On Sumset Inequalities: A mathematical approach

Original research paper

Received:
Accepted:
Online Ready:

Abstract

In this paper, we study sumset inequalities concerning cardinalities of sumsets and functions of finite subsets of integers such as entropy and additive energy. We also identify sufficient conditions for existence of arithmetic progressions in difference sets and three-term progressions in differences of squares of prime numbers.

Keywords: Additive combinatorics; Sumsets; Entropy; Additive energy
2010 Mathematics Subject Classification: 11B13; 11B25; 11B30

1 Introduction

When studying sumsets of subsets of the natural numbers, one has a myriad number of ways to define functions of the sets involved. For given sets $X, Y, Z \subseteq \mathbb{N}$, it is then possible to construct relations and inequalities which include different functions of the sets (Ruzsa [2009]). Examples of such functions include cardinality, entropy (Ruzsa [2008]) and additive energy (Balog and Szemerédi [1994], Gowers [1998]) and the extent to which a natural number is represented as a sum of two elements from the set. We prove some theorems, by elementary arguments, which exhibit certain inequalities. One such theorem yields an inequality with additive energy, by means of concentration of probability measures. One notes also the connection with ideas in probability and statistics such as entropy, concentration, frequency and density. We then prove theorems about existence of arithmetic progressions in difference set. When dealing with the set of prime numbers or squares of primes, by counting the number of primes less than a number, one may show existence arbitrarily long and three-term progressions in difference sets. In the context of difference sets generated by squares of primes, we need a hypothesis which requires the non-existence of two distinct increasing triples of prime squares, which are not translations of each other. Prior research in this topic may be found in Ruzsa [1978a], Ruzsa [1989], Ruzsa [1996], Ruzsa [2008], Tao [2010], Green et al. [2025], Gowers et al. [2025], Balog and Szemerédi [1994], Gowers [1998], Bourgain [1999], Ruzsa [1978b], Ruzsa [1994], Chang [2002], Green [2005], Green and Tao [2008], Basu [2024], Tao and Vu [2016] and Tao and Vu [2006].

2 On certain sumset inequalities

In this section, we will consider some inequalities involving sumsets and relations between functions of the sets such as set cardinality, additive energy etc., within additive combinatorics.

1) We consider a stronger conjecture than Ruzsa's entropic conjecture for sumsets (Ruzsa [2008]) and identify certain sumset inequalities when specific sufficient conditions are met, such as arithmetic progressions. The entropic counterpart involves the joint entropy of the pair of sums of random variables.

2) When considering the sum with the largest number of combinations possible, we derive an inequality concerning this object and sumsets by applying a concentration theorem by Markov's inequality. The probabilities in the concentration will lead to sumset inequalities.

3) We will then consider the problem of existence of arbitrarily long arithmetic progressions in the difference set $X - X$, when X is dense. Particularly, we will be interested in situations when X is the set of primes and three-term progressions when X is the set of squares of primes.

The following notation is for asymptotic order of growth. For a subset of the Euclidean space $\mathcal{X} \subseteq \mathbb{R}^{d_1}$ and functions $f, g : \mathcal{X} \to \mathbb{R}^{d_2}$ we say $f = O(g)$ if there exists a constant $C > 0$ such that $||f(x)|| \leq C||g(x)||$ for each $x \in \mathcal{X}$. We say that $f = \Omega(g)$ if there exists an unbounded subset $\mathcal{X}' \subseteq \mathcal{X}$ and a constant $C > 0$ such that $||f(x)|| \geq C||g(x)||$ for each $x \in \mathcal{X}'$.

We will apply the same notation for X, Y, Z when X, Y, Z are finite sets or random variables. Suppose that X, Y, Z are $\mathbb{N}$-valued random variables which are finitely supported. The definition of entropy is as follows.

$$H(X) = \sum_x p_x \ln(\frac{1}{p_x}) \tag{2.1}$$

and one denotes by p_x, the probability that the outcome x will be realised. The summation is over x such that $p_x > 0$. Then, a conjecture (Ruzsa [2008]) involving entropy, when X, Y, Z are independent, is

$$H(X + Y) + H(Z) \leq H(X + Z) + H(Y + Z). \tag{2.2}$$

A stronger version of this conjecture would be the following, when X, Y, Z are independent random variables which satisfy certain sufficient conditions.

$$H(X + Y) + H(Z) \leq H(X + Z, Y + Z). \tag{2.3}$$

In the above inequalities, H is the entropy function defined on random variables. This would be an expectation of the logarithm of the probabilities. In this situation that X, Y, Z are finite subsets of the natural numbers $\mathbb{N}$, we have the corresponding sumset inequality involving cardinalities which is of the following form.

$$|X + Y||Z| \leq |\mathcal{E}_{X,Y,Z}|. \tag{2.4}$$

The set $\mathcal{E}_{X,Y,Z}$ is defined as

$$\mathcal{E}_{X,Y,Z} = \{(x + z, y + z) : x \in X, y \in Y, z \in Z\} \tag{2.5}$$

Of course, we note that it is known (Ruzsa [1978a], Ruzsa [1989],Ruzsa [1996]) that

$$|X + Y||Z| \leq |X + Z||Y + Z|. \tag{2.6}$$

Hence, any counterexamples to the entropic conjecture must involve non-uniform probabilities. This leads us to understanding whether the same is true for the stronger conjecture.

We prove the following theorems.

Proposition 2.1. *Suppose that $c > 0$. Further, suppose that $X, Y, Z \subseteq \mathbb{N}$ are finite sets such that*

$$c|Z| \leq \min\{|X|, |Y|\} \tag{2.7}$$

and

$$|X + Y| \leq c \min |X|, |Y|. \tag{2.8}$$

Then,

$$|X + Y||Z| \leq |\mathcal{E}_{X,Y,Z}|. \tag{2.9}$$

Proof. Note that

$$\begin{aligned} c|X + Y||Z| &\leq c(\min |X|, |Y|)^2 && (2.10)\\ &\leq c|X||Y| && (2.11)\\ &\leq |\mathcal{E}_{X,Y,Z}|. && (2.12) \end{aligned}$$

□

The next theorem is regarding arithmetic progressions.

Proposition 2.2. *Suppose that X and Y finite subsets of $\mathbb{N}$, which are arithmetic progressions with the same difference term i.e. $\min\{|x - x'| : x, x' \in X; x \neq x'\} = \min\{|y - y'| : y, y' \in Y; y \neq y'\}$. Then,*

$$|X + Y||Z| \leq |\mathcal{E}_{X,Y,Z}|. \tag{2.13}$$

Proof. Since X, Y are arithmetic progressions with the same difference term,

$$|X + Y| = |X| + |Y| - 1. \tag{2.14}$$

Note that $\mathcal{E}_{X,Y,Z} = \bigcup_{z\in Z}(X \times Y + \{(z, z)\})$. We define the set

$$F = \{(x, y) \in X \times Y : x = \max x'_{x'\in X} \text{ or } y = \max y'_{y'\in Y}\} \tag{2.15}$$

which corresponds to the frontier of the set $X \times Y$ in $\mathbb{N}^2$. Note that $|F| = |X| + |Y| - 1$. Hence,

$$\begin{aligned} |X + Y||Z| &= |Z|(|X| + |Y| - 1). && (2.16)\\ &= |\bigcup_{z\in Z} F + \{(z, z)\}| && (2.17)\\ &\leq |\mathcal{E}_{X,Y,Z}|. && (2.18) \end{aligned}$$

□

The following theorem is a result concerning finite subsets with wide consecutive differences in X and Y relative to Z.

Proposition 2.3. *Suppose also that X and Y are finite subsets of the natural numbers. Further, suppose that $\min\{|x - x'| : x, x' \in X; x \neq x'\} > \max_{z\in Z} z$ and $\min\{|y - y'| : y, y' \in Y; y \neq y'\} > \max_{z\in Z} z$. Then,*

$$|X + Y||Z| \leq |\mathcal{E}_{X,Y,Z}|. \tag{2.19}$$

Proof. Note that given the condition is satisfied, for $(x,y) \neq (x',y')$, we have that $(\{(x,y)\}+Z) \cap (\{(x',y')\}+Z) \neq \emptyset$. Since $|X+Y| \leq |X||Y|$, the following equality proves the result.

$$\begin{aligned} |\mathcal{E}_{X,Y,Z}| &= |\bigcup_{(x,y)\in X\times Y} (\{(x,y)\}+Z)| && (2.20)\\ &= \sum_{(x,y)\in X\times Y} |\{(x,y)\}+Z| && (2.21)\\ &= |X||Y||Z|. && (2.22) \end{aligned}$$

□

We now define some functions of the set X such as additive energy and the sum which corresponds to the maximum number of additive quadruples, which are vectors (x_1,x_2,x_3,x_4) with the property that $x_1+x_2=x_3+x_4$ (Tao and Vu [2006]).

Additive energy E_X, is defined as follows.

$$E_X = \frac{|\{(x_1,x_2,x_3,x_4) \in X^4 : x_1+x_2=x_3+x_4\}|}{|X|^3}. \tag{2.23}$$

Amongst all sums that are generated by additive quadruples, we find the one which is generated the most number of times. For each $m \in X+X$, we define the function

$$n_m = |\{(x_1,x_2,x_3,x_4) \in X^4 : x_1+x_2=x_3+x_4=m\}| \tag{2.24}$$

and the maximum over m which is

$$n_X = \max_{m\in X+X} n_m. \tag{2.25}$$

Then, we define the following set. For a given $\varepsilon > 0$,

$$\mathcal{M}_\varepsilon = \{m : n_m \geq \varepsilon n_X\}. \tag{2.26}$$

We prove the following theorem.

Proposition 2.4. *Suppose that $X \subseteq \mathbb{N}$ is a finite subset of natural numbers and $0<c<1$. If $E_X \geq c$, then*

$$(1-\sqrt{1-\frac{c}{n_X}})|X+X| \leq |\mathcal{M}_{\sqrt{1-\sqrt{1-\frac{c}{n_X}}}}|. \tag{2.27}$$

Proof. The additive energy may be written as

$$E_X = \frac{\sum_{m\in X+X} n_m^2}{|X|^3}. \tag{2.28}$$

Now, we define the random variable $\varphi(m) = \frac{n_m^2}{n_X^2}$ on $X+X$, with the uniform probability measure on $X+X$. Denote as $\mathbb{P}$ and $\mathbb{E}$ the associated probability and expectation operators. Hence, if $E_X \geq c$, then expectation of φ may be lower bounded as $\mathbb{E}[\varphi] \geq \frac{c}{n_X}$, since $|X+X| \leq |X|^2$ and $n_X \leq |X|$. By Markov's inequality, we may show that

$$\mathbb{P}(\varphi \geq 1-\sqrt{1-\frac{c}{n_X}}) \geq 1-\sqrt{1-\frac{c}{n_X}}. \tag{2.29}$$

Then, the result is proved by the fact that the probability measure was the uniform probability measure. □

Note the connection also with the Balog-Szemerédi-Gowers theorems (Balog and Szemerédi [1994], Gowers [1998], Bourgain [1999], Gowers et al. [2025]).

We now move to some properties of the difference set $X - X$, for possibly infinite $X \subseteq \mathbb{N}$ in the following context. Denote by $\mathbb{P}$, the set of all prime numbers. Define

$$Q_n = X \cap \{1, ..., n\} \text{ and } Q'_n = (X - X) \cap \{1, ..., n\} \tag{2.30}$$

We will be interested in elementary arguments which allows us to reason about $n_{Q_n}, n_{Q'_n}$ (Basu [2024]) and stronger properties involving existence of arithmetic progressions in $X-X$ (Basu [2024],Ruzsa [1994], Chang [2002], Bourgain [1990], Szemerédi [1975], Gowers [2001], Tao and Vu [2006]). We prove the following theorems. A sequence $(x_1, x_2, ..., x_k)$ is said to be a *translation* of a sequence $(x'_1, x'_2, ..., x'_k)$ if there exists an integer d such that $x'_j = x_j + d$ for all $1 \leq j \leq k$.

Proposition 2.5. *Suppose that $X \subseteq \mathbb{N}$ and $\delta > 0$ such that*

$$\limsup_{n \to \infty} \frac{|Q_n|}{n} \geq \delta > 0. \tag{2.31}$$

Then, $X - X$ contains arbitrarily long arithmetic progressions. Moreover, if $|Q_n| \geq \delta n$ and $\delta = \Omega(\frac{\sqrt{\ln(n)}}{e^{\sqrt{\ln(n)}}})$, then Q'_n contains an arithmetic progression of length $\Omega(\sqrt{\ln(n)})$.

Proof. We shall denote as $k \in \mathbb{N}$, the length of the arithmetic progression. Define $q_n = |Q_n|$. The cardinality of the largest set of increasing sequences $(x_1, x_2, ..., x_k) \in Q_n^k$ (i.e. $x_j < x_{j+1}$) such that no two distinct sequences in the set are translations of each other, is at least $n^{-1}\binom{q_n}{k}$. Now, select n large enough such that

$$n^{-1}\binom{q_n}{k} > (8n+1)^{k-2}. \tag{2.32}$$

Note that the multiplier n^{-1} is present since the number of translations of a sequence in $\{1, ..., n\}$ is not more than n. Now, by the pigeon-hole principle, we find two distinct increasing sequences $(x_1, x_2, ...x_k)$ and $(x'_1, x'_2, ...x'_k)$ such that

$$x_{j-1} - 2x_j + x_{j+1} = x'_{j-1} - 2x'_j + x'_{j+1}, \text{ for all } j \in \{2, ..., k-1\}. \tag{2.33}$$

Then, the difference $(x_1 - x'_1, x_2 - x'_2, ...x_k - x'_k)$ would yield a non-trivial arithmetic progression of length k since the two sequences are not translations of each other.

Further, note that since $k^k \geq k!$, the above strict inequality is implied by

$$(q_n - k) > k(8n+1)^{\frac{k-1}{k}}. \tag{2.34}$$

Hence, by setting $k = \Omega(\sqrt{\ln(n)})$, we prove the second part of the theorem. □

The following theorems now consider the set of prime numbers and the set of squares of prime numbers (Green [2005], Green and Tao [2008], Tao and Vu [2006]). Define $P_n = \mathbb{P} \cap \{1, 2, ..., n\}$ to be the set of prime numbers less than or equal n. The arguments in these proofs also allows us to reason about n_{P_n} and $n_{P'_n}$

Proposition 2.6. *Suppose that $X = \mathbb{P}$. Then, $X - X$ contains arbitrarily long arithmetic progressions.*

Proof. Denote as $[\alpha]$, the largest natural number less than $\alpha > 0$. The number of the increasing sequences $(x_1, x_2, ..., x_k)$ i.e. $x_j < x_{j+1}$ in the set P_n is at least $\Omega(n^{-1}\binom{[\frac{n}{\ln(n)}]}{k})$, by Chebychev's theorem, which would be greater than $(8n+1)^{k-2}$ for large n, and which is an upper bound on the number of combinations of values of the vector $((x_{j-1} - 2x_j + x_{j+1}))_{j=2}^{k-1}$. Then, by the pigeon-hole principle there exist distinct $(x_1, x_2, ..., x_k), (x'_1, x'_2, ..., x'_k) \in P_n^k$ such that the difference $(x_1 - x'_1, x_2 - x'_2, ...x_k - x'_k)$ would again be an arithmetic progressions of length k. □

Proposition 2.7. *Suppose that $X = \{p^2 : p \in \mathbb{P}\}$. Suppose that no two distinct increasing triples in X are translations of each other. Then, $X - X$ contains an arithmetic progression of length three.*

Proof. The number of increasing triples $(x, y, z) \in X \cap \{1, 2, ..., n\}$ is at least $\Omega(\binom{[\frac{\sqrt{n}}{\ln(n)}]}{3})$, which would be greater than $8n$, for large n. By the same argument as the prior two propositions, we find an arithmetic progression of length three in $X - X$. □

3 Conclusion

In this paper, it was shown that sumset inequalities arise from inequalities from probabilistic concentration and also elementary combinatorial identities which are in the form of inequalities. It becomes possible to reason about certain functions of additive sets and identify specific statements that would hold within the theory, with subsets of natural numbers.

References

Antal Balog and Endre Szemerédi. A statistical theorem of set addition. *Combinatorica*, 14(3):263–268, 1994.

Pathikrit Basu. On the number of combinations generating a sum. *Asian Research Journal of Mathematics*, 20(9):98–102, August 2024.

Jean Bourgain. On arithmetic progressions in sums of sets of integers. *A tribute to Paul Erdos*, pages 105–109, 1990.

Jean Bourgain. On the dimension of kakeya sets and related maximal inequalities. *Geometric & Functional Analysis GAFA*, 9(2):256–282, 1999.

Mei-Chu Chang. A polynomial bound in freiman's theorem. *Duke Math. J.*, 115(1):399–419, 2002.

William T Gowers. A new proof of szemerédi's theorem. *Geometric & Functional Analysis GAFA*, 11 (3):465–588, 2001.

William Timothy Gowers. A new proof of szemerédi's theorem for arithmetic progressions of length four. *Geometric & Functional Analysis GAFA*, 8(3):529–551, 1998.

William Timothy Gowers, Ben Green, Freddie Manners, and Terence Tao. On a conjecture of marton. *Annals of Mathematics*, 201(2):515–549, 2025.

Ben Green. Roth's theorem in the primes. *Annals of mathematics*, pages 1609–1636, 2005.

Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of mathematics*, pages 481–547, 2008.

Ben Green, Freddie Manners, and Terence Tao. Sumsets and entropy revisited. *Random Structures & Algorithms*, 66(1):e21252, 2025.

Imre Z Ruzsa. On the cardinality of a+a and a-a. *Combinatorics Proc. Fifth Hungarian Colloq. (Keszthely 1976), Colloq. Math. Soc. J. Bolyai, North-Holland, Amsterdam*, 18(5):933–938, 1978a.

Imre Z Ruzsa. On difference sets. *Studia Sci. Math. Hungar*, 13(3-4):319–326, 1978b.

Imre Z Ruzsa. An application of graph theory to additive number theory. *Scientia, Ser. A*, 3(97-109): 9, 1989.

Imre Z Ruzsa. Generalized arithmetical progressions and sumsets. *Acta Mathematica Hungarica*, 65 (4):379–388, 1994.

Imre Z Ruzsa. Sums of finite sets. In *Number Theory: New York Seminar 1991–1995*, pages 281–293. Springer, 1996.

Imre Z. Ruzsa. Sumsets and entropy. *Random Structures &; Algorithms*, 34:1–10, November 2008.

Imre Z Ruzsa. Sumsets and structure. *Combinatorial number theory and additive group theory*, pages 87–210, 2009.

Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith*, 27(199-245):2, 1975.

Terence Tao. Sumset and inverse sumset theory for shannon entropy. *Combinatorics, Probability and Computing*, 19(4):603–639, 2010.

Terence Tao and Van Vu. Sum-avoiding sets in groups. *Discrete Analysis*, 2016.

Terence Tao and Van H Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.

©2025 Pathikrit Basu; This is an Open Access article distributed under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/2.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.