independently reduced this to

$$\mathcal{A}(n) > cn^{\frac{1}{2}} \log n.$$

Theorem (Van Vu-Sz. (2003)). *Erdős' conjecture holds.*

[Related results of Chen, different approach.]

Non-decreasing sequences

An infinite sequence $\mathcal{A}$ is sub-complete if $\mathcal{S}_{\mathcal{A}}$ contains an infinite **AP** .

Again, $\mathcal{A}(n)$ denotes the number of elements of $\mathcal{A}$ in $[1, n]$. This number could be larger than n as we allow A to contain the same number many times.

In 1966 Folkman made the following conjecture:

Conjecture (Folkman). *There is a constant $C > 0$ such that the following holds. If $\mathcal{A} = \{a_1 \leq a_2 \leq a_3 \leq \ldots\}$ is an infinite non-decreasing sequence of positive integers, and $\mathcal{A}(n) \geq Cn$ for all sufficiently large n, then $\mathcal{A}$ is subcomplete.*

(If true Folkman's conjecture is tight.)

Theorem (Van Vu-Sz. (2004)). *Folkman's conjecture is true*

The number of 0-sum-free sets

$\mathcal{A}$ is called *zero-sum-free* if $0 \notin \mathcal{S}_{\mathcal{A}}$, where $\mathcal{S}_{\mathcal{A}}$ is the collection of subset sums of $\mathcal{A}$ mod n.

Olson proved that a zero-sum-free set has at most $2n^{1/2}$ elements.

So the number of zero-sum-free sets is at most

$$\sum_{i=1}^{2\sqrt{n}} \binom{n}{i} = 2^{\Omega(\sqrt{n}\log n)}.$$

Theorem (Van Vu-Sz. (2003)). *Let n be a prime. The number of zero-sum-free sets (mod n) is*

$$2^{(\sqrt{1/3}\pi \log_2 e + o(1))\sqrt{n}}$$

Why?

$\mathcal{A}$ is n-small if the sum of the elements in $\mathcal{A}$ is less than n.

The number of representations of n as a sum of different positive integers is

$$2^{\left(\sqrt{1/3}\pi \log_2 e + o(1)\right)\sqrt{n}}$$

Consequently, the number of n-small sets is

$$2^{\left(\sqrt{1/3}\pi \log_2 e + o(1)\right)\sqrt{n}}$$

Theorem (Van Vu-Sz. (2003)). $\approx$ *Most of the 0-free sets are n-small, so their number is at most*

$$2^{\left(\sqrt{1/3}\pi \log_2 e + o(1)\right)\sqrt{n}}$$

The number of x-sum-free sets

Definition . Let $x \not\equiv 0 (\text{mod } n)$.

$\mathcal{A}$ is x-sum-free, if $x \notin \mathcal{S}_{\mathcal{A}}$. (The number of x-sum-free sets is the same for every $x \not\equiv 0$)

Theorem (Van Vu-Sz. (2003)). *The number of x-sum-free sets is*

$$2^{(\sqrt{2/3}\pi \log_2 e + o(1))\sqrt{n}}$$

The reason is that a typical $\frac{1}{2}n$-sum-free set is of the form

$$\mathcal{A}_1 \cup (n - \mathcal{A}_2),$$

where $\mathcal{A}_1$ and $\mathcal{A}_2$ are $\frac{1}{2}(n-1)$ small sets.

Proof ???

Lemma 1 (Fundamental theorem of G. Freiman)**.** *For every positive constant c there is a positive integer d and a positive constant k such that the following holds. If $\mathcal{A} \subseteq \mathbb{Z}$ and $|\mathcal{A} + \mathcal{A}| \leq c|\mathcal{A}|$, then $\mathcal{A} + \mathcal{A}$ is a subset of a* **GAP** *$\mathcal{P}$ of dimension d with volume at most $k|\mathcal{A}|$.*

Lemma 2 (Generalization of I. Ruzsa)**.** *For every positive constant c there is a positive integer d and a positive constant k such that the following holds. If $\mathcal{A}, \mathcal{B} \subseteq \mathbb{Z}$ of the same cardinality and $|\mathcal{A} + \mathcal{B}| \leq c|\mathcal{A}|$, then $\mathcal{A} + \mathcal{B}$ is a subset of a* **GAP** *$\mathcal{P}$ of dimension d with volume at most $k|\mathcal{A}|$.*

Definition 1. *A set $\mathcal{A}$ is a (δ, d)-set if one can find a* **GAP** $\mathcal{Q}$ *of dimension d such that $\mathcal{B} = \mathcal{Q} \cap \mathcal{A}$ satisfies $|\mathcal{B}| > \delta \max\{|\mathcal{A}|, \mathrm{Vol}(\mathcal{Q})\}$*

Lemma 3. *For any constant $\varepsilon > 0$, and integer d there exists a constant $\delta > 0$ such that the following holds. If $|\mathcal{A} + \mathcal{A}| \leq (2^d - \varepsilon)|\mathcal{A}|$, then $\mathcal{A}$ is a (δ, d)-set.*

This is in a paper of Bilu, a direct consequence of Freiman's cube-lemma and Freiman's theorem.

Lemmas, cont, II

Lemma 4. *For any positive integer d, there is a positive δ such that the following holds. If a* **GAP** $\mathcal{Q}$ *of dimension d is proper, but $2\mathcal{Q}$ is not, then $2\mathcal{Q}$ is a $(\delta, d-1)$-set.*

Lemma 5. *For any positive constant γ, and positive integer d there is a positive constant γ' and a positive integer g such that the following holds. If $X_1, X_2, \dots, X_g$ are subsets of a* **GAP** $\mathcal{P}$*, of dimension d and $\mathrm{Vol}(X_i) > \gamma \mathrm{Vol}(\mathcal{P})$, then $X_1 + X_2 + \dots + X_g$ contains a* **GAP** $\mathcal{Q}$ *of dimension d and cardinality at least $\gamma' \mathrm{Vol}(\mathcal{P})$. Moreover, the differences of $\mathcal{Q}$ are the multiples of the differences of P.*

Lemmas, cont, III

Lemma 6. *For any positive constant γ, and positive integer d there is a positive constant γ' and a positive integer h such that the following holds. If $\mathcal{P}$ is a* **GAP** *of dimension d, and $\mathcal{B} \subset \mathcal{P}$ for which $|\mathcal{B}| > \gamma \mathrm{Vol}(\mathcal{P})$, then $h\mathcal{B}$ contains a PROPER* **GAP** *$\mathcal{Q}$ of dimension d and volume at least $\gamma' |\mathcal{B}|$.*

Lemma 7. *Let*

$$\mathcal{P} = \{x_1 a_1 + x_2 a_2 : \ 0 \leq x_i \leq \ell_i\}.$$

Let $\mathcal{P}$ be a **GAP** *of dimension 2. The $\mathcal{P}$ contains* **AP** *of length $\frac{3}{5}|\mathcal{P}|$ and difference gcd(a_1, a_2).*