A structural approach to subset sum problems

Van H. Vu
Department of Mathematics
Rutgers
vanvu@math.rutgers.edu

Let A be a subset of an (abelian) additive group G, define

$$S_A := \{\sum_{x \in B} x | B \subset A, |B| < \infty\}.$$

Two related notions

$$lA := \{a_1 + \ldots + a_l | a_i \in A\}$$

$$l^*A := \{a_1 + \ldots + a_l | a_i \in A, i \neq j\}.$$

Trivial relations

$$l^*A \subset lA \quad \cup_l l^*A = S_A.$$

Similar definition for A being a sequence (repetition allowed).

Example.

$A = \{0, 1, 4\}$, $G = Z$, $2A = \{0, 1, 2, 4, 5, 8\}$, $2^*A = \{1, 4, 5\}$,
$S_A = \{0, 1, 4, 5\}$.

$A = \{0, 1, 4\}, G = Z_5$, $2A = G$, $2^*A = \{0, 1, 4\} = S_A$.

$A = \{1, 1, 9\}$, $G = Z$, $3A = \{3, 11, 19, 27\}$, $3^*A = \{11\}$,
$S_A = \{1, 1, 2, 9, 10, 11\}$.

Many basic questions/results in additive combinatorics have the form:

If A is sufficiently dense, then S_A (or l^*A or lA) contains a special element (such as 0 or a square) or a large structure (such as a long AP or G itself).

The main question is to find out a threshold for "dense".

The most basic groups: Z, Z_p (with p large prime).

Example1. (Erdős-Zinburg-Ziv theorem, 1960s) If A is a sequence of $2p-1$ elements in Z_p, then p^*A contains zero.

Construction: $\{0^{p-1}, 1^{p-1}\}$.

Example 2. (Olson, 1960s) Let A be a subset of Z_p with cardinality $cp^{1/2}$, for a sufficiently large c, then S_A contains zero.

Construction: $\{1, 2, \ldots, \lfloor\sqrt{2p}\rfloor - 1\}$. (The sum is less than p.)

Example 3. (Olson, 1960s) Let A be a subset of Z_p with cardinality $\sqrt{4p-3}$ then $S_A = Z_p$.

Construction: $\{-m, \ldots, -1, 0, 1, \ldots, m\}$ where m is about $\sqrt{p}$, $1 + \ldots + m < \lfloor p/2 \rfloor$.

$[n] := \{1, 2, \ldots, n\}$.

Example 4. (Folkman conjecture 1960s): The following holds for sufficiently large constant C. Let A be an strictly increasing sequence of positive integer with (asymptotic) density at least $Cn^{1/2}$ (namely $|A \cap [n]| > Cn^{1/2}$ for all sufficiently large n). Then S_A contains an infinite arithmetic progression.

Construction: Cassels (1960s): $\sqrt{n}$ is best possible.

Example 5. (Erdős problem 1980s) Let A be a subset of at least $Cn^{1/3}$ elements of $[n]$. Then S_A contains a square.

Construction: $A = q, 2q, \ldots, kq$ with q prime, $(k+1)k < 2q$, $kq \leq n$.

There are several results concerning these problems (and many other), using various techniques: combinatorial, harmonic analysis, algebraic etc.

Many problems can be solved using a “structural approach”, based on the following ideas:

If A is relatively dense (close to the desired threshold) and S_A does not contain the desired object, then A has a very special structure.

If A is relatively dense (close to the desired threshold) then S_A has a special structure.

By adding new elements to A, we can obtain the desired object.

Strengthening several existing results (with classification of the extremal constructions). Solving several open questions.

A generalized arithmetic progression (GAP) of dimension d is a set of the form

$$\{a_0 + a_1 x_1 + \ldots + a_d x_d | M_i \leq x_i \leq N_i.\}$$

It is intuitive to view a GAP Q as the image of a d-dimensional box under a linear map

$$\Phi(x_1, \ldots, x_d) = a_0 + a_1 x_1 + \ldots + a_d x_d.$$

Q is proper if Φ is one-to-one.

Freiman' s theorem. A be a subset of a torsion-free group G. If $|2A| \leq C|A|$, then there is a proper GAP Q of dimension $d = d(C), |Q| = O_C(|A|)$ such that $A \subset Q$.

Informally. Being a dense subset of a proper GAP is the only reason for $2A$ to be small.

One can develop theorems of similar spirit (but usually quite different at technical level) for S_A (or lA, l^*A with l growing with $|A|$).

(Szemeredi-V. 02, Szemerédi-Nguyen-V. 05, Nguyen-V. 07) We say A is zero-sum-free if S_A does not contain 0.

Let A be a subset (sequence) of Z_p, then the main reason for A to be zero-sum-free is that its elements are *small* after a proper dilation (thus do not add up to p).

Theorem. (Nguyen-V. 07) After a proper dilation, any zero-sum-free subset A of Z_p has the form

$$A = A' \cup A''$$

where the elements of A' (viewed as integers between 0 and $p-1$) are small, $\sum_{x \in A'} x < p$ and A'' is negligible, $|A''| \le p^{6/13} \ll \sqrt{p}$.

Similar results for lA and l^*A, and for A being a sequence.

Application: Edős-Ginburg-Ziv, together with classification of extremal sets:

Gao-Panigrahi-Thangdurai (2005) If A is a sequence of cardinality at least $3/2p$ and no p elements of A add up to zero, then A is basically a sequence of two elements with high multiplicities.

Nguyen-V. (2007) True if $|A|$ has at least $p + p^{.99}$ elements.

Application: size of the largest zero-sum-free set in Z_p.

Szemerédi:$C\sqrt{p}$ (1970), Olson $2\sqrt{p}$ (1968), Hamidoune-Zemor $\sqrt{2p} + 5\log p$.

Theorem. (Deshouillers et. al., Szemerédi-Nguyen-V. 06) Let $n(p)$ be the largest integer so that $1 + \ldots + (n-1) < p$.

If $p \neq \frac{n(p)(n(p)+1)}{2} - 1$, and A is a subset of Z_p with $n(p)$ elements, then $0 \in S_A$.

If $p = \frac{n(p)(n(p)+1)}{2} - 1$, and A is a subset of Z_p with $n(p) + 1$ elements, then $0 \in S_A$. Furthermore, up to a dilation, the only zero-sum-free set with $n(p)$ elements is $\{-2, 1, 3, 4, \ldots, n(p)\}$.

Application: Structure of relatively large zero-sum-free sets

Theorem. (Deshouillers 05) (Structure of relatively large zero-sum-free sets) Let A be a subset of Z_p such that S_A does not contain 0 and A is of size at least $\sqrt{p}$. Then (after a proper dilation)

$$\sum_{x \in A, x < p/2} \|x/p\| \leq 1 + O(p^{-1/4})$$

$$\sum_{x \in A, x > p/2} \|x/p\| \leq O(p^{-1/4})$$

It is conjectured that $p^{-1/2}$ is the right error term (with a matching construction).

(Szemerédi-Nguyen-V. 06) $O(p^{-1/2})$.

We say that A is complete if $S_A = G$ and incomplete otherwise.

Theorem. (Nguyen-V. 07) After a proper dilation, any incomplete subset A of Z_p has the form

$$A = A' \cup A^{''}$$

where the elements of A' (viewed as integers between 0 and $p-1$) are small (in the integer norm) $\sum_{x \in A'} \|x/p\| < 1$ and $A^{''}$ is negligible $|A^{''}| \leq p^{6/13}$.

Similar results for lA and l^*A, and for A being a sequence.

Application: Structure of relatively large incomplete set

Theorem. (Deshouillers-Freiman) Let A be a subset of Z_p such that S_A does not contain 0 and A is of size at least $\sqrt{2p}$. Then

$$\sum_{x \in A} \|x/p\| \leq 1 + O(p^{-1/4}).$$

Again it was conjectured that $p^{-1/2}$ is the right error term (with a matching construction).

(Nguyen-V. 07) True if $|A| \geq 1.99\sqrt{p}$.

Application: Structure of long incomplete sequences.

Let $1 \leq m \leq p$ be a positive integer and A be an incomplete sequence of Z_p with maximum multiplicity $m(A) \leq m$. Trying to make A as large as possible, we come up with the following example,

$$B^m = \{-n^{[k]}, (n-1)^{[m]}, \dots, -1^{[m]}, 0^{[m]}, 1^{[m]}, \dots, (n-1)^{[m]}, n^{[k]}\}$$

where $1 \leq k \leq m$ and n are the unique integers satisfying

$$2m(1+2+\dots+n-1) + 2kn < p \leq 2m(1+2+\dots+n-1) + 2(k+1)n.$$

Nguyen-V. 07: Any long incomplete sequence can be decomposed into a subset of B^m (for some m) and a set of small cardinality (after a proper dilation).

Application: Counting problems.

Szemerédi-V. (03) The number of zero-sum-free sets in Z_p is $\exp((\sqrt{\frac{1}{3}}\pi + o(1))\sqrt{p})$.

Let $m(A)$ be the highest multiplicity in a sequence A

Nguyen-V. (2007) The number of zero-sum-free sequences A satisfying $m(A) \leq m$ is $\exp((\sqrt{(1-\frac{1}{m+1})\frac{2}{3}}\pi + o(1))\sqrt{p})$.

Similar results for incomplete sets.

Let G be a general abelian group, find the maximum size $c(G)$ of an incomplete subset ?

Diderrich conjecture (1975) $|G| = ph$, where $p \geq 3$ is the smallest prime divisor of $|G|$ and h is composite, then $c(G) = h + p - 2$.

Proved by Gao-Hamidoune (1999).

Fact. If $S_{A \cap H} = H$ for some maximal subgroup H of (prime) index q, then $|A| \leq |H| + q - 2$.

Proof. A/H is a sequence in Z_q. If a sequence B contains $q-1$ non-zero elements in Z_q, then $S_B \cup 0 = Z_q$.

A set A is sub-complete if there is a subgroup H of prime index such that $S_{A \cap H} = H$.

Threshold for sub-completeness

Gao, Hamidoune, Lladó and Serra (03) showed that (under some weak assumption) any incomplete subset of at least $\frac{p}{p+2}h + p$ elements is sub-complete. Furthermore, one can choose H to have index p.

$|G| = p_1 \dots p_k$, $p = p_1 \leq p_2 \leq \dots \leq p_k$ primes.

V. (07) (Again under some weak assumption) $|A| \geq \frac{1+\epsilon}{p_2}h$ then A is sub-complete.

$1 + \epsilon$ cannot be replaced by $1 - \epsilon$.

$G = Z$. A dense subset of $[n] = \{1, 2, \ldots, n\}$.

Theorem. (Freiman, Sárközy 90s) $|A| \geq C\sqrt{n \log n}$, then S_A contains an AP of length $c|A|^2$.

Application: Folkman's conjecture

(Luczak-Schoen, Hegyvári 94): Let A be an increasing sequence of positive integers of asymptotic density $C\sqrt{n \log n}$, then S_A contains an infinite AP. (In the 60s, Erős proved for density $n^{(\sqrt{5}-1)/2}$, Folkman proved for $n^{1/2+\epsilon}$.)

Theorem. (Szemerédi-V. 03) If $|A| \geq C\sqrt{n}$, then S_A contains an AP of length $c|A|^2$.

Application: Confirming Folkman's conjecture: Let A be an increasing sequence of positive integers of asymptotic density $C\sqrt{n}$, then S_A contains an infinite AP. (Szemerédi-V 03).

Chen (2003) proved with a stronger assumption

$$|A \cap [n]| \geq \min\{C\sqrt{n}, n\}\,, \textbf{for all } n.$$

Sz-V. theorem is sharp, for both the length of the AP and the lower bound on $|A|$. There are sets of size $\sqrt{n}/100$ such that S_A does not contain any AP of length $n^{3/4}$.

For smaller density, we can prove that S_A contains a large proper GAP of constant dimension.

For example, if $|A| \geq Cn^{1/3}$, then S_A contains either an AP of length $c|A|^2$ or a proper GAP of dimension 2 and volume $c|A|^3$. (Szemerédi-V. 04)

Application: Edős square-free problem.

Let A be a subset of at least $Cn^{1/3}$ elements of $[n]$. Then S_A contains a square.

Erdős (1988): $n/\log n$, Alon-Freiman (1989): $n^{2/3}$, Sárközy (1994): $n^{1/2}\log n$.

Theorem. (Nguyen-V. 07) If $A \subset [n]$ has cardinality at least $n^{1/3}\log n$, then S_A contains a square.

Let A be a sequence of non-zero integers, view S_A as a multi-set of 2^n elements. Let M_A be the largest multiplicity in S_A. For example, $A = \{1, \ldots, 1\}$, $M_A = \binom{n}{\lfloor n/2 \rfloor} = \Theta(2^n/\sqrt{n})$.

Littlewood-Erdős-Offord (1940s) $M_A = O(2^n/\sqrt{n})$. Many extensions by Erdös-Moser, Sárközy-Szemerédi (1960s) Katona, Kleitman, Halász (1970s), Griggs et. al., Frankl-Füredi, Stanley (1980s).

Tao-V. (2005, 2008) If $M_A \geq 2^n/n^C$ for some constant C, then (most of) A is contained in a GAP of fixed dimension d and volume $n^{C'}$, with C', d depending on C.

Application:

Conjecture. (Circular Law Conjecture 1950s)Let $\xi_{ij}, 1 \leq i, j \leq n$ be i.i.d random variables with mean 0 and variance 1 and M_n be the random matrix whose entries are ξ_{ij}. Then the limiting distribution of the eigenvalues of $\frac{1}{\sqrt{n}} M_n = \{\xi_{ij}\}$ is uniform on the unit disk.

Previous works: Ginibre-Mehta (60s), Girko (84), Bai (97), Edelman (97), Bai-Silvestein (05), Götze-Tikhomirov, Pan-Zhou (07).

Tao-V. (07) The CL conjecture holds under an extra assumption that $2 + \epsilon$ moment of ξ_{ij} exists.