VSEVOLOD F. LEV

The structure of multisets with a small number of subset sums

Astérisque, tome 258 (1999), p. 179-186
<http://www.numdam.org/item?id=AST_1999__258__179_0>

Astérisque **258**, 1999, p. 179–186

THE STRUCTURE OF MULTISETS WITH A SMALL NUMBER OF SUBSET SUMS

by

Vsevolod F. Lev

Abstract. — We investigate multisets of natural numbers with relatively few subset sums. Namely, let A be a multiset such that the number of distinct subset sums of A is bounded by a fixed multiple of the cardinality of A (that is, $|P(A)| \ll |A|$). We show that the set $P(A)$ of subset sums is then a union of a small number of arithmetic progressions sharing a common difference.

Similar problems were considered by G. Freiman (see [1]) and M. Chaimovich (see [2]). Unlike those papers, our conditions are stated in terms of the cardinality of the subset sums set $P(A)$ only and not on the largest element of the original multiset A.

The result obtained is nearly best possible.

1. Notation and definitions

By a *multiset* we mean a finite collection of natural numbers with repetitions allowed: $A = \{a_1, \dots, a_k\}$, where $a_1 \le \dots \le a_k$ are the elements of A. The number of appearances of an element will be called its *multiplicity.*

As with "normal" sets, $|A| = k$ is called the *cardinality* of A. The sum of all elements of the multiset is $\sigma(A) = a_1 + \dots + a_k$, and its *subset sums set* is

$$P(A) = \{\varepsilon_1 a_1 + \dots + \varepsilon_k a_k \colon 0 \le \varepsilon_1, \dots, \varepsilon_k \le 1\}.$$

Notice that 0 and $\sigma(A)$ are both included in $P(A)$; generally, e belongs to $P(A)$ if and only if $\sigma(A) - e$ does.

Another useful notation:

$$A = \{a_1 \times k_1, \dots, a_s \times k_s\},$$

meaning that $a_1 < \dots < a_s$ are *distinct* elements of A with multiplicities $k_1, \dots, k_s \ge 1$. In these terms, the cardinality of A is $|A| = k_1 + \dots + k_s$, the sum of its elements is $\sigma(A) = k_1 a_1 + \dots + k_s a_s$, and its subset sums set is

$$P(A) = \{\kappa_1 a_1 + \dots + \kappa_s a_s \colon 0 \le \kappa_1 \le k_1, \dots, 0 \le \kappa_s \le k_s\}.$$

1991 ***Mathematics Subject Classification.*** — 11P99, 11B75.
Key words and phrases. — Subset sums, small doubling, multisets.

2. The main result

The following theorem is our main result.

Theorem 1. — *Let A satisfy*

$$|P(A)| \le C|A| - 4C^3, \tag{1}$$

where C is a natural number, and suppose that the cardinality of A is sufficiently large: $|A| \ge 8C^3$. Then $P(A)$ is a union of at most $C-1$ arithmetic progressions with the same common difference.

Theorem 1 (the proof of which will be given in Section 5) is somewhat unusual in describing the structure of the subset sums set $P(A)$ rather then the structure of the multiset A itself. As the reader will notice, this reflects the essence of the problem: one can change A substantially without affecting $P(A)$, and thus it seems impossible to describe the structure of A under any reasonable condition on $P(A)$.

I conjecture that (1) can be replaced by the weaker restriction

$$|P(A)| \le C|A| - (C-1)^2. \tag{2}$$

The following examples show that inequality (2) cannot be further relaxed.

Example 1. — *Let $A = \{1 \times (k-C+1), b \times (C-1)\}$, where $k = |A|$ and b are sufficiently large. Then $P(A)$ is the union of C progressions*

$$\begin{gathered} 0, 1, \dots, k-C+1, \\ b, b+1, \dots, b+(k-C+1), \\ \cdot \quad \cdot \quad \cdot \\ (C-1)b, (C-1)b+1, \dots, (C-1)b+(k-C+1), \end{gathered}$$

so that $|P(A)| = C(k-C+2) = Ck - (C-1)^2 + 1$. However, $P(A)$ cannot be represented as a union of at most $C-1$ arithmetic progressions with a common difference.

Example 2. — *Let $A = \{1 \times (C-1), b \times (k-C+1)\}$, where $k = |A|$ and b are sufficiently large. Then $P(A)$ is the union of C progressions*

$$\begin{gathered} 0, b, \dots, (k-C+1)b, \\ 1, 1+b, \dots, 1+(k-C+1)b, \\ \cdot \quad \cdot \quad \cdot \\ C-1, C-1+b, \dots, C-1+(k-C+1)b, \end{gathered}$$

so that $|P(A)| = Ck - (C-1)^2 + 1$, and again $P(A)$ cannot be represented as a union of at most $C-1$ arithmetic progressions with a common difference.

Note that in view of Lemma 2 below, the inequality $|P(A)| \ge |A| + 1$ is always true. Hence, the conditions of Theorem 1 are never satisfied for $C = 1$, and from now on we assume $C \ge 2$.

3. Small values of C

For A satisfying (1) (or even (2)) with small values of C ($C = 2, 3$) the structure of $P(A)$, as well as the structure of A itself, can be completely described.

We begin with some basic properties of subset sums set. First, we estimate by how much $|P(A)|$ increases if one adds an element to A.

Lemma 1. — *Let $A = \{a_1 \times k_1, \ldots, a_s \times k_s\}$, $A^+ = A \cup \{a\}$, and suppose that A contains at least $i-1$ different elements less then a (that is, $a > a_{i-1}$ unless $i = 1$). Then*

$$|P(A^+)| \geq |P(A)| + i.$$

Proof. — $P(A^+)$ contains all the elements of $P(A)$, as well as the i additional elements

$$\sigma(A) + a, \sigma(A) + a - a_1, \ldots, \sigma(A) + a - a_{i-1}.$$

□

As a direct corollary, we obtain a lower-bound estimate for $|P(A)|$.

Lemma 2. — *The cardinality of the subset sums set $P(A)$ of the multiset*

$$A = \{a_1 \times k_1, \ldots, a_s \times k_s\}$$

satisfies

$$|P(A)| \geq 1 + k_1 + 2k_2 + \cdots + sk_s.$$

In particular, $|P(A)| \geq 1 + |A|$.

Proof. — The assertion is obviously true for $|A| = 1$, and we use induction on $|A|$. Denote by A^- the multiset obtained by removing from A its largest element a_s. Applying Lemma 1, we obtain then

$$\begin{aligned} |P(A)| &\geq |P(A^-)| + s \geq (1 + k_1 + 2k_2 + \cdots + s(k_s - 1)) + s \\ &= 1 + k_1 + 2k_2 + \cdots + sk_s. \end{aligned}$$

□

It follows from Lemma 2 that a multiset A with relatively small value of $|P(A)|$ has at least one element with large multiplicity.

Lemma 3. — *Let $A = \{a_1 \times k_1, \ldots, a_s \times k_s\}$, and let $k_0 = \max_{1 \leq i \leq s} k_i$ be the maximal multiplicity of an element of A. Then*

$$k_0 > \frac{k^2}{2|P(A)|}.$$

Proof. — For $1 \leq i \leq s$ we have:

$$\begin{aligned} |P(A)| &\geq 1 + k_1 + 2k_2 + \cdots + ik_i + (i+1)(k_{i+1} + \cdots + k_s) \\ &> (i+1)k - (k_i + 2k_{i-1} + \cdots + ik_1) \\ &\geq (i+1)k - \frac{1}{2}i(i+1)k_0. \end{aligned}$$

The resulting estimate

$$|P(A)| > (i+1)k - \frac{1}{2}i(i+1)k_0$$

also holds for $i > s$, as in this case the expression in the right-hand side, considered as a function of real i, has a negative derivative:

$$k - \frac{1}{2}(2i+1)k_0 < k - sk_0 \leq 0.$$

Hence,

$$k_0 > \frac{2}{i}\left(k - \frac{|P(A)|}{i+1}\right)$$

for every $i = 1, 2, \dots$ We choose i under the condition

$$2\frac{|P(A)|}{k} - 1 \leq i < 2\frac{|P(A)|}{k}.$$

Then

$$\frac{2}{i} > \frac{k}{|P(A)|}, \quad \frac{|P(A)|}{i+1} \leq \frac{k}{2},$$

and so

$$k_0 > \frac{k}{|P(A)|} \cdot \frac{k}{2} = \frac{k^2}{2|P(A)|}.$$

□

We now construct multisets whose subset sums sets have a particularly simple structure.

Example 3. — *Let $A = \{a_1, \dots, a_k\}$ be a multiset such that*

i) $a_2, \dots, a_k \equiv 0 \pmod{a_1}$;

ii) $a_{i+1} \leq a_1 + \dots + a_i$ for $i = 1, \dots, k-1$.

Then $P(A)$ is an arithmetic progression: $P(A) = \{0, a_1, 2a_1, \dots, \sigma(A)\}$.

This easily follows by induction on k: if $A^- = \{a_1, \dots, a_{k-1}\}$, then

$$\begin{aligned} P(A) &= P(A^-) \cup (a_k + P(A^-)) \\ &= \{0, a_1, \dots, \sigma(A^-)\} \cup \{a_k, a_k + a_1, \dots, a_k + \sigma(A^-)\} \\ &= \{0, a_1, \dots, \sigma(A)\}, \end{aligned}$$

since $a_k \leq \sigma(A^-)$ and $a_k + \sigma(A^-) = \sigma(A)$.

Proposition 1. — *Any multiset A, satisfying $|P(A)| \leq 2|A| - 1$ (that is satisfying (2) with $C = 2$) has the structure, described in Example 3.*

Proof. — Suppose, on the contrary, that there exists an index $2 \leq i \leq k$ for which either $a_i \not\equiv 0 \pmod{a_1}$ or $a_i > a_1 + \dots + a_{i-1}$; we assume, moreover, that i is the *minimum* index with this property. Then, writing $A_j = \{a_1, \dots, a_j\}$ $(j = 1, \dots, k)$ and applying Lemma 2, we obtain

$$|P(A_i)| = 2|P(A_{i-1})| \geq 2i$$

(since $P(A_i) = P(A_{i-1}) \cup (a_i + P(A_{i-1}))$, and $P(A_{i-1})$ is disjoint with $a_i + P(A_{i-1})$), therefore

$$|P(A)| = |P(A_k)| \geq |P(A_{k-1})| + 2 \geq \cdots \geq |P(A_i)| + 2(k-i) \geq 2k.$$

$\square$

The following example describes the construction of multisets whose subset sums set consists of exactly *two* arithmetic progressions.

Example 4. — *Let* $A = \{a_1, \ldots, a_m\} \cup \{b_1, b_2\} \cup \{c_1, \ldots, c_n\}$, *where*
- $a_m < b_1 \leq c_1$;
- $A_m = \{a_1, \ldots, a_m\}$ *satisfies conditions (i) and (ii) of Example 3 with* $a_1 = 2$;
- $b_1, b_2 \not\equiv 0 \pmod 2$;
- $b_1 + b_2 \leq \sigma(A_m) + 2$;
- $c_{i+1} \leq \sigma(C_i) - 2b_1 + 3 \quad (0 \leq i \leq n-1)$, *where* $C_i = \{a_1, \ldots, a_m\} \cup \{b_1, b_2\} \cup \{c_1, \ldots, c_i\}$.

Then $P(A)$ *is a union of two progressions with the common difference* 2: *if* $\sigma(A)$ *is even, then*

$$P(A) = \{0, 2, \ldots, \sigma(A) - 2, \sigma(A)\} \cup \{b_1, b_1 + 2, \ldots, \sigma(A) - b_1\},$$

and if $\sigma(A)$ *is odd, then*

$$P(A) = \{0, 2, \ldots, \sigma(A) - b_1\} \cup \{b_1, b_1 + 2, \ldots, \sigma(A)\}.$$

In either case, $|P(A)| = \sigma(A) - b_1 + 2$.

The verification is left to the interested reader.

Proposition 2. — *Any multiset* A *with co-prime elements satisfying* $|P(A)| \leq 3|A| - 4$ *(that is satisfying (2) with* $C = 3$*) has either the structure described in Example 3, or the structure described in Example 4.*

This proposition will not be used in the sequel and is given just for completeness. Its proof (which is rather long and tedious) is available from the author.

4. More lemmas and properties of $P(A)$

In this section, we prepare for the proof of Theorem 1. To this end, we first determine the value of $|P(A)|$ for multisets A with only two different elements. Without loss of generality we can restrict ourselves to the case when these two elements are co-prime.

Lemma 4. — *Let* $A = \{a_1 \times k_1, a_2 \times k_2\}$, *where* $(a_1, a_2) = 1$. *Then*

i) *if* $k_1 \leq a_2 - 1$ *or* $k_2 \leq a_1 - 1$, *then*

$$|P(A)| = (k_1 + 1)(k_2 + 1);$$

ii) *if* $k_1 \geq a_2 - 1$ *and* $k_2 \geq a_1 - 1$, *then*

$$|P(A)| = a_1 k_1 + a_2 k_2 - (a_1 - 1)(a_2 - 1) + 1.$$

Proof. — In Case i), the assertion follows from the fact that all values of the linear form

$$a_1x + a_2y; \quad 0 \le x \le k_1,\ 0 \le y \le k_2$$

are pairwise distinct: if, for instance, $k_1 \le a_2 - 1$, and $a_1x + a_2y = a_1x' + a_2y'$, then $x \equiv x' \pmod{a_2}$; therefore (in view of $0 \le x, x' \le k_1 < a_2$) we have $x = x'$, whence $y = y'$.

In Case ii) we use induction on k_2. For $k_2 = a_1 - 1$ we apply the above proved:

$$\begin{aligned} |P(A)| &= (k_1+1)(k_2+1) = a_1k_1 + a_1 \\ &= a_1k_1 + a_2k_2 - a_2k_2 + a_1 \\ &= a_1k_1 + a_2k_2 - a_1a_2 + a_1 + a_2 \\ &= a_1k_1 + a_2k_2 - (a_1-1)(a_2-1) + 1. \end{aligned}$$

Suppose now that $k_2 > a_1 - 1$. Write $A^- = \{a_1 \times k_1, a_2 \times (k_2 - 1)\}$, so that

$$|P(A^-)| = a_1k_1 + a_2(k_2 - 1) - (a_1 - 1)(a_2 - 1) + 1.$$

We have to prove, therefore, that $|P(A)| = |P(A^-)| + a_2$. Obviously, the difference $|P(A)| - |P(A^-)|$ counts the numbers of the form

$$xa_1 + k_2a_2;\ 0 \le x \le k_1, \tag{3}$$

which cannot be represented in the form

$$xa_1 + ya_2; \quad 0 \le x \le k_1,\ 0 \le y \le k_2 - 1.$$

We show that this particular subset of (3) is obtained when $k_1 - a_2 + 1 \le x \le k_1$; that is, there exist exactly a_2 such numbers. Indeed, if $x < k_1 - a_2 + 1$ then the number $e = xa_1 + k_2a_2$ possesses the representation $e = (x + a_2)a_1 + (k_2 - a_1)a_2$. On the other hand, for $x \ge k_1 - a_2 + 1$ the equality $xa_1 + k_2a_2 = x'a_1 + ya_2$ is impossible: otherwise $x' \equiv x \pmod{a_2}$, meaning that $x' \le x$, and then $xa_1 + k_2a_2 > x'a_1 + ya_2$, a contradiction. □

The following lemma shows that under certain conditions, a multiset can be slightly modified in such a way that the number of its elements will increase while its subset sums set will not change. Once again, we start with multisets with exactly two distinct elements.

Lemma 5. — *Let* $A = \{a_1 \times k_1, a_2 \times k_2\}$, *where*

$$a_1 < a_2,\ k_1 \ge a_2 - 1,\ k_2 \ge 2a_1 - 1.$$

Then there exist k_1', k_2' *such that the multiset* $A' = \{a_1 \times k_1', a_2 \times k_2'\}$ *satisfies*

$$P(A') = P(A), \quad |A'| > |A|.$$

Proof. — We set $k_1' = k_1 + a_2$, $k_2' = k_2 - a_1$. Since $k_1' + k_2' = k_1 + k_2 + (a_2 - a_1) > k_1 + k_2$, we have only to prove that $P(A') = P(A)$.

1) Suppose that $e = xa_1 + ya_2 \in P(A)$, where $0 \le x \le k_1,\ 0 \le y \le k_2$, and show that $e \in P(A')$. Indeed, this is trivial if $y \le k_2 - a_1$, and otherwise it follows from $e = (x + a_2)a_1 + (y - a_1)a_2$.

2) Suppose that $e = xa_1 + ya_2 \in P(A')$, where $0 \le x \le k_1'$, $0 \le y \le k_2'$, and show that $e \in P(A)$. Indeed, this is trivial if $x \le k_1$, and otherwise this follows from $e = (x - a_2)a_1 + (y + a_1)a_2$. □

We now wish to bring the assumptions of Lemma 5 to a more convenient form, as well as to extend this lemma for the case of multisets with arbitrarily many distinct elements.

Lemma 5'. — *Let $A = \{a_1 \times k_1, \dots, a_s \times k_s\}$, and suppose that some two multiplicities k_i, k_j $(1 \le i < j \le s)$ satisfy $k_i k_j \ge 2(|P(A)| - k)$. Then there exists a multiset A' such that*

$$P(A') = P(A), \quad |A'| > |A|.$$

Proof. — Write $A_0 = \{a_i \times k_i, a_j \times k_j\}$ and $A_1 = A \setminus A_0$ (so that $A = A_0 \cup A_1$). We denote $d = (a_i, a_j)$ and set $a_i' = a_i/d$, $a_j' = a_j/d$. Clearly,

$$|P(A)| \ge |P(A_0)| + |P(A_1)| - 1 \ge |P(A_0)| + |A_1|,$$
$$|P(A_0)| \le (\tfrac{1}{2}k_i k_j + k) - (k - k_i - k_j) = \tfrac{1}{2}k_i k_j + k_i + k_j < (k_i + 1)(k_j + 1),$$

whence, in view of Lemma 4, $k_i \ge a_j'$ and $k_j \ge a_i'$. Moreover, applying Lemma 4 once more (this time part (ii)) we obtain:

$$\begin{aligned} |P(A_0)| &= k_i a_i' + k_j a_j' - (a_i' - 1)(a_j' - 1) + 1 \\ &> k_i a_i' + k_j + k_j(a_j' - 1) - (a_i' - 1)(a_j' - 1) \\ &\ge k_i a_i' + k_j, \end{aligned}$$

which implies

$$\tfrac{1}{2}k_i k_j + k_i + k_j > k_i a_i' + k_j,$$
$$k_j > 2a_i' - 2.$$

This allows us to apply Lemma 5 to A_0 (more precisely, to the multiset $\{a_i' \times k_i, a_j' \times k_j\}$) to find A_0' with $P(A_0') = P(A_0)$, $|A_0'| > |A_0|$. Then the multiset $A' = A_0' \cup A_1$ will obviously satisfy the required conditions $P(A') = P(A)$, $|A'| > |A|$. □

5. Proof of the main theorem

Two multisets A and A' will be called *equivalent*, if $P(A) = P(A')$. Without loss of generality we can assume that A is a multiset of the maximum possible cardinality of all equivalent multisets. We write A in the form

$$A = \{a_0 \times k_0\} \cup B, \ B = \{b_1 \times k_1, \dots, b_s \times k_s\},$$

where $k_1, \dots, k_s \le k_0$, and $b_1, \dots, b_s \ne a_0$.

By a *chain* we will mean a sequence $E = \{e_1, \dots, e_t\}$ of the elements of $P(B)$, satisfying the two following conditions:

i) $0 < e_{i+1} - e_i \le k_0 a_0$; $\quad i = 1, \dots, t-1$;

ii) $e_1 \equiv \dots \equiv e_t \pmod{a_0}$.

The chain E will be referred to as *maximal* if no more elements of $P(B)$ can be added to E without violating either (i) or (ii).

Let $S = P(\{a_0 \times k_0\}) = \{0, a_0, \ldots, k_0 a_0\}$. It is obvious that:

- if E is a chain, then the sum $S + E$ is an arithmetic progression with the difference a_0;
- if E_1 and E_2 are two distinct maximal chains, then the progressions $S + E_1$ and $S + E_2$ are disjoint.

Clearly, there is exactly one way to decompose $P(B)$ into maximal chains, and we denote the number of these chains by N. We assume $N \geq C$ (since otherwise $P(A)$ consists of at most $C - 1$ progressions with the difference a_0) and show that this assumption leads to a contradiction.

Since obviously $|P(A)| - |P(B)| \geq N k_0$, we obtain

$$|P(B)| \leq |P(A)| - N k_0 < C(|A| - k_0) = C|B|$$

(in fact, one can easily prove that B satisfies (2)). Therefore, by Lemma 3,

$$\max_{1 \leq i \leq s} k_i \geq \frac{|B|^2}{2|P(B)|} > \frac{|B|}{2C}.$$

By Lemma 5′ and in view of the maximality of A,

$$k_0 \cdot \tfrac{|B|}{2C} < 2(|P(A)| - k) < 2(C-1)k,$$
$$k_0(k - k_0) < 4C(C-1)k.$$

The left-hand side of the last inequality is a quadratic polynomial of k_0 with zeroes at 0 and k, maximum at $k/2$, and attaining both at $4C^2$ and $k - 4C^2$ the same common value

$$4C^2(k - 4C^2) \geq 4C(C-1)k.$$

Therefore, either $k_0 < 4C^2$ or $k_0 > k - 4C^2$ holds true.

The first is actually impossible, since by Lemma 3, $k_0 > \frac{k^2}{2Ck} \geq 4C^2$. Hence $k_0 > k - 4C^2$, and it follows that

$$|P(A)| \geq N(k_0 + 1) > C(k - 4C^2) = Ck - 4C^3,$$

a contradiction with (1). (Notice that this is the only place where we use (1) instead of the weaker (2).) This completes the proof of Theorem 1.

References

[1] Freiman G.A., *Subset-sum problem with different summands*, Congressus Numerantium, **70**, 1990, 207–215.

[2] Chaimovich M., *Solving a value-independent knapsack problem with use of methods of additive number theory*, Congressus Numerantium, **72**, 1990, 115–123.

V.F. Lev, Department of Mathematics, University of Georgia, Athens, GA 30602
E-mail : seva@math.tau.ac.il • *Url :* http://www.math.uga.edu:80/~seva/