Sets with More Sums Than Differences

Thao Do, Archit Kulkarni, David Moon, Jake Wellens;
Advisor: Steven J. Miller (sjm1@williams.edu)
Number Theory and Probability Group - SMALL 2013 - Williams College

1. MSTDs in higher dimensional lattices

1.1 Introduction

MSTD: A set $A \subset \mathbb{N}$ is called a *more sums than differences set* or MSTD if $|A+A| = |\{a+b : a, b \in A\}| > |A-A| = |\{a-b : a, b \in A\}|$.

Martin and O'Bryant (2006): For $n > 15$, the proportion of subsets $A \subset \{0, \dots, n\}$ with $|A+A| > |A-A|$ is at least $4 \cdot 10^{-7}$.

Example: Conway discovered the small MSTD set $\{0, 2, 3, 4, 7, 11, 12, 14\}$. Nathanson proved that this is in fact the smallest MSTD subset of $\mathbb{N}$.

Explicit Constructions: Despite the fact that a uniformly random subset of $\{0, \dots, n\}$ is MSTD with positive probability, no one knows of an explicit family of MSTDs which has $O(1)$ density. Miller, Orosz and Sheinerman have constructed an explicit family of subsets of $\{0, \dots, n\}$ of density $O(1/n^4)$. Zhao later used the combinatorial idea of bidirectional ballot sequences to improve this density to $O(1/n)$.

We can extend some of these results to subsets of higher dimensional lattices!

1.2 Our Results

Theorem (Positive percent in $\mathbb{Z}^2$) : Let A be a uniformly random subset of the square $\{0, \dots, n\} \times \{0, \dots, n\}$. Then with positive probability, $|A+A| > |A-A|$.

Proof Idea: The probability that a given lattice point is in the sum or difference set approaches 1 as the point moves away from the corners, since points away from corners have many possible representations as sums and differences. Thus, if we demand that the corners of our sets A have a given fringe profile which is MSTD, then with high probability, all middle sums and differences will be present, and so the sumset will win. By picking the fringes large enough, we can make sure the probability that $A+A$ contains all middle sums is at least some strictly positive constant independent of n.

Note: This technique, which is a generalization of Martin and O'Bryant's technique for one dimensional MSTD sets, also generalizes to higher dimensions and other shapes to give the following theorem:

Theorem (Arbitrary lattices): Let A be a uniformly random subset of some Cartesian product of intervals in d dimensions. Then with positive probability, $|A+A| > |A-A|$.

We also generalize the results of Zhao, who showed that in 1 dimension, the proportion of MSTD subsets of $\{0, \dots, n\}$ converges to a limit as $n \to \infty$.

Theorem (Limiting proportions exist): The proportion of MSTD subsets of the d-cube $\{0, \dots, n\}^d$ converges to a positive limit.

From Monte Carlo experiments, the limiting proportion when $d = 1$ appears to be roughly $4.5 \cdot 10^{-4}$, while when $d = 2$, it appears to be more like $5 \cdot 10^{-8}$. This and other evidence leads us to conjecture...

Conjecture: The limiting proportion of subsets of the d-cube which are MSTD is monotonically decreasing in d.

We have also explicitly constructed a family of MSTD subsets of the square with polynomial density:

Theorem: Set $A \subset \mathbb{Z}$ has property P_n (or is a P_n*-set*) if both its sum set and difference set contain all but the first and last n possible elements (and of course they may or may not contain some of these fringe elements). Explicitly, let $a = \min A$ and $b = \max A$. Then, A is a P_n-set if

$$[2a+n, 2b-n] \subset A+A$$

and

$$[(-b-a)+n, (b-a)-n] \subset A-A.$$

Let $A = L \cup R$ be an MSTD P_n-set with $L \subset [0, n-1]$, $R \subset [n, 2n-1]$, and $0, 2n-1 \in A$. Fix some $k \geq n$, and let $s > 2n+2k$ be arbitrary. Let $A(\mathcal{M}; k) = \mathcal{A} \cup \mathcal{K} \cup \mathcal{M}$, where $\mathcal{A}$, $\mathcal{K}$, and $\mathcal{M}$ are defined as follows:

- Let $A' = L \cup R'$, where $R' = (s-2n) + R$. Let $\mathcal{A} = (A')^2$.
- Let $V = \{(0,0), (s-1, 0), (0, s-1), (s-1, s-1)\}$. Let $k(0) = [n, n+k-1]$ and $k(s-1) = [s-n-k, s-n+1]$. For each $(v_1, v_2) \in V$, let

  $$K_x(v_1, v_2) = k(v_1) \times [0, k-1],$$
  $$K_y(v_1, v_2) = [0, k-1] \times k(v_2).$$

  Finally, let

  $$\mathcal{K} = \bigcup_{(v_1, v_2) \in V} (K_x(v_1, v_2) \cup K_y(v_1, v_2))$$

- Let $B = ([0, s-1] \times \{0, s-1\}) \cup (\{0, s-1\} \times [0, s-1])$. Let $\mathcal{M} \subset [0, s-1]^2 \setminus (\mathcal{A} \cup \mathcal{K})$ such that it has no k-square of missing elements, and $\mathcal{M} \cap B$ has no run of more than k missing elements. To be precise, every subset of M of the form $[a, a+k-1] \times [b, b+k-1]$ must contain at least one element of H, and further, every k consecutive points in M that lie along an edge of our square S must contain at least one element of H.

Then,

(1) $A(\mathcal{M}; k)$ is an MSTD set, and so varying $\mathcal{M}$ gives us an infinite family of distinct MSTD sets in $\mathbb{Z}^2$;

(2) there is a constant $c > 0$ such that the percentage of subsets of $[0, s-1]^2$ that are in this family (and thus are MSTD sets) is at least c/s^{32}.

2. Sums and Differences of Correlated Random Subsets

2.1 Introduction

(p, ρ_1, ρ_2)**-Pairs:** Consider the following variation of the MSTD problem. Instead of placing each element in A with probability p, and then comparing $|A+A|$ and $|A-A|$, we construct a pair of *correlated* random subsets (A, B) as follows: with probability p, an element a goes in A. Then, if $a \in A$, we place it in B with probability ρ_1. If $a \notin A$, we place a in B with probability ρ_2. Then we compare $|A+B|$ and $|\pm(A-B)|$.

Remark: In the special case of $(p, 1, 0)$, this is the standard (A, A) problem. When $\rho_1 = \rho_2$, this is the case of (A, B), where A and B are independently chosen random subsets. The special case $(p, 0, 1)$ corresponds to the interesting case (A, A^c).

2.2 Results

Theorem (Positive Limiting Probabilities Exist): For any $\vec{\rho} = (p, \rho_1, \rho_2)$ with $p(\rho_1 + \rho_2) > 0$, the probability that a $\vec{\rho}$-correlated pair of subsets $(A, B) \subset \{0, \dots, n\}$ is MSTD approches a positive limit $P(\vec{\rho}) > 0$ as $n \to \infty$.

Proof Idea: We follow the main ideas of Zhao, who proved the result in the (A, A) case. We show that in the limit, an MSTD pair is *rich* (meaning the sumset and difference set contain all middle sums and differences) with probability 1. Rich MSTDs are easy to count by their minimal fringe profiles, which we use to show the limit exists.

Conjecture: If we constrain $\rho_1 = 1 - \rho_2$, then for any fixed p, the function $P(\vec{\rho})$ is continuous and strictly increasing in ρ_2.

2.3 Probability Decaying in N

Hegarty and Miller, 2008: When elements chosen with probability $p(N) \to 0$ as $N \to \infty$, then $|A-A| > |A+A|$ almost surely.

Our goal: Extend these results to correlated random subsets.

Lemma: The probability for the event $a_i \in A, b_i \in B$ or $a_i \in B, b_i \in A$ happens for any given (a_i, b_i) is $p^* = p^2(2\rho_1 - \rho_2^2 + 2p(1-p)\rho_2$.

We believe the phase transition happens when $Np^* = \Theta(1)$. When $(\rho_1, \rho_2) = (1, 0)$, $p^* = p^2$, so the phase transition is likely to occur when $p = \Theta(N^{-1/2})$. When $\rho_1 = \rho_2 = p$, this is equivalent to $Np^2(1-p^2) = \Theta(1)$. Since $p \to 0$ as $N \to \infty$, this implies $Np^2 = \Theta(1)$ and so again the phase transition is at $p = \Theta(N^{-1/2})$.

3. References

I. Asimov, *The Endochronic Properties of Resublimated Thiotimoline*, ASF, March 1948.

E. Fein and S. Schneider, *The Rules for Online Dating: Capturing the Heart of Mr. Right in Cyberspace*, page 23.

P. V. Hegarty and S. J. Miller, *When almost all sets are difference dominated*, Random Structures and Algorithms **35** (2009), no. 1, 118–136. http://arxiv.org/abs/0707.3417

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.

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

G. Martin and K. O'Bryant, *Many sets have more sums than differences*, in Additive Combinatorics, CRM Proc. Lecture Notes, vol. 43, AMS, Providence, RI, 2007, pp. 287–305.

J. Stanfield *Utilizing Infinite Loops to Compute an Approximate Value of Infinity*, the Journal of Irreproducible Results **50** (2006), no. 2, page 26.

M. M. Tai, *A Mathematical Model for the Determination of Total Area Under Glucose Tolerance and Other Metabolic Curves*, Diabetes Care February 1994 vol. 17 no. 2 **152-154**

V. Ortiz-Somovilla, F. Espana-Espana, E.J. Pedro-Sanz, and A.J. Gaitan-Juarado, *Meat Mixture Detection in Iberian Pork Sausages*, Meat Science, Volume 71, Issue 3, November 2005, pages **490-497**

Y. Zhao, *Sets Characterized by Missing Sums and Differences*, Journal of Number Theory **131** (2011), 2107–2134.