



ON NATHANSON'S TRIANGULAR NUMBER PHENOMENON

Kevin O'Bryant

*Department of Mathematics, City University of New York, The Graduate Center
and The College of Staten Island, New York, New York*

kevin.obryant@csi.cuny.edu

Received: 6/25/25, Accepted: 2/11/26, Published: 9/28/26

Abstract

For a finite set $A \subseteq \mathbb{Z}$, the h -fold sumset is $hA := \{x_1 + \cdots + x_h : x_i \in A\}$. We interpret the beginning of the sequence of sumset sizes $(|hA|)_{h=1}^\infty$ in terms of the successive L^1 -minima of the lattice of points in $\mathbb{Z}^{|A|}$ whose coordinates sum to 0 and which are perpendicular to $\langle a_1, \dots, a_{|A|} \rangle$. In particular, if h_1, h_2 are the first and second minima, and $1 \leq h < h_1$, then $|hA| = \binom{h+|A|-1}{|A|-1}$, while if $h_1 \leq h < h_2$, then $|hA| = \binom{h+|A|-1}{|A|-1} - \binom{h-h_1+|A|-1}{|A|-1}$. This explains the appearance of triangular numbers in the sequence of sumset sizes, an observation related to a recent experiment of Nathanson.

1. Introduction

The h -fold sumset of a set A of k integers is $hA := \{x_1 + \cdots + x_h : x_i \in A\}$. By elementary counting, its size satisfies both

$$|hA| \leq h \cdot \text{diam}(A) + 1 \quad \text{and} \quad |hA| \leq \binom{h+k-1}{k-1}.$$

By Nathanson's Theorem [1], there is a $C \leq 1$ so that for sufficiently large h we have the equality $|hA| = h \cdot \text{diam}(A) + C$. For sufficiently small h , we have $|hA| = \binom{h+k-1}{k-1}$. In this work, we address the meaning of "sufficiently small", and give an explicit formula for $|hA|$ for h that is slightly larger than sufficiently small.

For example, consider 4-element sets. As there are $\binom{h+3}{3}$ unordered h -tuples of elements of A , the quantity $\binom{h+3}{3} - |hA|$ counts the "missing" sums. A specific set A is considered in Table 1. The appearance of the triangular numbers (A000217) in the start of the bottom row of Table 1 was surprising, and the main theorem of this work explains its appearance for all 4-element sets A . Namely, we prove that the second row of such a table begins with tetrahedral numbers (A000292).

h	1	2	3	4	5	6	7	8	9	10	11	12
$ hA $	4	10	20	34	52	74	100	130	162	193	222	249
$\binom{h+3}{3} - hA $	0	0	0	1	4	10	20	35	58	93	142	206
first differences		0	0	1	3	6	10	15	23	35	49	64

Table 1: The size of some sumsets of $A = \{0, 2, 18, 25\}$, and a surprising appearance of many triangular numbers $1, 3, 6, 10, \dots$

The string of numbers in the first row, which are of the form $2(h^2 + 1)$, is also interesting, not typical, and explained below in Corollary 1. In Section 5.1 below, we repeat a related experiment of Nathanson [8] in which the triangular numbers appear. Indeed, this work was directly inspired by that work.

1.1. Notation and Conventions

We use $\mathbb{N}$ for the positive integers, and $\mathbb{N}_0$ for the set $\{0\} \cup \mathbb{N}$. For each finite set $A \subseteq \mathbb{R}$, we denote its elements as $a_1 < \dots < a_k$, so that $|A| = k$ and $\text{diam}(A) = a_k - a_1$. We use the convention that binomial coefficients are zero if their top argument is smaller than their bottom argument, and the convention that if X is a set and k a nonnegative integer, then $\binom{X}{k}$ is the set of subsets of X with cardinality k .

The sumset hA is the set $\{b_1 + \dots + b_h : b_i \in A\}$. Recursively, $1A := A$ and $hA := \{s + b : s \in (h-1)A, b \in A\}$. The size of hA is invariant under translation and dilation of A , often enabling us to assume without loss of generality that $\min A = 0$ and that $\gcd(A) = 1$.

Let

$$\mathcal{X}_{t,k} := \{\langle c_1, \dots, c_k \rangle : c_i \in \mathbb{N}_0, c_1 + \dots + c_k = t\}.$$

With $\bar{\mathbf{a}} := \langle a_1, a_2, \dots, a_k \rangle$, we have

$$hA = \{\bar{\mathbf{x}} \cdot \bar{\mathbf{a}} : \bar{\mathbf{x}} \in \mathcal{X}_{h,k}\}. \quad (1)$$

By stars-and-bars,

$$|\mathcal{X}_{t,k}| = \binom{t+k-1}{k-1}.$$

For each set $A \subseteq \mathbb{Z}$, we let $\mathcal{L}$ be the $(k-2)$ -dimensional lattice in $\mathbb{Z}^k$ of vectors that are perpendicular to both $\bar{\mathbf{1}}$ (the vector of all 1's) and $\bar{\mathbf{a}}$. This lattice, which we call *the coefficient lattice of A* , has dimension $k-2$, and consequently for $k \geq 4$ this dimension is at least 2. The entries of an element of $\mathcal{L}$ sum to 0 and are integers, so the sum of their absolute values is an even integer.

We define the i -th successive L^1 -minima λ_i of $\mathcal{L}$ to be the infimum of λ such that $\{\bar{\mathbf{c}} : \|\bar{\mathbf{c}}\|_1 \leq \lambda\}$ contains i linearly independent vectors. We let $\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \dots, \bar{\mathbf{y}}_{k-2}$ be

(the nonuniquely defined) linearly independent vectors in $\mathcal{L}$ with $\|\bar{\mathbf{y}}_i\|_1 = \lambda_i$, and call them *minimizers*.

For the coefficient lattices we consider, λ_i must be an even integer since it is the sum of the absolute values of the entries of an element of $\mathcal{L}$. We define h_1, h_2 so that $\lambda_1 = 2h_1, \lambda_2 = 2h_2$ are the first and second minima of the lattice with respect to the L^1 -norm. That is, every nonzero vector in $\mathcal{L}$ has L^1 -norm at least $2h_1$, there is a first minimizer $\bar{\mathbf{y}}_1 \in \mathcal{L}$ with $\|\bar{\mathbf{y}}_1\|_1 = 2h_1$, and every vector in $\mathcal{L}$ that is linearly independent of $\bar{\mathbf{y}}_1$ has L^1 -norm at least $2h_2$.

By Lemma 1 below, every pair of even numbers $2h_1, 2h_2$ with $2 \leq h_1 \leq h_2$ arises as the successive minima of such a lattice, even with $k = 4$. The proof of Lemma 1 is delayed to §2.1.

Lemma 1. *For every integer $k \geq 4$ and every pair of integers $2 \leq h_1 \leq h_2$, there is a k -element set A whose coefficient lattice $\mathcal{L}$ has first L^1 -minimum $2h_1$ and second L^1 -minimum $2h_2$.*

We conjecture that for every $1 \leq h_1 \leq h_2 \leq \dots \leq h_{k-2}$, there is a set A of k integers whose coefficient lattice has successive L^1 -minima $2h_1, \dots, 2h_{k-2}$.

1.2. Theorems and Corollaries

Our main theorem connects the minima of the coefficient lattice $\mathcal{L}$ to the sequence of sumset sizes.

Theorem 1. *Let $A = \{a_1, \dots, a_k\}$ be a set of $k \geq 4$ integers. Let $\mathcal{L}$ be the sublattice of $\mathbb{Z}^k$ that is perpendicular to both $\bar{\mathbf{1}}$ and $\bar{\mathbf{a}} = \langle a_1, \dots, a_k \rangle$. Let $2h_1, 2h_2$ be the first and second minima of $\mathcal{L}$ with respect to the L^1 -norm. If $1 \leq h < h_1$, then*

$$|hA| = \binom{h+k-1}{k-1}. \quad (2)$$

If $h_1 \leq h < h_2$, then

$$|hA| = \binom{h+k-1}{k-1} - \binom{h-h_1+k-1}{k-1}. \quad (3)$$

Thus, A is a B_h -set if $h < h_1$. Further, for $h_1 < h < h_2$ the difference between the number of consecutive missing sums is

$$\begin{aligned} & \left(\binom{h+k-1}{k-1} - |hA| \right) - \left(\binom{(h-1)+k-1}{k-1} - |(h-1)A| \right) \\ &= \binom{h-h_1+k-1}{k-1} - \binom{(h-1)-h_1+k-1}{k-1} = \binom{h-h_1+k-2}{k-2}. \end{aligned}$$

If $k = 4$, with $t = h - h_1 + 2$, this is $\binom{t}{2}$, the triangular numbers seen in the bottom row of the table above. We comment that with $k = 4$, both $\binom{h+k-1}{k-1}$ and $\binom{h-h_1+k-1}{k-1}$ are tetrahedral numbers [A000292](#).

Theorem 1 tells us that the sequence of sumset sizes of a set of k integers is given exactly by a degree $k - 1$ polynomial for the first several h , and exactly by a degree $k - 2$ polynomial for the next several h . The author suspects, but has not proven, that if $\lambda_{i-1} \leq h < \lambda_i$, then $|hA|$ is “essentially” a degree $k - i$ polynomial whose coefficients are determined by the first $i - 1$ minimizers. The vague “essentially” qualifier is necessary to handle congruence issues that can arise for larger i , as seen in the Frobenius Problem.

Returning to the example above, $A = \{0, 2, 18, 25\}$, examining the coefficient lattice $\mathcal{L}$ we find that $h_1 = 4$ and $h_2 = 9$. By Theorem 1, $|hA| = \binom{h+3}{3}$ for $1 \leq h < 4$, and $|hA| = \binom{h+3}{3} - \binom{h-1}{3} = 2(h^2 + 1)$ for $4 \leq h < 9$. That $|hA| = 2(h^2 + 1)$ also for $1 \leq h \leq 3$ is a coincidence arising from the coincidence $\binom{h+3}{3} = 2(h^2 + 1)$ (for $1 \leq h \leq 3$). The set $A = \{0, 1, 3b + 1, 3b + 4\}$ gives us the pleasant Corollary 1.

Corollary 1. *For all $b > 1$, there is a set A of integers with $|hA| = 2(h^2 + 1)$ for $1 \leq h \leq b$.*

Call a polynomial $p(h)$ *cute* if for all $b > 1$, there is a set A of integers with $|hA| = p(h)$ for $1 \leq h \leq b$. We are unable to enumerate the cute polynomials.

In [8], Nathanson lists a number of problems concerning

$$\mathcal{R}_{\mathbb{Z}}(h, k) := \{|hA| : A \subseteq \mathbb{Z}, |A| = k\}.$$

Problem 8 of his program is solved by Corollary 2, whose proof gives an explicit set A for each of the claimed sizes of $|hA|$.

Corollary 2. *Let h be a positive integer. Then*

$$\left\{ \binom{h+k-1}{k-1} - \binom{h+k-j}{k-1} : 1 < j \leq h \right\} \subseteq \mathcal{R}_{\mathbb{Z}}(h, k).$$

While of a different flavor, the proof of the following result has some conceptual and stylistic overlap with our proof of Theorem 1. In [4], Nathanson writes “it is not known if $\mathcal{R}_{\mathbb{N}_0}(h, k) = \mathcal{R}_{\mathbb{R}}(h, k)$.” As $\mathcal{R}_{\mathbb{N}_0}(h, k) = \mathcal{R}_{\mathbb{N}}(h, k)$ trivially, the first equality in our Theorem 2 fills this gap in the literature. Let $A^h := \{a_1 \cdots a_h : a_i \in A\}$ be the h -fold product set.

Theorem 2. *We have*

$$\begin{aligned} \{|hA| : |A| = k, A \subseteq \mathbb{N}\} &= \{|hA| : |A| = k, A \subseteq \mathbb{R}\} \\ &= \{|A^h| : |A| = k, A \subseteq \mathbb{N}\} = \{|A^h| : |A| = k, A \subseteq \mathbb{R}\}. \end{aligned}$$

2. Proof of Theorem 1 and Corollaries

Proof of Theorem 1. Suppose that we have distinct $\overline{\mathbf{c}}_1, \overline{\mathbf{c}}_2 \in \mathcal{X}_{h,k}$ with $\overline{\mathbf{c}}_1 \cdot \overline{\mathbf{a}} = \overline{\mathbf{c}}_2 \cdot \overline{\mathbf{a}}$. Then $\overline{\mathbf{c}}_1 - \overline{\mathbf{c}}_2 \in \mathcal{L}$, and

$$2h_1 \leq \|\overline{\mathbf{c}}_1 - \overline{\mathbf{c}}_2\|_1 \leq \|\overline{\mathbf{c}}_1\|_1 + \|\overline{\mathbf{c}}_2\|_1 = 2h. \quad (4)$$

If $h < h_1$, this is impossible, and so (using Equation (1)) $|hA| = |\mathcal{X}_{h,k}| = \binom{h+k-1}{k-1}$, as claimed.

We now assume that $h_1 \leq h < h_2$, and take $\overline{\mathbf{y}} \in \mathcal{L}$ with $\|\overline{\mathbf{y}}\|_1 = 2h_1$.

The h -type of A is the equivalence relation on $\mathcal{X}_{h,k}$ defined by

$$\overline{\mathbf{c}}_1 \equiv \overline{\mathbf{c}}_2 \quad \text{if} \quad \overline{\mathbf{c}}_1 \cdot \overline{\mathbf{a}} = \overline{\mathbf{c}}_2 \cdot \overline{\mathbf{a}}.$$

The number of classes in the h -type of A is $|hA|$.

We define a second equivalence relation on $\mathcal{X}_{h,k}$ by defining $\overline{\mathbf{c}}_1 \cong \overline{\mathbf{c}}_2$ if $\overline{\mathbf{c}}_1 - \overline{\mathbf{c}}_2$ is a multiple of $\overline{\mathbf{y}}$. We will first prove that the two equivalence relations, denoted with $\cong$ and $\equiv$, are identical. Then we will prove that the $\cong$ equivalence relation has $\binom{h+k-1}{k-1} - \binom{h-h_1+k-1}{k-1}$ classes.

First, suppose that $\overline{\mathbf{c}}_1 \equiv \overline{\mathbf{c}}_2$ and $\overline{\mathbf{c}}_1 \neq \overline{\mathbf{c}}_2$. Then $(\overline{\mathbf{c}}_1 - \overline{\mathbf{c}}_2) \cdot \overline{\mathbf{a}} = 0$, and

$$(\overline{\mathbf{c}}_1 - \overline{\mathbf{c}}_2) \cdot \overline{\mathbf{1}} = \overline{\mathbf{c}}_1 \cdot \overline{\mathbf{1}} - \overline{\mathbf{c}}_2 \cdot \overline{\mathbf{1}} = h - h = 0,$$

so that $\overline{\mathbf{c}}_1 - \overline{\mathbf{c}}_2 \in \mathcal{L}$. But since $\|\overline{\mathbf{c}}_1 - \overline{\mathbf{c}}_2\|_1 = 2h$, and $2h < 2h_2$, it must be that $\overline{\mathbf{c}}_1 - \overline{\mathbf{c}}_2 = \alpha \overline{\mathbf{y}}$ for some $\alpha \in \mathbb{Z}$. Ergo, $\overline{\mathbf{c}}_1 \cong \overline{\mathbf{c}}_2$.

Now, suppose that $\overline{\mathbf{c}}_1 \cong \overline{\mathbf{c}}_2$ and $\overline{\mathbf{c}}_1 \neq \overline{\mathbf{c}}_2$. Then $\overline{\mathbf{c}}_1 - \overline{\mathbf{c}}_2 = \alpha \overline{\mathbf{y}}$. Consequently, $(\overline{\mathbf{c}}_1 - \overline{\mathbf{c}}_2) \cdot \overline{\mathbf{a}} = \alpha \overline{\mathbf{y}} \cdot \overline{\mathbf{a}} = 0$, since $\overline{\mathbf{y}} \in \mathcal{L}$. Thus, $\overline{\mathbf{c}}_1 \equiv \overline{\mathbf{c}}_2$.

It remains only to count the $\cong$ -classes. Each class is finite, so has a representative $\overline{\mathbf{c}}$ with $\overline{\mathbf{c}} - \overline{\mathbf{y}}$ not in the class. Every vector in the class with $\overline{\mathbf{c}}$ has the form $\overline{\mathbf{c}} + i\overline{\mathbf{y}}$, and we let α be the maximal such i . The classes are convex, being the intersection of a space with the convex $\mathcal{X}_{h,k}$, so the entire class is $\overline{\mathbf{c}}, \overline{\mathbf{c}} + \overline{\mathbf{y}}, \overline{\mathbf{c}} + 2\overline{\mathbf{y}}, \dots, \overline{\mathbf{c}} + \alpha\overline{\mathbf{y}}$ for some nonnegative integer α . Each class with $\alpha + 1$ elements has exactly α solutions to $\overline{\mathbf{c}}_2 - \overline{\mathbf{c}}_1 = \overline{\mathbf{y}}$. It follows¹ that the number of classes is

$$|\mathcal{X}_{h,k}| - \#\{(\overline{\mathbf{c}}_1, \overline{\mathbf{c}}_2) \in \mathcal{X}_{h,k}^2 : \overline{\mathbf{c}}_2 - \overline{\mathbf{c}}_1 = \overline{\mathbf{y}}\}.$$

It now remains to count the number of $\overline{\mathbf{c}}_1, \overline{\mathbf{c}}_2 \in \mathcal{X}_{h,k}$ with $\overline{\mathbf{c}}_2 - \overline{\mathbf{c}}_1 = \overline{\mathbf{y}}$. Let

¹This is identical to the theorem that for a forest $G = (V, E)$, the number of connected components is $|V| - |E|$.

$\bar{\mathbf{y}} = \langle y_1, \dots, y_k \rangle$, and define

$$\begin{aligned} \bar{\mathbf{u}} &= \langle u_1, \dots, u_k \rangle, & u_i &:= \begin{cases} 0, & y_i \leq 0; \\ y_i, & y_i > 0; \end{cases} \\ \bar{\mathbf{v}} &= \langle v_1, \dots, v_k \rangle, & v_i &:= \begin{cases} -y_i, & y_i \leq 0; \\ 0, & y_i > 0. \end{cases} \end{aligned}$$

We have $\|\bar{\mathbf{u}}\|_1 = \|\bar{\mathbf{v}}\|_1 = h_1$, since $\bar{\mathbf{y}} \cdot \bar{\mathbf{1}} = 0$, and $\bar{\mathbf{y}} = \bar{\mathbf{u}} - \bar{\mathbf{v}}$. Every solution to $\bar{\mathbf{c}}_2 - \bar{\mathbf{c}}_1 = \bar{\mathbf{y}}$ corresponds to a solution to $\bar{\mathbf{c}}_2 - \bar{\mathbf{u}} = \bar{\mathbf{c}}_1 - \bar{\mathbf{v}}$, and vice versa. Now consider the i -th coordinate of $\bar{\mathbf{r}} := \bar{\mathbf{c}}_2 - \bar{\mathbf{u}}$. If $y_i \leq 0$, then $u_i = 0$, and so

$$r_i = (\bar{\mathbf{c}}_2 - \bar{\mathbf{u}})_i = (\bar{\mathbf{c}}_2)_i - y_i = (\bar{\mathbf{c}}_2)_i,$$

which is nonnegative since $\bar{\mathbf{c}}_2 \in \mathcal{X}_{h,k}$. On the other hand, if $y_i > 0$, then $v_i = 0$ and

$$r_i = (\bar{\mathbf{c}}_2 - \bar{\mathbf{u}})_i = (\bar{\mathbf{c}}_1 - \bar{\mathbf{v}})_i = (\bar{\mathbf{c}}_1)_i - v_i = (\bar{\mathbf{c}}_1)_i,$$

which is nonnegative since $\bar{\mathbf{c}}_1 \in \mathcal{X}_{h,k}$. Thus, $\bar{\mathbf{c}}_2 - \bar{\mathbf{u}}$ has nonnegative integer coordinates summing $h - h_1$, and so $\bar{\mathbf{r}} \in \mathcal{X}_{h-h_1,k}$.

That is, the solutions to $\bar{\mathbf{c}}_2 - \bar{\mathbf{c}}_1 = \bar{\mathbf{y}}$ are in 1-1 correspondence with the elements of $\mathcal{X}_{h-h_1,k}$. We now have

$$\# \{(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2) \in \mathcal{X}_{h,k}^2 : \bar{\mathbf{c}}_2 - \bar{\mathbf{c}}_1 = \bar{\mathbf{y}}\} = |\mathcal{X}_{h-h_1,k}|$$

and consequently

$$|hA| = (\text{number of } \equiv\text{-classes}) = (\text{number of } \cong\text{-classes}) = |\mathcal{X}_{h,k}| - |\mathcal{X}_{h-h_1,k}|,$$

as claimed. $\square$

2.1. Proof of Lemma 1

Proof of Lemma 1. Consider the set

$$A = \{0, 1, (a-1)(b-1) + 1, (a-1)(b-1) + a\},$$

where $b \geq a \geq 2$. The lattice $\mathcal{L}$ has first minimizer $\bar{\mathbf{y}}_1 = \langle 1-a, a-1, 1, -1 \rangle$ (with norm $2a$) and the second minimizer $\bar{\mathbf{y}}_2 = \langle 0, 1, -b, b-1 \rangle$ (with norm $2b$). The resulting $\mathcal{L}$ has minima $2a, 2b$, as we now prove.

To see that the first and second minimizers are as claimed, we observe that the vector space perpendicular to $\bar{\mathbf{1}}$ and $\bar{\mathbf{a}}$ has dimension 2, and $\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2$ are linearly independent and therefore span the space in question (over the reals). We are then tasked with identifying all real numbers α, β with $\alpha\bar{\mathbf{y}}_1 + \beta\bar{\mathbf{y}}_2 \in \mathbb{Z}^4$. As

$$\alpha\bar{\mathbf{y}}_1 + \beta\bar{\mathbf{y}}_2 = \langle \alpha(1-a), \alpha(a-1) + \beta, \alpha - \beta b, -\alpha + (b-1)\beta \rangle,$$

the first and second components being integers forces $\beta \in \mathbb{Z}$, from which the third component being integral forces $\alpha \in \mathbb{Z}$. Ergo, $\overline{\mathbf{y}_1}, \overline{\mathbf{y}_2}$ are a *lattice* basis for the lattice $\mathcal{L}$.

Thus, the L^1 -norm of a typical member of $\mathcal{L}$ has the form

$$|\alpha(1-a)| + |\alpha(a-1) + \beta| + |\alpha - \beta b| + |\alpha - (b-1)\beta|. \quad (5)$$

For each fixed β , the norm form (5) is a piecewise linear function of α whose minimum happens at one of the values

$$\alpha = 0, \quad \alpha = -\beta/(a-1), \quad \alpha = \beta b, \quad \alpha = (b-1)\beta.$$

Thus, the minimum norm is positive and is one of

$$2b\beta, \quad 2\beta \left(\frac{1}{a-1} + b \right), \quad 2\beta((a-1)b+1), \quad 2\beta((a-1)(b-1)+1),$$

which is at least $2b$ in all cases. If $\beta = 0$, without loss of generality $\alpha > 0$, and the norm form simplifies to $2a\alpha$. This is minimized with $\alpha = 1$, which gives us the first minimum of $2a$ and the first minimizer as $\overline{\mathbf{y}_1}$. The second minimizer must be linearly independent of the first, i.e., $\beta \neq 0$, and from the above computation we now know that the second minimizer is $2b$, achieved by $\overline{\mathbf{y}_2}$.

This proves Lemma 1 for $k = 4$. For larger k , we need only adjoin a sufficiently quickly growing sequence to A . As a specific instance of “sufficiently quick”, it suffices to take $a_{i+1} > (h_2 - 1) \cdot a_i$ for $4 \leq i < k$. $\square$

2.2. Proof of Corollary 2

Proof of Corollary 2. This is a consequence of Theorem 1 and Lemma 1. With $b \geq a \geq 2$, the set

$$A = \{0, 1, (a-1)(b-1)+1, (a-1)(b-1)+a\} \cup \{a_5, \dots, a_k\},$$

where $a_{i+1} = 2b \cdot a_i$ for $4 \leq i < k$, has first minimizer $\langle 1-a, a-1, 1, -1 \rangle$ (with norm $2a$) and second minimizer $\langle 0, 1, -b, b-1 \rangle$ (with norm $2b$). The resulting coefficient lattice $\mathcal{L}$ has minima $2a, 2b$. For a given h, j with $1 \leq j \leq h$, take $a = h-j+1 \geq 2$ and any $b > \max\{h, a+1\}$, and Theorem 1 now says

$$|hA| = \binom{h+k-1}{k-1} - \binom{h-a+k-1}{k-1}. \quad \square$$

3. Types of Tables

3.1. General Notation for “Additive” Combinatorial Number Theory

Let G be a group (for example, $\mathbb{R}$), and let $\varphi : G^h \rightarrow G^i$ be any function (for example, $\sigma_h(\overline{\mathbf{x}}) = x_1 + \dots + x_h$). Let $X \subseteq G$ (for example, $X = \mathbb{N}$, the positive

integers). If $i = 1$, we define

$$\begin{aligned}\varphi X &:= \{\varphi(\bar{\mathbf{g}}) : \bar{\mathbf{g}} \in X^h\} \\ \mathcal{R}_X(\varphi, k) &:= \{|\varphi A| : |A| = k, A \subseteq X\} \\ m_X(\varphi, k) &:= \min\{|A| : A \subseteq X, |A| = k\} \\ M_X(\varphi, k) &:= \max\{|A| : A \subseteq X, |A| = k\}\end{aligned}$$

If $A \subseteq X$ has $|A| = k$ and $|\varphi A| = M_X(\varphi, k)$, then we call A a φ -Sidon set. If $A \subseteq X$ has $|A| = k$ and $|\varphi A| = m_X(\varphi, k)$, then we call A a φ -progression.

When X is clear from the context, it is omitted from the notation. When $\varphi(\bar{\mathbf{a}}) = a_1 + \cdots + a_h$, we abbreviate it in the notation as simply h : the set hA is the h -fold sumset of A and $\mathcal{R}_{\mathbb{N}}(h, k)$ is the set of sizes of h -fold sumsets of sets of k positive integers. In Table 3.1, some well-studied examples are listed.

$\varphi(\bar{\mathbf{a}})$	X	$m_X(\varphi, k)$	$M_X(\varphi, k)$	$\mathcal{R}_X(\varphi, k)$	φ -Sidon	φ -progression
$a_1 + a_2$	$\mathbb{Z}$	$2k - 1$	$\binom{k+1}{2}$	$[m, M]$	Sidon set	arithmetic prog.
$a_1 - a_2$	$\mathbb{Z}$	$2k - 1$	$k^2 - k + 1$	unknown	Sidon set	arithmetic prog.
$a_1 a_2$	$\mathbb{N}$	$2k - 1$	$\binom{k+1}{2}$	$[m, M]$	mult. Sidon set	geometric prog.
$a_1 + a_2 + a_3$	$\mathbb{Z}$	$3k - 2$	$\binom{k+2}{3}$	unknown	B_3 -set	arithmetic prog.
$a_1 + \cdots + a_h$	$\mathbb{Z}$	$hk - h + 1$	$\binom{h+k-1}{k-1}$	unknown	B_h -set	arithmetic prog.
$a_1 + 2a_2$	$\mathbb{Z}$	$3k - 2$	k^2	unknown	unnamed	???

Table 2: A few functions φ and some known values of m_X , M_X , $\mathcal{R}_X$, and extremal sets.

There are interesting questions with $i \geq 2$. For example, with $\varphi(x_1, x_2) = (x_1 + x_2, x_1 - x_2)$, one is led to questions about the relative sizes of sumsets and difference sets, for example [3]. The work [9] concerns $\varphi(x_1, x_2) = (u_1 x_1 + u_2 x_2, v_1 x_1 + v_2 x_2)$ in full generality. The function $\varphi(x_1, x_2) = (x_1 + x_2, x_1 x_2)$ leads one to the infamous Sum-Product Conjecture [2, 10].

3.2. Types of Tables

If G is an orderable group and $\varphi : G^h \rightarrow G$, we can enumerate the elements of A as $\{a_1 < \cdots < a_k\}$, and define the φ -table as the function T from $[k]^h \rightarrow G$ with

$$T(\bar{\mathbf{x}}) = \varphi(a_{x_1}, a_{x_2}, \dots, a_{x_h}).$$

For example, with $\varphi(\bar{\mathbf{x}}) = x_1 + x_2$, the table is $T(i, j) = a_i + a_j$.

We define the (weak) *type* of the φ -table to be the equivalence relation on $[k]^h$ induced by φ , i.e., we have $\bar{\mathbf{x}} \stackrel{\varphi, A}{\equiv} \bar{\mathbf{y}}$ if and only if $T(\bar{\mathbf{x}}) = T(\bar{\mathbf{y}})$.

We set $\mathcal{T}_X(\varphi, k)$ to be the set of all types of φ -tables for $A \subseteq X$ with $|A| = k$.

We note that $|\varphi A|$ is the number of equivalence classes of the type of the φ -table, so that if A and B have the same type of φ -table, then $|\varphi A| = |\varphi B|$.

We comment that “type” as we have defined it here is somewhat weaker than the closely related “type” in [10], and enumerated in [A378609](#). There, “type” includes an ordering of the equivalence classes. The version in this work is easier to work with theoretically, while the additional information in the ordering is computationally useful.

Question 1. How many types of addition table are there? That is, what is $|\mathcal{T}_{\mathbb{Z}}(h, k)|$?

4. An Application of Types

Nathanson [4] asked if $\mathcal{R}_{\mathbb{N}}(h, k) = \mathcal{R}_{\mathbb{R}}(h, k)$. We provide the stronger equality of types.

Theorem 3. *Let h, k be positive integers. Then, $\mathcal{T}_{\mathbb{N}}(h, k) = \mathcal{T}_{\mathbb{R}}(h, k)$. In particular, $\mathcal{R}_{\mathbb{N}}(h, k) = \mathcal{R}_{\mathbb{R}}(h, k)$*

Definition 1. For $\varphi : \mathbb{R}^h \rightarrow \mathbb{R}$, the *separation* of a set A is

$$\text{sep}_{\varphi}(X) := \min \left\{ |\phi(\bar{\mathbf{a}}) - \phi(\bar{\mathbf{b}})| : \bar{\mathbf{a}}, \bar{\mathbf{b}} \in X^h, \varphi(\bar{\mathbf{a}}) \neq \varphi(\bar{\mathbf{b}}) \right\}.$$

If $\varphi(\bar{\mathbf{x}}) = \sigma_h(\bar{\mathbf{x}}) = x_1 + \cdots + x_h$, we write “ h ” in place of “ φ ” in the notation, and note that $\text{sep}_1(X) \geq \text{sep}_2(X) \geq \cdots$.

Proof of Theorem 3. For $h = 1$ this is trivial, so we take $h \geq 2$. Let $X \subseteq \mathbb{R}$ with $|X| = k$. We build a set $A \subseteq \mathbb{N}_0$ with $|A| = k$ and A, X having the same type of h -fold addition table. By affine invariance, we may assume that

$$X = \{0 = x_1 < \cdots < x_k = 1\}.$$

Let $q_0 \in \mathbb{N}$ be sufficiently large that $\text{sep}_h(X) \cdot q_0 \geq 2$.

Consider the sequence of vectors $d_q = \langle \{qq_0x_2\}, \dots, \{qq_0x_{k-1}\} \rangle$ of fractional parts of dilations of $x_2, \dots, x_{k-1}$. Our pigeons are d_q for $0 \leq q \leq (k-2)^{2h}$. Our pigeonholes are the $(k-2)$ -fold cartesian products of $[\frac{i}{2h}, \frac{i+1}{2h})$ with $0 \leq i < 2h$. We have $(k-2)^{2h} + 1$ pigeons, and $(k-2)^{2h}$ pigeonholes, so there are q', q'' with

$$0 \leq q' < q'' \leq (k-2)^{2h},$$

and $d_{q'}, d_{q''}$ in the same pigeonhole. That is, with $Q := q''q_0 - q'q_0 \in [q_0, q_0(k-2)^{2h}]$ and a_i the integer nearest Qx_i , we have

$$-\frac{1}{2h} < Qx_i - a_i < \frac{1}{2h}, \quad q_0 \leq Q \leq q_0(k-2)^{2h}$$

simultaneously for $1 \leq i \leq k$. Set $\epsilon_i \in (-1/(2h), 1/(2h))$ so that $Qx_i = a_i + \epsilon_i$. Notice $\epsilon_1 = \epsilon_k = 0$.

We claim that $A = \{a_1, \dots, a_k\}$ has $|A| = |X|$ and A, X have the same type of addition table. Since $x_1 < \dots < x_k$ and $Q \geq q_0 \geq 1$, we clearly have $a_1 \leq \dots \leq a_k$. More precisely, we have

$$a_{i+1} - a_i = Q(x_{i+1} - x_i) + (\epsilon_i - \epsilon_{i+1}) \geq \text{sep}_1(X)Q - \frac{1}{h} \geq \text{sep}_h(X)q_0 - \frac{1}{2} > 1$$

so that $a_{i+1} > a_i$. In particular, $|A| = |X|$.

We will argue that the h -type of X is the same as the h -type of A , whence $|hX| = |hA|$. First, assume that $(b_1, \dots, b_h) \stackrel{h,X}{\equiv} (c_1, \dots, c_h)$, so that

$$x_{b_1} + \dots + x_{b_h} = x_{c_1} + \dots + x_{c_h}.$$

One therefore has

$$a_{b_1} + \dots + a_{b_h} = a_{c_1} + \dots + a_{c_h} + (\epsilon_{c_1} + \dots + \epsilon_{c_h}) - (\epsilon_{b_1} + \dots + \epsilon_{b_h}).$$

From the definition of ϵ_i , we have

$$-1 < (\epsilon_{c_1} + \dots + \epsilon_{c_h}) - (\epsilon_{b_1} + \dots + \epsilon_{b_h}) < 1$$

and by the integrality of the a_i , this must be 0. That is,

$$a_{b_1} + \dots + a_{b_h} = a_{c_1} + \dots + a_{c_h}$$

and consequently $(b_1, \dots, b_h) \stackrel{h,A}{\equiv} (c_1, \dots, c_h)$.

Now, assume that $(b_1, \dots, b_h) \stackrel{h,A}{\equiv} (c_1, \dots, c_h)$, from which we have

$$a_{b_1} + \dots + a_{b_h} = a_{c_1} + \dots + a_{c_h}.$$

As $a_i = Qx_i - \epsilon_i$, we have

$$x_{b_1} + \dots + x_{b_h} = x_{c_1} + \dots + x_{c_h} - \left(\frac{\epsilon_{b_1} + \dots + \epsilon_{b_h} - \epsilon_{c_1} - \dots - \epsilon_{c_h}}{Q} \right).$$

As

$$\left| \frac{\epsilon_{b_1} + \dots + \epsilon_{b_h} - \epsilon_{c_1} - \dots - \epsilon_{c_h}}{Q} \right| < \frac{1}{Q} \leq \frac{\text{sep}_h(X)}{2},$$

it must be that

$$x_{b_1} + \dots + x_{b_h} = x_{c_1} + \dots + x_{c_h}.$$

Thus, $(b_1, \dots, b_h) \stackrel{h,X}{\equiv} (c_1, \dots, c_h)$. □

Theorem 4. Let h be positive integer, and let $\sigma(\bar{x}) = \sum_{i=1}^h x_i$ and $\pi(\bar{x}) = \prod_{i=1}^h x_i$. For every positive integer k

$$\mathcal{T}_{\mathbb{N}}(\sigma, k) = \mathcal{T}_{\mathbb{N}}(\pi, k).$$

In particular, $\mathcal{R}_{\mathbb{N}}(\sigma, k) = \mathcal{R}_{\mathbb{N}}(\pi, k)$.

Proof. Suppose that S is a k -element set of positive integers, and let $P := \{2^s : s \in S\}$. Then $a_1 + \cdots + a_h = b_1 + \cdots + b_h$ if and only if $2^{a_1} \cdots 2^{a_h} = 2^{b_1} \cdots 2^{b_h}$. Thus every type in $\mathcal{T}_{\mathbb{N}}(\sigma, k)$ is also in $\mathcal{T}_{\mathbb{N}}(\pi, k)$.

Now, suppose that P is a set of positive integers with $|P| = k$. We must exhibit a set A of positive integers whose σ -type is the same as the π -type of P .

Clearly, the π -type of P is the same as the σ -type of $L_1 := \{\log_2 p : p \in P\}$, but L is not typically a set of integers. For $c > 0$, we set $L_c := \{c \log_2 p : p \in P\}$, which has the same σ -type as L_1 and π -type as P . We will show that it is possible to take c so that $A_c := \{c + \lfloor c \log_2 p \rfloor : p \in P\} \subseteq \mathbb{N}$, where $\lfloor x \rfloor$ is the integer nearest to x , has the same σ -type as L_c .

If c is sufficiently large, then clearly A_c will be a set of $|P|$ distinct positive integers. As in the above argument, we can take c both arbitrarily large and so that $\{c + \lfloor c \log_2 p \rfloor : p \in P\}$ has arbitrarily large separation, and so that each $c + \lfloor c \log p \rfloor$ is within $1/(2h)$ of an integer. The proof goes through as above. $\square$

5. Experimental Thoughts

5.1. Nathanson's Experiment

Consider the following experiment conducted by Nathanson [8]. We start by generating 10^7 random subsets of $\{1, 2, \dots, 1000\}$ of size 4. We have $|10A|$ taking 213 distinct values between 31 and 286, inclusive. Roughly 78% of them are B_{10} -sets, i.e., $|hA| = \binom{10+4-1}{4-1} = 286$. Of the 2 202 261 other sets in our sample, we plot the proportion that have each possible size $|hA|$ in Figure 1

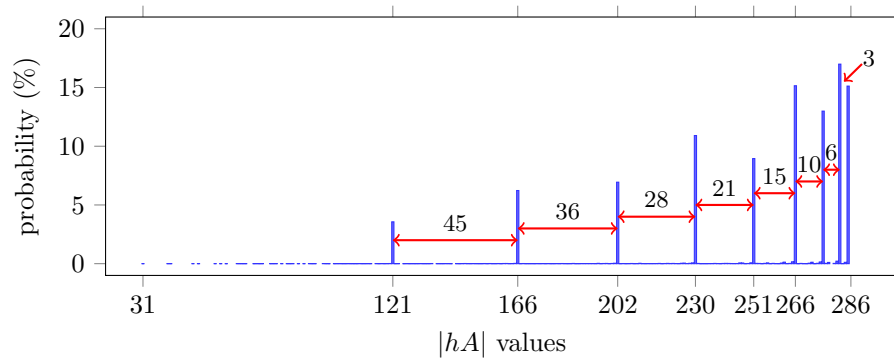


Figure 1: The experimental conditional probability of $|10A|$, where $A \subseteq \{1, 2, \dots, 1000\}$ uniformly, conditioned on $|10A| \neq 286$.

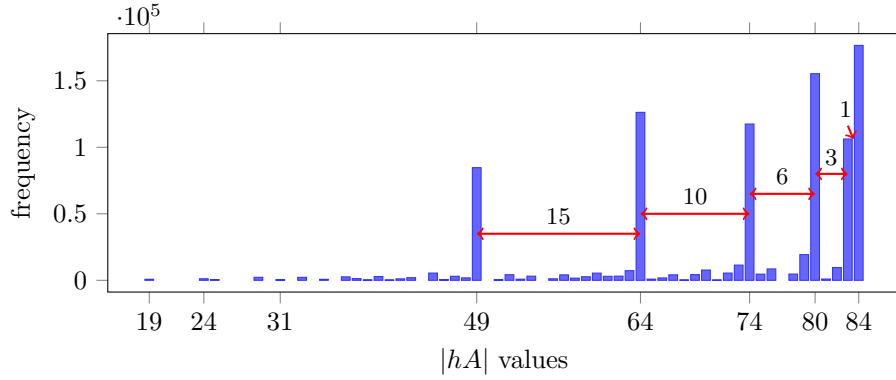


Figure 2: The bar chart of the 6-fold sumset sizes of all 4-element subsets $A \subseteq [70] := \{1, 2, \dots, 70\}$. As Nathanson observed, the distribution is strongly concentrated on sizes of the form $\binom{h+3}{3} - \binom{t+2}{3}$ for $0 \leq t < h$.

We find that 97% of the 2 202 261 non- B_{10} -sets² have $|hA|$ among the 9 sizes

$$285, 282, 276, 266, 251, 230, 202, 166, 121.$$

The first differences of the list of common values of $|10A|$ that are less than 286 are

$$45, 36, 28, 21, 15, 10, 6, 3,$$

i.e., the triangular numbers. It is this mystery — the appearance of triangular numbers — to which we have given a partial explanation in this work.

We note that the partial sums of the triangular numbers are the tetrahedral numbers given by $\binom{t+2}{3}$. Nathanson's experiment indicates that a random 4-element subset A of $[n]$ is highly likely to have $|hA| = \binom{h+3}{3}$, but if it does not then it is highly likely to have $|hA|$ among $\binom{h+3}{3} - \binom{t+2}{3}$, with $1 \leq t < h$.

5.2. Distribution of Minima

This author believes that the appearance of triangular and tetrahedral numbers in Nathanson's experiment is a consequence of Theorem 1 and the as-yet-unproven reality that roughly 80% of $\binom{[1000]}{4}$ have $h_1 > 6$, and of the remaining 20%, almost all have $h_2 > 6$. More generally, the following conjecture combined with Theorem 1 would be a satisfactory explanation.

²It is expected that most of our sample are B_{10} -sets: with fixed h, k , if n is sufficiently large then almost all sets in $\binom{[n]}{k}$ are B_h -sets.

Conjecture 1. Fix $h \geq 3$, and let $\epsilon > 0$. There is an n_0 such that for all $n \geq n_0$, at least $(1 - \epsilon)\binom{n}{4}$ of the subsets of $\binom{[n]}{4}$ have $h_1 > h$, and of the sets with $h_1 \leq h$, the proportion with $h_2 \leq h$ is at least $1 - \epsilon$.

This is really a question about a “random coefficient lattice”, the lattice perpendicular to $\bar{\mathbf{1}}$ and a random $\bar{\mathbf{a}}$. Experiments indicate that if A is chosen uniformly from $\binom{[n]}{4}$, then the expected value of $h_1(A)$ is close to $\frac{1}{2}n^{1/2}$, with standard deviation around $\frac{1}{5}n^{1/2}$. If A is chosen uniformly from $\binom{[n]}{5}$, the expected value seems to approach $\frac{5}{8}n^{1/3}$, with the standard deviation in the neighborhood of $\frac{1}{5}n^{1/3}$.

Another concrete example is instructive. Let $A = \{1, 5, 96, 100\}$. The set A is not a B_2 -set as $1 + 100 = 5 + 96$. In lattice terms, $h_1 = 2$ and $\bar{\mathbf{y}}_1 = \langle 1, -1, -1, 1 \rangle$. However, A does have

$$|hA| = \binom{h+3}{3} - \binom{h+1}{3}$$

for $2 \leq h \leq 94$. For $h = 95$, though, we encounter the new equality:

$$95 \cdot 96 = 91 \cdot 100 + 4 \cdot 5.$$

In lattice terms, $h_2 = 95$ and $y_2 = \langle 0, -4, 95, -91 \rangle$.

This experimental work would benefit from a positive answer to the question:

Question 2. Is there a way to generate k -element subsets of $[n]$ that are not B_h -sets uniformly that is more efficient than simply generating k -element subsets of $[n]$ and testing for the B_h -property?

5.3. Nathanson’s Program

In recent works, Nathanson [4, 5, 6, 7, 8] has given a program to develop a deeper understanding of $|hA|$, beyond merely the extreme values. The current work fits within his framework.

Remark 1 (A gap in the literature.). In [4], Nathanson writes “it is not known if $\mathcal{R}_{\mathbb{N}_0}(h, k) = \mathcal{R}_{\mathbb{Z}}(h, k)$.” Our Theorem 2 fills this gap in the literature.

Remark 2 (A Problem from Nathanson.). In [8], Problem 8 is to prove that

$$\left\{ \binom{h+3}{3} - \binom{j+2}{3} : 1 \leq j \leq h \right\} \subseteq \mathcal{R}_{\mathbb{Z}}(h, 4).$$

This is a consequence of Theorem 1 with the set

$$A = \{0, 1, (a-1)(b-1) + 1, (a-1)(b-1) + a\},$$

where $b \geq a \geq 2$. The first minimizer is $\langle 1-a, a-1, 1, -1 \rangle$ (with norm $2a$) and the second minimizer is $\langle 0, 1, -b, b-1 \rangle$ (with norm $2b$). The resulting $\mathcal{L}$ has minima

2a, 2b. For a given h, j with $1 \leq j \leq h$, choose any $b > h$ and $a = h - j + 1$ and Theorem 1 now says

$$|hA| = \binom{h+3}{3} - \binom{h-a+3}{3} = \binom{h+3}{3} - \binom{j+2}{3}.$$

Remark 3 (Another Problem from Nathanson.). In [8], Problem 9 is to “Discover the pattern of popular sumset sizes for sets of size $k \geq 5$.” The current work suggests that the popular values will be

$$\left\{ \binom{h+k-1}{k-1} - \binom{j+k-1}{k-1} : 0 \leq j < h \right\},$$

provided that n is sufficiently large in terms of h, k . A modest amount of experimentation agrees with this suggestion.

Acknowledgements. This work was inspired by Melvyn B. Nathanson’s many articles on the sequence of sumset sizes, a few of the most recent of which are cited in the bibliography. Moreover, the prose has benefited from his questions and suggestions.

References

- [1] M. B. Nathanson, Sums of finite sets of integers, *Amer. Math. Monthly* **79** (1972), 1010–1012.
- [2] M. B. Nathanson, On sums and products of integers, *Proc. Amer. Math. Soc.* **125**(1) (1997), 9–16.
- [3] M. B. Nathanson, Sets with more sums than differences, *Integers* **7**, (2007) #A5, 24 pp.
- [4] M. B. Nathanson, Compression and complexity for sumset sizes in additive number theory, *J. Number Theory* **281** (2026), 321–343.
- [5] M. B. Nathanson, Explicit sumset sizes in additive number theory, *Canad. Math. Bull.* (2025), 1–12.
- [6] M. B. Nathanson, Inverse problems for sumset sizes of finite sets of integers, preprint, [arXiv:2412.16154](#).
- [7] M. B. Nathanson, Problems in additive number theory, VI: sizes of sumsets, *Acta Math. Hungar.* **176** (2025), 498–521.
- [8] M. B. Nathanson, Triangular and tetrahedral number differences of sumset sizes in additive number theory, preprint, [arXiv:2506.15015](#).
- [9] M. B. Nathanson, K. O’Bryant, B. Orosz, I. Ruzsa, and M. Silva, Binary linear forms over finite sets of integers, *Acta Arith.*, **129**(4) (2007), 341–361.
- [10] K. O’Bryant, Visualizing the sum-product conjecture, *Combinatorial and additive number theory, VIII*, to appear.