Phase Transitions in the Distribution of Missing Sums

Hung V. Chu (Washington and Lee University)
Email: chuh19@mail.wlu.edu
Lily Shao (Williams College)
Email: ls12@williams.edu

In joint work with Noah Luntlaza, Steven J. Miller, Victor Xu.

Young Mathematicians Conference
Ohio State University

Introduction

Statement

- A finite set of integers, $|A|$ its size.

Statement

- A finite set of integers, $|A|$ its size.
- The sumset: $A + A = \{a_i + a_j | a_i, a_j \in A\}$.

Statement

- A finite set of integers, $|A|$ its size.
- The sumset: $A + A = \{a_i + a_j | a_i, a_j \in A\}$.
- The difference set: $A - A = \{a_i - a_j | a_i, a_j \in A\}$.

Statement

- A finite set of integers, $|A|$ its size.
- The sumset: $A + A = \{a_i + a_j | a_i, a_j \in A\}$.
- The difference set: $A - A = \{a_i - a_j | a_i, a_j \in A\}$.

Definition

A finite set of integers. A is called **sum-dominated** or MSTD (more-sum-than-difference) if $|A + A| > |A - A|$, **balanced** if $|A + A| = |A - A|$ and **difference-dominated** if $|A + A| < |A - A|$.

False conjecture

- Natural to think that $|A + A| \leq |A - A|$.

False conjecture

- Natural to think that $|A + A| \leq |A - A|$.
- Each pair (x, y), $x \neq y$ gives two differences: $x - y \neq y - x$, but only one sum $x + y$.

False conjecture

- Natural to think that $|A+A| \leq |A-A|$.
- Each pair (x, y), $x \neq y$ gives two differences: $x - y \neq y - x$, but only one sum $x + y$.
- However, sets A with $|A+A| > |A-A|$ do exist!

Examples

- Conway: $A_1 = \{0, 2, 3, 4, 7, 11, 12, 14\}$
- Marica: $A_2 = \{0, 1, 2, 4, 7, 8, 12, 14, 15\}$
- Pigarev and Freiman: $A_3 = \{0, 1, 2, 4, 5, 9, 12, 13, 14, 16, 17, 21, 24, 25, 26, 28, 29\}$
- Hegarty: $A_4 = \{0, 1, 2, 4, 5, 9, 12, 13, 17, 20, 21, 22, 24, 25, 29, 32, 33, 37, 40, 41, 42, 44, 45\}$

Examples

- Conway: $A_1 = \{0, 2, 3, 4, 7, 11, 12, 14\}$
- Marica: $A_2 = \{0, 1, 2, 4, 7, 8, 12, 14, 15\}$
- Pigarev and Freiman: $A_3 = \{0, 1, 2, 4, 5, 9, 12, 13, 14, 16, 17, 21, 24, 25, 26, 28, 29\}$
- Hegarty: $A_4 = \{0, 1, 2, 4, 5, 9, 12, 13, 17, 20, 21, 22, 24, 25, 29, 32, 33, 37, 40, 41, 42, 44, 45\}$
- Hegarty proved that the smallest cardinality of MSTD sets is 8.

Martin and Obryant '06

Theorem

Consider $I_n = \{0, 1, ..., n-1\}$. The proportion of MSTD subsets of I_n is bounded below by a positive constant $c \approx 2 \cdot 10^{-7}$.

Martin and Obryant '06

Theorem

Consider $I_n = \{0, 1, ..., n-1\}$. The proportion of MSTD subsets of I_n is bounded below by a positive constant $c \approx 2 \cdot 10^{-7}$.

- Later, Zhao improved the bound to $4.28 \cdot 10^{-4}$ and proved that the limiting proportion exists.

Martin and Obryant '06

Theorem

Consider $I_n = \{0, 1, ..., n-1\}$. The proportion of MSTD subsets of I_n is bounded below by a positive constant $c \approx 2 \cdot 10^{-7}$.

- Later, Zhao improved the bound to $4.28 \cdot 10^{-4}$ and proved that the limiting proportion exists.
- Probabilistic method.

Results

Overview

- Behavior of the distribution of the number of missing sums.

Overview

- Behavior of the distribution of the number of missing sums.

- The distribution has some strange behavior:

Overview

- Behavior of the distribution of the number of missing sums.
- The distribution has some strange behavior:
  - Not unimodal

Overview

- Behavior of the distribution of the number of missing sums.
- The distribution has some strange behavior:
  - Not unimodal
  - Against missing certain number of sums.

Sets of Missing Sums

- Let $I_n = \{0, 1, 2, ..., n-1\}$.

Sets of Missing Sums

- Let $I_n = \{0, 1, 2, ..., n-1\}$.
- Form $S \subseteq I_n$ **randomly with probability p of picking an element in I_n**. ($q = 1 - p$: the probability of not chosing an element.)

Sets of Missing Sums

- Let $I_n = \{0, 1, 2, ..., n-1\}$.
- Form $S \subseteq I_n$ **randomly with probability p of picking an element in I_n**. ($q = 1 - p$: the probability of not chosing an element.)
- $B_n = (I_n + I_n) \backslash (S + S)$ is the set of missing sums, $|B_n|$: the number of missing sums.

Distribution of Missing Sums

- Fix $p \in (0,1)$, study $\mathbb{P}(|B| = k) = \lim_{n\to\infty} \mathbb{P}(|B_n| = k)$. (Zhao proved that the limit exists.)

Distribution of Missing Sums

- Fix $p \in (0, 1)$, study $\mathbb{P}(|B| = k) = \lim_{n\to\infty} \mathbb{P}(|B_n| = k)$. (Zhao proved that the limit exists.)
- $\mathbb{P}(|B| = k)$: the limiting distribution of missing sums.

Distribution of Missing Sums

- Fix $p \in (0,1)$, study $\mathbb{P}(|B| = k) = \lim_{n\to\infty} \mathbb{P}(|B_n| = k)$. (Zhao proved that the limit exists.)
- $\mathbb{P}(|B| = k)$: the limiting distribution of missing sums.

Divot

For some $k \geq 1$, if
$\mathbb{P}(|B| = k-1) > \mathbb{P}(|B| = k) < \mathbb{P}(|B| = k+1)$, then the distribution of sums has a **divot** at k.

Example of Divot

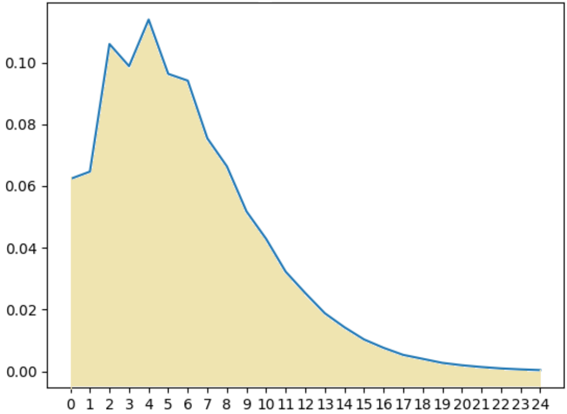


Figure: Frequency of the number of missing sums for subsets of $\{0, 1, 2, ..., 400\}$ by simulating 1,000,000 subsets with $p = 0.6$.

Why Divot?

- Closely related to the behavior of missing sums in the sumset.

Why Divot?

- Closely related to the behavior of missing sums in the sumset.
  Many famous problems can be stated in the language of sumsets: Goldbach's Conjecture and Fermat's Last Theorem.

Why Divot?

- Closely related to the behavior of missing sums in the sumset.
  Many famous problems can be stated in the language of sumsets: Goldbach's Conjecture and Fermat's Last Theorem.
  For example, let P_n be the set $\{1^n, 2^n, 3^n, ...\}$. Then Fermat's Last Theorem is equivalent to
  $(P_n + P_n) \cap P_n = \emptyset$ for all $n \geq 3$.

Why Divot?

- Closely related to the behavior of missing sums in the sumset.
  Many famous problems can be stated in the language of sumsets: Goldbach's Conjecture and Fermat's Last Theorem.
  For example, let P_n be the set $\{1^n, 2^n, 3^n, ...\}$. Then Fermat's Last Theorem is equivalent to $(P_n + P_n) \cap P_n = \emptyset$ for all $n \geq 3$.

- Interesting itself: two-bump distribution.

Numerical Analysis for $p = 1/2$

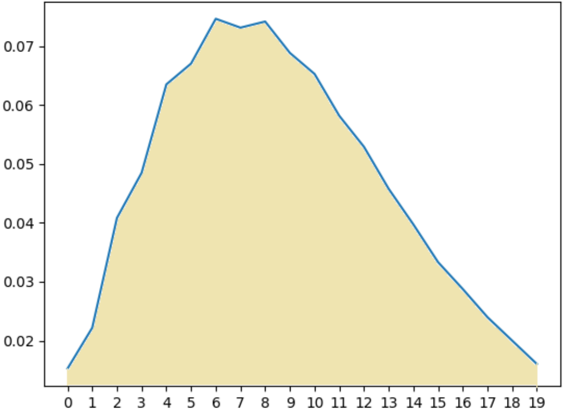


Figure: Frequency of the number of missing sums for all subsets of $\{0, 1, 2, ..., 25\}$.

Lazarev-Miller-O'Bryant '11

Divot at 7

For $p = 1/2$, there is a divot at 7, i.e.
$\mathbb{P}(|B| = 6) > \mathbb{P}(|B| = 7) < \mathbb{P}(|B| = 8)$.

Question

Existence of Divots

For a fixed different value of p, are there other divots?

Question

Existence of Divots

For a fixed different value of p, are there other divots?

Answer: Yes!

Numerical analysis for different $p \in (0, 1)$: $p = 0.6$

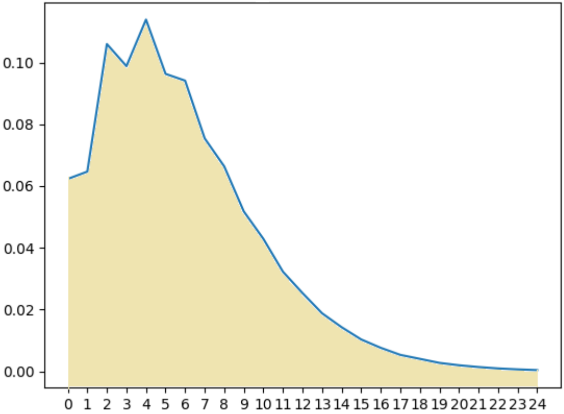


Figure: Distribution of $|B| = k$ by simulating 1,000,000 subsets of $\{0, 1, 2, \ldots, 400\}$ with $p = 0.6$.

Numerical analysis for different $p \in (0, 1) : p = 0.7$

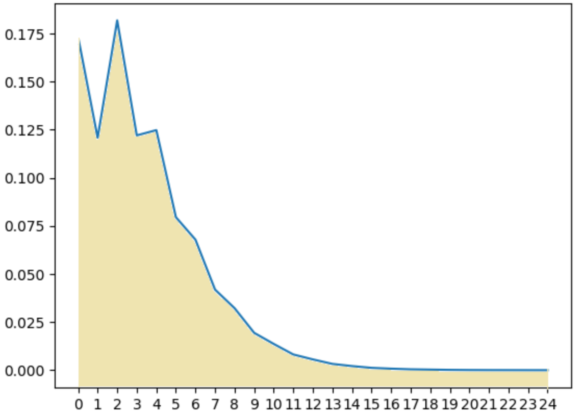


Figure: Distribution of $|B| = k$ by simulating 1,000,000 subsets of $\{0, 1, 2, \ldots, 400\}$ with $p = 0.7$.

Numerical analysis for different $p \in (0, 1)$: $p = 0.8$

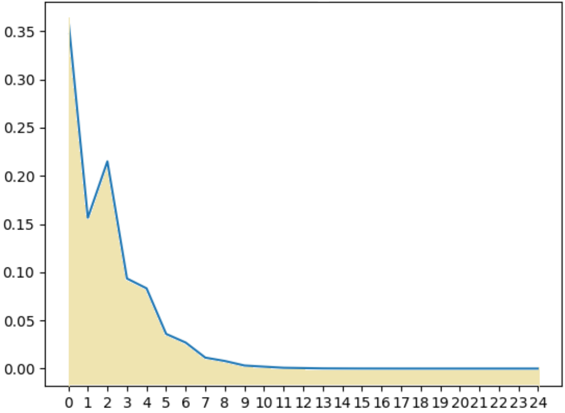


Figure: Distribution of $|B| = k$ by simulating 1,000,000 subsets of $\{0, 1, 2, \ldots, 400\}$ with $p = 0.8$.

Numerical analysis for different $p \in (0, 1) : p = 0.9$

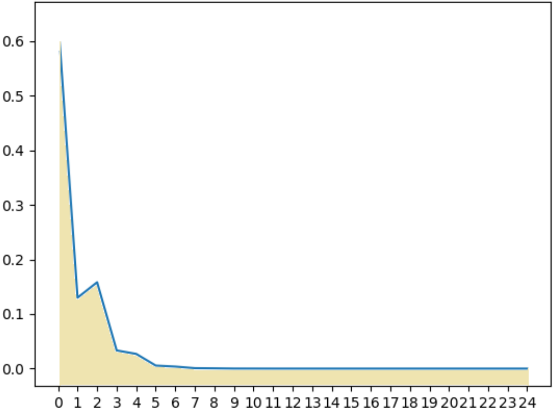


Figure: Distribution of $|B| = k$ by simulating 1,000,000 subsets of $\{0, 1, 2, \ldots, 400\}$ with $p = 0.9$.

Main Result

Chu-Luntzlara-Miller-Shao-Xu

For $p \geq 0.68$, there is a divot at 1, i.e.
$\mathbb{P}(|B| = 0) > \mathbb{P}(|B| = 1) < \mathbb{P}(|B| = 2)$.

Main Result

Chu-Luntzlara-Miller-Shao-Xu

For $p \geq 0.68$, there is a divot at 1, i.e.
$\mathbb{P}(|B| = 0) > \mathbb{P}(|B| = 1) < \mathbb{P}(|B| = 2)$.

- Empirical evidence predicts the value of p such that the divot at 1 starts to exist is between 0.6 and 0.7.

Sketch of Proof

Key Ideas

- Want $\mathbb{P}(|B| = 0) > \mathbb{P}(|B| = 1) < \mathbb{P}(|B| = 2)$.

Key Ideas

- Want $\mathbb{P}(|B| = 0) > \mathbb{P}(|B| = 1) < \mathbb{P}(|B| = 2)$.
- Establish an upper bound T^1 for $\mathbb{P}(|B| = 1)$.

Key Ideas

- Want $\mathbb{P}(|B| = 0) > \mathbb{P}(|B| = 1) < \mathbb{P}(|B| = 2)$.
- Establish an upper bound T^1 for $\mathbb{P}(|B| = 1)$.
- Establish lower bounds T_0 and T_2 for $\mathbb{P}(|B| = 0)$ and $\mathbb{P}(|B| = 2)$, respectively.

Key Ideas

- Want $\mathbb{P}(|B| = 0) > \mathbb{P}(|B| = 1) < \mathbb{P}(|B| = 2)$.
- Establish an upper bound T^1 for $\mathbb{P}(|B| = 1)$.
- Establish lower bounds T_0 and T_2 for $\mathbb{P}(|B| = 0)$ and $\mathbb{P}(|B| = 2)$, respectively.
- Find values of p such that $T_2 > T^1 < T_0$.

Fringe Analysis

- Most of the missing sums come from the fringe: many more ways to form middle elements than fringe elements.

Fringe Analysis

- Most of the missing sums come from the fringe: many more ways to form middle elements than fringe elements.

- Example: let $S \subseteq \{0, 1, ..., 10\}$. Then $S + S \subseteq [0, 20]$. Consider $0 = 0 + 0$ and $20 = 10 + 10$ while $10 = 0 + 10 = 1 + 9 = 2 + 8 = 3 + 7 = 4 + 6 = 5 + 5$.

Fringe Analysis

- Most of the missing sums come from the fringe: many more ways to form middle elements than fringe elements.
- Example: let $S \subseteq \{0, 1, ..., 10\}$. Then $S + S \subseteq [0, 20]$. Consider $0 = 0 + 0$ and $20 = 10 + 10$ while $10 = 0 + 10 = 1 + 9 = 2 + 8 = 3 + 7 = 4 + 6 = 5 + 5$.
- Fringe analysis is enough to find good lower bounds and upper bounds for $\mathbb{P}(|B| = k)$.

Setup

- Consider $S \subseteq \{0, 1, 2, ..., n - 1\}$ with probability p of each element being picked.

Setup

- Consider $S \subseteq \{0, 1, 2, ..., n-1\}$ with probability p of each element being picked.
- Analyze fringe of size 30.

Setup

- Consider $S \subseteq \{0, 1, 2, ..., n-1\}$ with probability p of each element being picked.
- Analyze fringe of size 30.
- Write $S = L \cup M \cup R$, where $L \subseteq [0, 29]$, $M \subseteq [30, n-31]$ and $R \subseteq [n-30, n-1]$.

Notation

- Write $S = L \cup M \cup R$, where
  $L \subseteq [0, 29]$, $M \subseteq [30, n-31]$ and $R \subseteq [n-30, n-1]$.

Notation

- Write $S = L \cup M \cup R$, where
  $L \subseteq [0, 29]$, $M \subseteq [30, n-31]$ and $R \subseteq [n-30, n-1]$.

- L_k : the event that $L + L$ misses k sums in $[0, 29]$.

Notation

- Write $S = L \cup M \cup R$, where $L \subseteq [0, 29]$, $M \subseteq [30, n-31]$ and $R \subseteq [n-30, n-1]$.
- L_k : the event that $L + L$ misses k sums in $[0, 29]$.
- L_k^a: the event that $L + L$ misses k sums in $[0, 29]$ and contains $[30, 48]$.

Notation

- Write $S = L \cup M \cup R$, where $L \subseteq [0, 29]$, $M \subseteq [30, n - 31]$ and $R \subseteq [n - 30, n - 1]$.
- L_k : the event that $L + L$ misses k sums in $[0, 29]$.
- L_k^a: the event that $L + L$ misses k sums in $[0, 29]$ and contains $[30, 48]$.
- Similar notations applied for R.

Upper Bound

Given $0 \le k \le 30$,

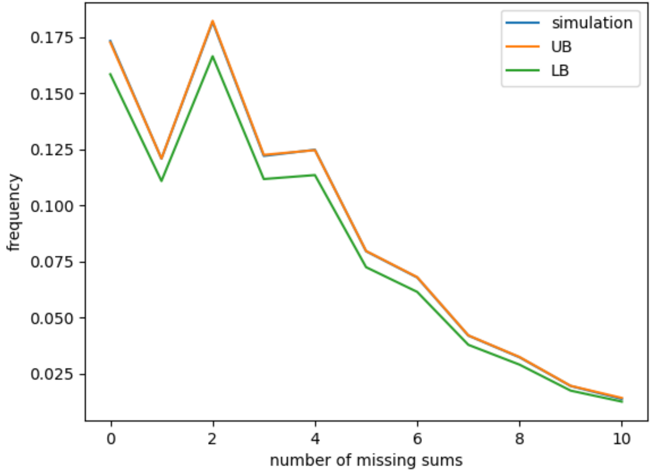
$$\mathbb{P}(|B| = k) \le \sum_{i=0}^{k} \mathbb{P}(L_i) P(L_{k-i}) + \frac{2(2q - q^2)^{15}(3q - q^2)}{(1 - q)^2}. \tag{1}$$

Lower Bound

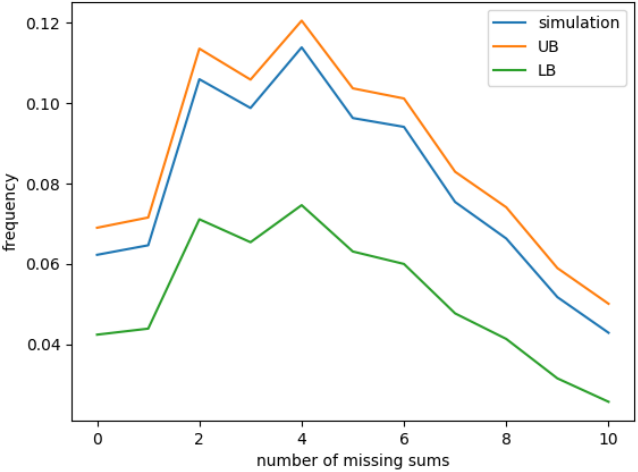
Given $0 \leq k \leq 30$,

$$\mathbb{P}(|B| = k) \geq \sum_{i=0}^{k} \Big[1 - (a-2)(q^{\tau(L_i^a)} + q^{\tau(L_{k-i}^a)}) - \frac{1+q}{(1-q)^2}(q^{\min L_i^a} + q^{\min L_{k-i}^a}) \Big] \mathbb{P}(L_i^a)\mathbb{P}(L_{k-i}^a). \quad (2)$$

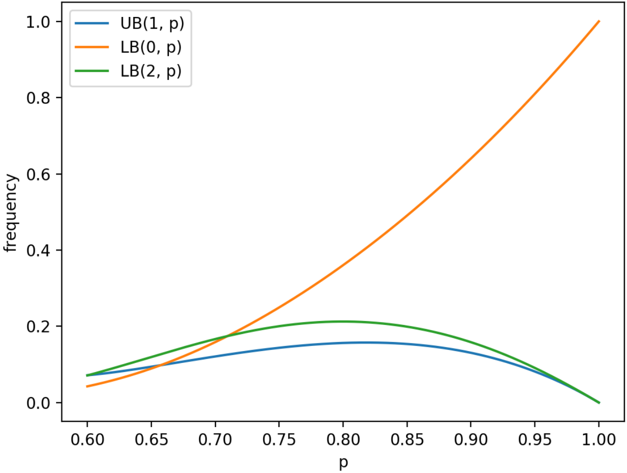
Our Bounds Are Sharp ($p \geq 0.7$)



Our Bounds Are Bad ($p \leq 0.6$)



Divot at 1


UB(1, p)
LB(0, p)
LB(2, p)
frequency
1.0
0.8
0.6
0.4
0.2
0.0
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00
p

Future Research

Question

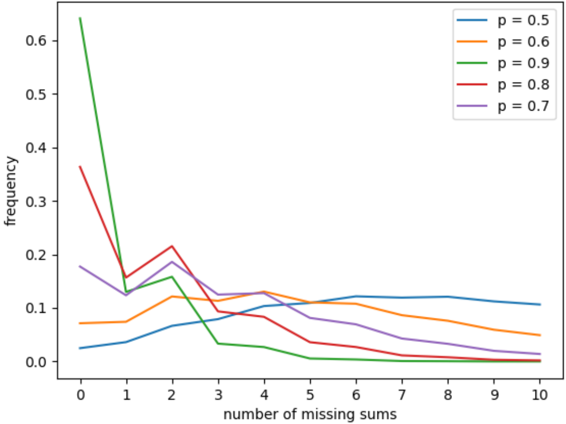


Figure: Shift of Divots

Question

Conjecture

There are no divots at even numbers.

Question

Is there a value of p such that there are no divots?

Bibliography

Bibliography

- O. Lazarev, S. J. Miller, K. O'Bryant, *Distribution of Missing Sums in Sumsets* (2013), Experimental Mathematics **22**, no. 2, 132–156.

- P. V. Hegarty, *Some explicit constructions of sets with more sums than differences* (2007), Acta Arithmetica **130** (2007), no. 1, 61–77.

- P. V. Hegarty and S. J. Miller, *When almost all sets are difference dominated*, Random Structures and Algorithms **35** (2009), no. 1, 118–136.

- G. Iyer, O. Lazarev, S. J. Miller and L. Zhang, *Generalized more sums than differences sets,* Journal of Number Theory **132** (2012), no. 5, 1054–1073.

Bibliography

- J. Marica, *On a conjecture of Conway*, Canad. Math. Bull. **12** (1969), 233–234.
- G. Martin and K. O'Bryant, *Many sets have more sums than differences*, in Additive Combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 287–305.
- M. Asada, S. Manski, S. J. Miller, and H. Suh, *Fringe pairs in generalized MSTD sets*, International Journal of Number Theory **13** (2017), no. 10, 2653â2675.
- S. J. Miller, B. Orosz and D. Scheinerman, *Explicit constructions of infinite families of MSTD sets*, Journal of Number Theory **130** (2010), 1221–1233.

Bibliography

- S. J. Miller and D. Scheinerman, *Explicit constructions of infinite families of mstd sets,* Additive Number Theory, Springer, 2010, pp. 229-248.
- S. J. Miller, S. Pegado and L. Robinson, *Explicit Constructions of Large Families of Generalized More Sums Than Differences Sets*, Integers **12** (2012), #A30.
- M. B. Nathanson, *Problems in additive number theory, 1*, Additive combinatorics, 263–270, CRM Proc. Lecture Notes **43**, Amer. Math. Soc., Providence, RI, 2007.

Bibliography

- M. B. Nathanson, *Sets with more sums than differences*, Integers : Electronic Journal of Combinatorial Number Theory **7** (2007), Paper A5 (24pp).
- Y. Zhao, *Constructing MSTD sets using bidirectional ballot sequences*, Journal of Number Theory **130** (2010), no. 5, 1212–1220.
- Y. Zhao, *Sets characterized by missing sums and differences*, Journal of Number Theory **131** (2011), no. 11, 2107–2134.

THANK YOU!