When Almost All Sets Are Difference Dominated

Steven J Miller
Williams College

Steven.J.Miller@williams.edu
http://web.williams.edu/Mathematics/sjmiller/public_html/

University of Illinois at Urbana-Champaign
Number Theory Seminar, March 26, 2013

Gameplan

- History of the subject.
- Main results and proofs:
  - ◇ Constructing Families
  - ◇ Phase transition
  - ◇ More summands
  - ◇ k-Generational.
- Describe open problems.

Joint with: Peter Hegarty, Ginny Hogan, Geoffrey Iyer, Oleg Lazarev, Brooke Orosz, Dan Scheinerman, Liyang Zhang.

Introduction

Statement

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

- Sumset: $A + A = \{a_i + a_j : a_j, a_j \in A\}$.
- Difference set: $A - A = \{a_i - a_j : a_j, a_j \in A\}$.

Arise in Goldbach's Problem, Twin Primes, Fermat's Last Theorem, ...

Statement

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

- Sumset: $A + A = \{a_i + a_j : a_j, a_j \in A\}$.
- Difference set: $A - A = \{a_i - a_j : a_j, a_j \in A\}$.

Arise in Goldbach's Problem, Twin Primes, Fermat's Last Theorem, ...

Definition

We say A is difference dominated if $|A - A| > |A + A|$, balanced if $|A - A| = |A + A|$ and sum dominated (or an MSTD set) if $|A + A| > |A - A|$.

Questions

Expect generic set to be difference dominated:

- addition is commutative, subtraction isn't:
- Generic pair (x, y) gives 1 sum, 2 differences.

Questions

Expect generic set to be difference dominated:

- addition is commutative, subtraction isn't:
- Generic pair (x, y) gives 1 sum, 2 differences.

Questions

- Do there exist sum-dominated sets?
- If yes, how many?

Examples

Examples

- Conway: $\{0, 2, 3, 4, 7, 11, 12, 14\}$.
- Marica (1969): $\{0, 1, 2, 4, 7, 8, 12, 14, 15\}$.
- Freiman and Pigarev (1973): $\{0, 1, 2, 4, 5, 9, 12, 13, 14, 16, 17, 21, 24, 25, 26, 28, 29\}$.
- Computer search of random subsets of $\{1, \dots, 100\}$: $\{2, 6, 7, 9, 13, 14, 16, 18, 19, 22, 23, 25, 30, 31, 33, 37, 39, 41, 42, 45, 46, 47, 48, 49, 51, 52, 54, 57, 58, 59, 61, 64, 65, 66, 67, 68, 72, 73, 74, 75, 81, 83, 84, 87, 88, 91, 93, 94, 95, 98, 100\}$.
- Recently infinite families (Hegarty, Nathanson).

Infinite Families

Key observation

If A is an arithmetic progression, $|A+A| = |A-A|$.

Infinite Families

Key observation

If A is an arithmetic progression, $|A+A| = |A-A|$.

Proof:

- WLOG, $A = \{0, 1, \dots, n\}$ as $A \to \alpha A + \beta$ doesn't change $|A+A|$, $|A-A|$.

Infinite Families

Key observation

If A is an arithmetic progression, $|A+A| = |A-A|$.

Proof:

- WLOG, $A = \{0, 1, \dots, n\}$ as $A \to \alpha A + \beta$ doesn't change $|A+A|$, $|A-A|$.
- $A + A = \{0, \dots, 2n\}$, $A - A = \{-n, \dots, n\}$, both of size $2n+1$. $\square$

Previous Constructions

Most constructions perturb an arithmetic progression.

Example:

- MSTD set $A = \{0, 2, 3, 4, 7, 11, 12, 14\}$.
- $A = \{0, 2\} \cup \{3, 7, 11\} \cup (14 - \{0, 2\}) \cup \{4\}$.

Example (Nathanson)

Theorem

$m, d, k \in \mathbb{N}$ with $m \geq 4$, $1 \leq d \leq m-1$, $d \neq m/2$, $k \geq 3$ if $d < m/2$ else $k \geq 4$. Let

- $B = [0, m-1] \backslash \{d\}$.
- $L = \{m-d, 2m-d, \dots, km-d\}$.
- $a^* = (k+1)m - 2d$.
- $A^* = B \cup L \cup (a^* - B)$.
- $A = A^* \cup \{m\}$.

Then A is an MSTD set.

Note: gives *exponentially* low density of MSTD sets.

New Explicit Constructions: Results and Notation

Previous best explicit sub-family of $\{1, \dots, n\}$ gave density of $C_1 n^d / 2^{n/2}$.

Our new family gives C_2/n^2, almost a positive percent.

Current record by Zhao: C_3/n.

Notation:

- $[a, b] = \{k \in \mathbb{Z} : a \leq k \leq b\}$.
- A is a P_n-set if its sumset and difference sets contain all but the first and last n possible elements (may or may not contain some of these fringe elements).

New Construction

Theorem (Miller-Orosz-Scheinerman '09)

- *$A = L \cup R$ be a P_n, MSTD set where $L \subset [1, n]$, $R \subset [n+1, 2n]$, and $1, 2n \in A$.*
- *Fix a $k \geq n$ and let m be arbitrary.*
- *M any subset of $[n+k+1, n+k+m]$ st no run of more than k missing elements. Assume $n+k+1 \notin M$.*
- *Set $A(M) = L \cup O_1 \cup M \cup O_2 \cup R'$, where $O_1 = [n+1, n+k]$, $O_2 = [n+k+m+1, n+2k+m]$, and $R' = R + 2k + m$.*

Then $A(M)$ is an MSTD set, and $\exists C > 0$ st the percentage of subsets of $\{0, \dots, r\}$ that are in this family (and thus are MSTD sets) is at least C/r^2.

Phase Transition

Probability Review

X random variable with density $f(x)$ means

- $f(x) \geq 0$;
- $\int_{-\infty}^{\infty} f(x) = 1$;
- $\text{Prob}(X \in [a, b]) = \int_a^b f(x)dx$.

Key quantities:

- Expected (Average) Value: $\mathbb{E}[X] = \int xf(x)dx$.
- Variance: $\sigma^2 = \int (x - \mathbb{E}[X])^2 f(x)dx$.

Binomial model

Binomial model, parameter $p(n)$

Each $k \in \{0, \ldots, n\}$ is in A with probability $p(n)$.

Consider uniform model ($p(n) = 1/2$):

- Let $A \in \{0, \ldots, n\}$. Most elements in $\{0, \ldots, 2n\}$ in $A + A$ and in $\{-n, \ldots, n\}$ in $A - A$.
- $\mathbb{E}[|A + A|] = 2n - 11$, $\mathbb{E}[|A - A|] = 2n - 7$.

Martin and O'Bryant '06

Theorem

Let A be chosen from $\{0, \dots, N\}$ according to the binomial model with constant parameter p (thus $k \in A$ with probability p). At least $k_{\mathrm{SD};p} 2^{N+1}$ subsets are sum dominated.

Martin and O'Bryant '06

Theorem

Let A be chosen from $\{0, \dots, N\}$ according to the binomial model with constant parameter p (thus $k \in A$ with probability p). At least $k_{\mathrm{SD};p} 2^{N+1}$ subsets are sum dominated.

- $k_{\mathrm{SD};1/2} \geq 10^{-7}$, expect about 10^{-3}.

Martin and O'Bryant '06

Theorem

Let A be chosen from $\{0, \dots, N\}$ according to the binomial model with constant parameter p (thus $k \in A$ with probability p). At least $k_{\mathrm{SD};p} 2^{N+1}$ subsets are sum dominated.

- $k_{\mathrm{SD};1/2} \geq 10^{-7}$, expect about 10^{-3}.
- Proof ($p = 1/2$): Generically $|A| = \frac{N}{2} + O(\sqrt{N})$.
  - ◇ about $\frac{N}{4} - \frac{|N-k|}{4}$ ways write $k \in A + A$.
  - ◇ about $\frac{N}{4} - \frac{|k|}{4}$ ways write $k \in A - A$.
  - ◇ Almost all numbers that can be in $A \pm A$ are.
  - ◇ Win by controlling fringes.

Notation

- $X \sim f(N)$ means $\forall \epsilon_1, \epsilon_2 > 0$, $\exists N_{\epsilon_1,\epsilon_2}$ st $\forall N \geq N_{\epsilon_1,\epsilon_2}$

$$\text{Prob}\left(X \not\in [(1-\epsilon_1)f(N), (1+\epsilon_1)f(N)]\right) \; < \; \epsilon_2.$$

Notation

- $X \sim f(N)$ means $\forall \epsilon_1, \epsilon_2 > 0$, $\exists N_{\epsilon_1,\epsilon_2}$ st $\forall N \geq N_{\epsilon_1,\epsilon_2}$

$$\text{Prob}\left(X \notin [(1-\epsilon_1)f(N),(1+\epsilon_1)f(N)]\right) < \epsilon_2.$$

- $\mathcal{S} = |A+A|$, $\mathcal{D} = |A-A|$,
  $\mathcal{S}^{\mathrm{c}} = 2N+1-\mathcal{S}$, $\mathcal{D}^{\mathrm{c}} = 2N+1-\mathcal{D}$.

Notation

- $X \sim f(N)$ means $\forall \epsilon_1, \epsilon_2 > 0$, $\exists N_{\epsilon_1,\epsilon_2}$ st $\forall N \geq N_{\epsilon_1,\epsilon_2}$

$$\text{Prob}\left(X \notin [(1-\epsilon_1)f(N), (1+\epsilon_1)f(N)]\right) \;<\; \epsilon_2.$$

- $\mathcal{S} = |A+A|$, $\mathcal{D} = |A-A|$,
  $\mathcal{S}^{\text{c}} = 2N+1-\mathcal{S}$, $\mathcal{D}^{\text{c}} = 2N+1-\mathcal{D}$.

New model: Binomial with parameter $p(N)$:

- $1/N = o(p(N))$ and $p(N) = o(1)$;
- $\text{Prob}(k \in A) = p(N)$.

Conjecture (Martin-O'Bryant)

As $N \to \infty$, A is a.s. difference dominated.

Main Result

Theorem (Hegarty-Miller)

$p(N)$ as above, $g(x) = 2\frac{e^{-x}-(1-x)}{x}$.

- $p(N) = o(N^{-1/2})$: $\mathcal{D} \sim 2\mathcal{S} \sim (Np(N))^2$;
- $p(N) = cN^{-1/2}$: $\mathcal{D} \sim g(c^2)N$, $\mathcal{S} \sim g\left(\frac{c^2}{2}\right)N$ ($c \to 0$, $\mathcal{D}/\mathcal{S} \to 2$; $c \to \infty$, $\mathcal{D}/\mathcal{S} \to 1$);
- $N^{-1/2} = o(p(N))$: $\mathcal{S}^c \sim 2\mathcal{D}^c \sim 4/p(N)^2$.

Can generalize to binary linear forms or arbitrarily many summands, still have critical threshold.

Inputs

Key input: recent strong concentration results of Kim and Vu (Applications: combinatorial number theory, random graphs, ...).

Need to allow dependent random variables.

Inputs

Key input: recent strong concentration results of Kim and Vu (Applications: combinatorial number theory, random graphs, ...).

Need to allow dependent random variables.

Sketch of proofs: $\mathcal{X} \in \{\mathcal{S}, \mathcal{D}, \mathcal{S}^{c}, \mathcal{D}^{c}\}$.

1. Prove $\mathbb{E}[\mathcal{X}]$ behaves asymptotically as claimed;
2. Prove $\mathcal{X}$ is strongly concentrated about mean.

Setup

Only need new strong concentration for $N^{-1/2} = o(p(N))$.

Will assume $p(N) = o(N^{-1/2})$ as proofs are elementary (i.e., Chebyshev: $\mathrm{Prob}(|Y - \mathbb{E}[Y]| \geq k\sigma_Y) \leq 1/k^2$).

Setup

Only need new strong concentration for $N^{-1/2} = o(p(N))$.

Will assume $p(N) = o(N^{-1/2})$ as proofs are elementary (i.e., Chebyshev: $\text{Prob}(|Y - \mathbb{E}[Y]| \geq k\sigma_Y) \leq 1/k^2$)).

For convenience let $p(N) = N^{-\delta}$, $\delta \in (1/2, 1)$.

IID binary indicator variables:

$$X_{n;N} = \begin{cases} 1 & \text{with probability } N^{-\delta} \\ 0 & \text{with probability } 1 - N^{-\delta}. \end{cases}$$

$X = \sum_{i=1}^{N} X_{n;N}$, $\mathbb{E}[X] = N^{1-\delta}$.

Proof

Lemma

$P_1(N) = 4N^{-(1-\delta)}$,
$\mathcal{O} = \#\{(m,n) : m < n \in \{1,\dots,N\} \bigcap A\}$.
With probability at least $1 - P_1(N)$ *have*

1. $X \in \left[\frac{1}{2}N^{1-\delta}, \frac{3}{2}N^{1-\delta}\right]$.
2. $\frac{\frac{1}{2}N^{1-\delta}(\frac{1}{2}N^{1-\delta}-1)}{2} \le \mathcal{O} \le \frac{\frac{3}{2}N^{1-\delta}(\frac{3}{2}N^{1-\delta}-1)}{2}$.

Proof

Lemma

$P_1(N) = 4N^{-(1-\delta)}$,
$\mathcal{O} = \#\{(m,n) : m < n \in \{1,\dots,N\} \bigcap A\}$.
With probability at least $1 - P_1(N)$ have

1. $X \in \left[\frac{1}{2}N^{1-\delta}, \frac{3}{2}N^{1-\delta}\right]$.
2. $\frac{\frac{1}{2}N^{1-\delta}(\frac{1}{2}N^{1-\delta}-1)}{2} \le \mathcal{O} \le \frac{\frac{3}{2}N^{1-\delta}(\frac{3}{2}N^{1-\delta}-1)}{2}$.

Proof:

- (1) is Chebyshev: $\mathrm{Var}(X) = N\mathrm{Var}(X_{n;N}) \le N^{1-\delta}$.
- (2) follows from (1) and $\binom{r}{2}$ ways to choose 2 from r.

Concentration

Lemma

- $f(\delta) = \min\left(\frac{1}{2}, \frac{3\delta-1}{2}\right)$, $g(\delta)$ *satisfies* $0 < g(\delta) < f(\delta)$.
- $p(N) = N^{-\delta}$, $\delta \in (1/2, 1)$, $P_1(N) = 4N^{-(1-\delta)}$, $P_2(N) = CN^{-(f(\delta)-g(\delta))}$.

With probability at least $1 - P_1(N) - P_2(N)$ *have* $\mathcal{D}/\mathcal{S} = 2 + O(N^{-g(\delta)})$.

Concentration

Lemma

- $f(\delta) = \min\left(\frac{1}{2}, \frac{3\delta-1}{2}\right)$, $g(\delta)$ *satisfies* $0 < g(\delta) < f(\delta)$.
- $p(N) = N^{-\delta}$, $\delta \in (1/2, 1)$, $P_1(N) = 4N^{-(1-\delta)}$, $P_2(N) = CN^{-(f(\delta)-g(\delta))}$.

With probability at least $1 - P_1(N) - P_2(N)$ *have* $\mathcal{D}/\mathcal{S} = 2 + O(N^{-g(\delta)})$.

Proof: Show $\mathcal{D} \sim 2\mathcal{O} + O(N^{3-4\delta})$, $\mathcal{S} \sim \mathcal{O} + O(N^{3-4\delta})$.

As $\mathcal{O}$ is of size $N^{2-2\delta}$ with high probability, need $2 - 2\delta > 3 - 4\delta$ or $\delta > 1/2$.

Analysis of $\mathcal{D}$

Contribution from ‘diagonal’ terms lower order, ignore.

Difficulty: (m, n) and (m', n') could yield same differences.

Analysis of $\mathcal{D}$

Contribution from ‘diagonal’ terms lower order, ignore.

Difficulty: (m, n) and (m', n') could yield same differences.

Notation: $m < n$, $m' < n'$, $m \leq m'$,

$$Y_{m,n,m',n'} = \begin{cases} 1 & \text{if } n - m = n' - m' \\ 0 & \text{otherwise.} \end{cases}$$

Analysis of $\mathcal{D}$

Contribution from 'diagonal' terms lower order, ignore.

Difficulty: (m, n) and (m', n') could yield same differences.

Notation: $m < n$, $m' < n'$, $m \leq m'$,

$$Y_{m,n,m',n'} = \begin{cases} 1 & \text{if } n - m = n' - m' \\ 0 & \text{otherwise.} \end{cases}$$

$\mathbb{E}[Y] \leq N^3 \cdot N^{-4\delta} + N^2 \cdot N^{-3\delta} \leq 2N^{3-4\delta}$. As $\delta > 1/2$, $\#\{\text{bad pairs}\} \lll \mathcal{O}$.

Claim: $\sigma_Y \leq N^{r(\delta)}$with $r(\delta) = \frac{1}{2}\max(3 - 4\delta, 5 - 7\delta)$. This and Chebyshev conclude proof of theorem.

Proof of claim

Cannot use CLT as $Y_{m,n,m',n'}$ are not independent.

Use $\mathrm{Var}(U+V) \leq 2\mathrm{Var}(U) + 2\mathrm{Var}(V)$.

Proof of claim

Cannot use CLT as $Y_{m,n,m',n'}$ are not independent.

Use $\mathrm{Var}(U+V) \le 2\mathrm{Var}(U) + 2\mathrm{Var}(V)$.

Write

$$\sum Y_{m,n,m',n'} \;=\; \sum U_{m,n,m',n'} + \sum V_{m,n,n'}$$

with all indices distinct (at most one in common, if so must be $n = m'$).

$$\mathrm{Var}(U) = \sum \mathrm{Var}(U_{m,n,m',n'}) + 2 \sum_{\substack{(m,n,m',n')\neq \\ (\widetilde{m},\widetilde{n},\widetilde{m}',\widetilde{n}')}} \mathrm{CoVar}(U_{m,n,m',n'}, U_{\widetilde{m},\widetilde{n},\widetilde{m}',\widetilde{n}'})$$

Analyzing $\mathrm{Var}(U_{m,n,m',n'})$

At most N^3 tuples.

Each has variance $N^{-4\delta} - N^{-8\delta} \le N^{-4\delta}$.

Thus $\sum \mathrm{Var}(U_{m,n,m',n'}) \le N^{3-4\delta}$.

Analyzing $\mathrm{CoVar}(U_{m,n,m',n'}, U_{\widetilde{m},\widetilde{n},\widetilde{m}',\widetilde{n}'})$

- All 8 indices distinct: independent, covariance of 0.

- 7 indices distinct: At most N^3 choices for first tuple, at most N^2 for second, get

  $$\mathbb{E}[U_{(1)}U_{(2)}] - \mathbb{E}[U_{(1)}]\mathbb{E}[U_{(2)}] = N^{-7\delta} - N^{-4\delta}N^{-4\delta} \le N^{-7\delta}.$$

- Argue similarly for rest, get $\ll N^{5-7\delta} + N^{3-4\delta}$.

Generalizations

Notation

- As adding sets and not multiplying, set

$$kA \;=\; \underbrace{A + \cdots + A}_{\text{k times}}.$$

- $[a, b] \;=\; \{a, a + 1, \dots, b\}$.

Questions

- Can we find a set A such that $|kA+kA| > |kA-kA|$?
- Can we find a set A such that $|A+A| > |A-A|$ and $|2A+2A| > |2A-2A|$?
- Can we find a set A such that $|kA+kA| > |kA-kA|$ for all k?

Questions

- Can we find a set A such that $|kA + kA| > |kA - kA|$? Yes.

- Can we find a set A such that $|A + A| > |A - A|$ and $|2A + 2A| > |2A - 2A|$? Yes.

- Can we find a set A such that $|kA + kA| > |kA - kA|$ for all k? No. (No such set exists, but can do for arbitrarily many k.)

Initial Observations

Question: Can we find A with $|kA + kA| > |kA - kA|$?

Initial Observations

Question: Can we find A with $|kA + kA| > |kA - kA|$?

- One set gives infinitely many (generalize Miller-Orosz-Scheinerman), more work get positive percentage.

Initial Observations

Question: Can we find A with $|kA + kA| > |kA - kA|$?

- One set gives infinitely many (generalize Miller-Orosz-Scheinerman), more work get positive percentage.
- How do we find one set?

Initial Observations

Question: Can we find A with $|kA + kA| > |kA - kA|$?

- One set gives infinitely many (generalize Miller-Orosz-Scheinerman), more work get positive percentage.
- How do we find one set?
- Computer simulations? We couldn't find a set for $k = 2$; the probability of finding some of these sets is less than 10^{-16}.

Initial Observations

Question: Can we find A with $|kA + kA| > |kA - kA|$?

- One set gives infinitely many (generalize Miller-Orosz-Scheinerman), more work get positive percentage.
- How do we find one set?
- Computer simulations? We couldn't find a set for $k = 2$; the probability of finding some of these sets is less than 10^{-16}.

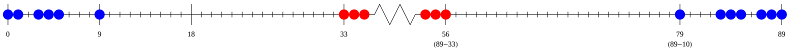
If A is symmetric ($A = c - A$ for some c) then

$$|A + A| = |A + (c - A)| = |A - A|.$$

Example: $|2A+2A| > |2A-2A|$

$$A \;=\; \{0,1,3,4,5,9\} \cup [33,56] \cup \{79,83,84,85,87,88,89\}$$

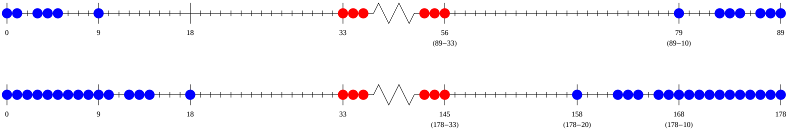
A



Example: $|2A+2A| > |2A-2A|$

$$A = \{0,1,3,4,5,9\} \cup [33,56] \cup \{79,83,84,85,87,88,89\}$$

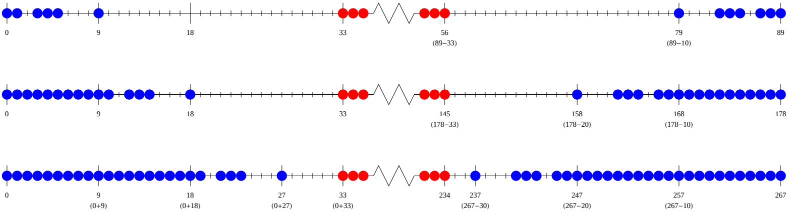
$A + A$



Example: $|2A+2A| > |2A-2A|$

$$A = \{0,1,3,4,5,9\} \cup [33,56] \cup \{79,83,84,85,87,88,89\}$$

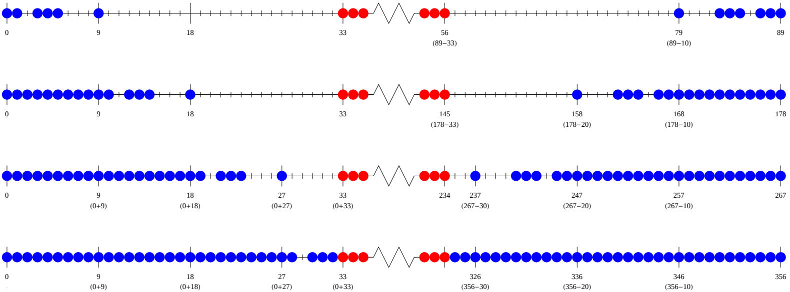
$A+A+A$



Example: $|2A + 2A| > |2A - 2A|$

$$A = \{0, 1, 3, 4, 5, 9\} \cup [33, 56] \cup \{79, 83, 84, 85, 87, 88, 89\}$$

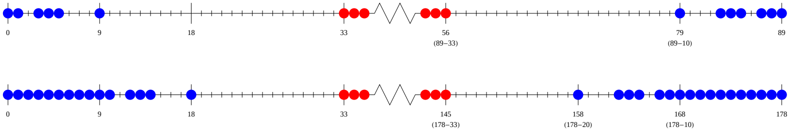
$A + A + A + A$



Example: $|2A+2A| > |2A-2A|$

$$A = \{0,1,3,4,5,9\} \cup [33,56] \cup \{79,83,84,85,87,88,89\}$$

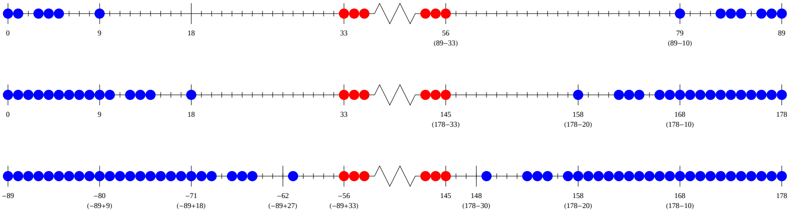
$A+A$



Example: $|2A+2A| > |2A-2A|$

$$A = \{0,1,3,4,5,9\} \cup [33,56] \cup \{79,83,84,85,87,88,89\}$$

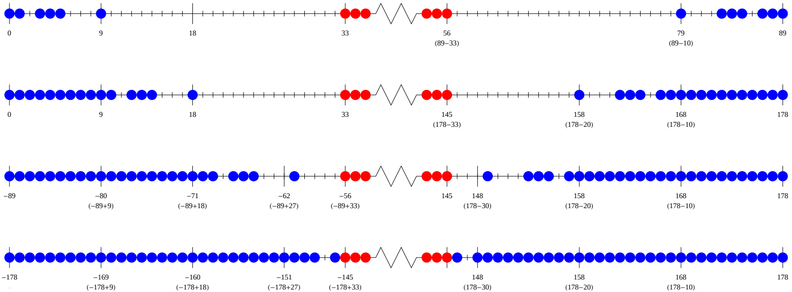
$A + A - A$



Example: $|2A + 2A| > |2A - 2A|$

$$A = \{0, 1, 3, 4, 5, 9\} \cup [33, 56] \cup \{79, 83, 84, 85, 87, 88, 89\}$$

$$A + A - A - A$$



Why the construction worked: Generalization to $xA - yA$

- Write $A = L \sqcup R$ (left and right).

Why the construction worked: Generalization to $xA - yA$

- Write $A = L \sqcup R$ (left and right).
- L, R almost symmetric, R slightly longer.

Why the construction worked: Generalization to $xA - yA$

- Write $A = L \sqcup R$ (left and right).
- L, R almost symmetric, R slightly longer.
- Left of $xA - yA$ is $xL - yR$ (right is $yL - xR$). As $|R| > |L|$, length of fringe depends on number of copies of L, R.

Why the construction worked: Generalization to $xA - yA$

- Write $A = L \sqcup R$ (left and right).
- L, R almost symmetric, R slightly longer.
- Left of $xA - yA$ is $xL - yR$ (right is $yL - xR$). As $|R| > |L|$, length of fringe depends on number of copies of L, R.
- Our example: (1) In $A + A + A + A$ right hits middle, no gaps, left 1 gap. (2) In $A + A - A - A$ left is $L + L - R - R$, length b/w $L + L + L + L$ and $R + R + R + R$ and 1 gap. Right is $R + R - L - L$, also 1 short, so $A + A - A - A$ misses 2.

Generalization

After dealing with some technical obstructions, we can generalize:

Generalization

After dealing with some technical obstructions, we can generalize:

Theorem

For all nontrivial choices of s_1, d_1, s_2, d_2, $\exists A \subseteq \mathbb{Z}$ *such that* $|s_1A - d_1A| > |s_2A - d_2A|$.

Generalization

After dealing with some technical obstructions, we can generalize:

Theorem

For all nontrivial choices of s_1, d_1, s_2, d_2, $\exists A \subseteq \mathbb{Z}$ *such that* $|s_1A - d_1A| > |s_2A - d_2A|$.

Example: We can have $|A + A + A + A| > |A + A + A - A|$:

$$A = \{0, 1, 3, 4, 5, 9, 33, 34, 35, 50, 54, 55, 56, 58, 59, 60\}.$$

k-Generational Sets

Question: Does a set A exist such that $|A+A| > |A-A|$ and $|A+A+A+A| > |A+A-A-A|$?

k-Generational Sets

Question: Does a set A exist such that $|A+A| > |A-A|$ and $|A+A+A+A| > |A+A-A-A|$?

Equivalently: $A, 2A$ are sum-dominant. We say A is 2-generational.

k-Generational Sets

Question: Does a set A exist such that $|A+A| > |A-A|$ and $|A+A+A+A| > |A+A-A-A|$?

Equivalently: $A, 2A$ are sum-dominant. We say A is 2-generational.

More generally, A is k-generational if $|cA+cA| > |cA-cA|$ for all $1 \leq c \leq k$.

k-Generational Sets

Question: Does a set A exist such that $|A+A| > |A-A|$ and $|A+A+A+A| > |A+A-A-A|$?

k-Generational Sets

Question: Does a set A exist such that $|A+A| > |A-A|$ and $|A+A+A+A| > |A+A-A-A|$? Yes!

$$A = \{0, 1, 3, 4, 7, 26, 27, 29, 30, 33, 37, 38, 40, 41, 42, 43, 46, 49, 50, 52, 53, 54, 72, 75, 76, 78, 79, 80\}$$

Theorem

We can find a k-generational set for all k.

k-Generational Sets

Question: Does a set A exist such that $|A+A| > |A-A|$ and $|A+A+A+A| > |A+A-A-A|$? Yes!

$$A = \{0, 1, 3, 4, 7, 26, 27, 29, 30, 33, 37, 38, 40, 41, 42, 43, 46, 49, 50, 52, 53, 54, 72, 75, 76, 78, 79, 80\}$$

Theorem

We can find a k-generational set for all k.

Idea of proof: Find A_j such that $|jA_j + jA_j| > |jA_j - jA_j|$ for each $1 \leq j \leq k$.

Combine the A_j's using the method of base expansion.

Base Expansion

Base Expansion: For sets A_1, A_2 and $m \in \mathbb{N}$ sufficiently large (relative to A_1, A_2) the set

$$A = m \cdot A_1 + A_2$$

behaves like the direct product $A_1 \times A_2 \subseteq \mathbb{Z} \times \mathbb{Z}$.
(here multiplication is the usual scalar multiplication)

Base Expansion

Base Expansion: For sets A_1, A_2 and $m \in \mathbb{N}$ sufficiently large (relative to A_1, A_2) the set

$$A = m \cdot A_1 + A_2$$

behaves like the direct product $A_1 \times A_2 \subseteq \mathbb{Z} \times \mathbb{Z}$.
(here multiplication is the usual scalar multiplication)

In particular:

$$|xA - yA| = |xA_1 - yA_1| \cdot |xA_2 - yA_2|$$

whenever $x + y$ is small relative to m.

Generalization

Theorem

For nontrivial x_j, y_j, w_j, z_j $(2 \leq j \leq k)$, we can find an A such that $|x_jA - y_jA| > |w_jA - z_jA|$ for all j.

Generalization

Theorem

For nontrivial x_j, y_j, w_j, z_j $(2 \leq j \leq k)$, we can find an A such that $|x_jA - y_jA| > |w_jA - z_jA|$ for all j.

Example: We can find an A such that

$$\begin{aligned}
|A+A| &> |A-A| \\
|A+A-A| &> |A+A+A| \\
|5A-2A| &> |A-6A| \\
&\vdots \\
|1870A-141A| &> |1817A-194A|.
\end{aligned}$$

Open Problems

Current and Open Problems

- Similar results for arbitrary finite groups (with Kevin Vissuet).
- Generalize phase transition results for more summands (SMALL ’13 hopefully).
- Generalize to subsets of $\mathbb{Z}^+ \times \mathbb{Z}^+$ (SMALL ’13 hopefully).
- Study the dependence of the divot on $p(N)$.

Divot: Lazarev - Miller - O'Bryant

Let $m(k)$ be the probability a uniformly drawn subset A of $[0, n]$ has $A + A$ missing exactly k summands as $n \to \infty$.

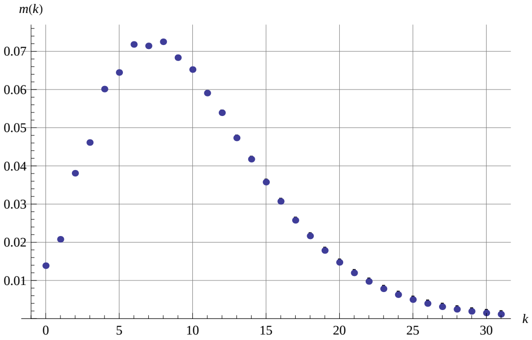


Figure: Experimental values of $m(k)$, vertical bars error (often smaller than dot!).

What happens if draw A from binomial with parameter $p(N)$?

Generalization of main result

Theorem (Hegarty-M): Binomial model with parameter $p(N)$ as before, u, v be non-zero integers with $u \geq |v|$, $\gcd(u, v) = 1$ and $(u, v) \neq (1, 1)$. Put $f(x, y) := ux + vy$ and let $\mathcal{D}_f$ denote the random variable $|f(A)|$. Then the following three situations arise:

1. $p(N) = o(N^{-1/2})$: Then
$$\mathcal{D}_f \sim (N \cdot p(N))^2.$$
2. $p(N) = c \cdot N^{-1/2}$ for some $c \in (0, \infty)$: Define the function $g_{u,v} : (0, \infty) \to (0, u + |v|)$ by
$$g_{u,v}(x) := (u + |v|) - 2|v| \left(\frac{1 - e^{-x}}{x} \right) - (u - |v|)e^{-x}.$$
Then
$$\mathcal{D}_f \sim g_{u,v}\left(\frac{c^2}{u}\right) N.$$
3. $N^{-1/2} = o(p(N))$: Let $\mathcal{D}_f^c := (u + |v|)N - \mathcal{D}_f$. Then $\mathcal{D}_f^c \sim \frac{2u|v|}{p(N)^2}$.

Generalization of Hegarty-Miller

Let f, g be two binary linear forms. Say f dominates g for the parameter $p(N)$ if, as $N \to \infty$, $|f(A)| > |g(A)|$ almost surely when A is a random subset (binomial model with parameter $p(N)$).
Theorem (Hegarty-M): $f(x, y) = u_1 x + u_2 y$ and $g(x, y) = u_2 x + g_2 y$, where $u_i \geq |v_i| > 0$, $\gcd(u_i, v_i) = 1$ and $(u_2, v_2) \neq (u_1, \pm v_1)$. Let

$$\alpha(u, v) := \frac{1}{u^2}\left(\frac{|v|}{3} + \frac{u - |v|}{2}\right) = \frac{3u - |v|}{6u^2}.$$

The following two situations can be distinguished :

- $u_1 + |v_1| \geq u_2 + |v_2|$ and $\alpha(u_1, v_1) < \alpha(u_2, v_2)$. Then f dominates g for all p such that $N^{-3/5} = o(p(N))$ and $p(N) = o(1)$. In particular, every other difference form dominates the form $x - y$ in this range.
- $u_1 + |v_1| > u_2 + |v_2|$ and $\alpha(u_1, v_1) > \alpha(u_2, v_2)$. Then there exists $c_{f,g} > 0$ such that one form dominates for $p(N) < cN^{-1/2}$ $(c < c_{f,g})$ and other dominates for $p(N) > cN^{-1/2}$ $(c > c_{f,g})$.

Open Problems

- One unresolved matter is the comparison of arbitrary difference forms in the range where $N^{-3/4} = O(p)$ and $p = O(N^{-3/5})$. Note that the property of one binary form dominating another is not monotone, or even convex.
- A very tantalizing problem is to investigate what happens while crossing a sharp threshold.
- One can ask if the various concentration estimates can be improved (i.e., made explicit).

Bibliography

Bibliography

- P. V. Hegarty, *Some explicit constructions of sets with more sums than differences* (2007). Acta Arithmetica **130** (2007), no. 1, 61–77. http://arxiv.org/abs/math/0611582
- 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. Hogan and S. J. Miller, *When Generalized Sumsets are Difference Dominated*, preprint. http://arxiv.org/abs/1301.5703
- G. Iyer, O. Lazarev, S. J. Miller and L. Zhang, *Finding and Counting MSTD sets*, to appear in the conference proceedings of the 2011 Combinatorial and Additive Number Theory Conference. http://arxiv.org/abs/1107.2719
- 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. http://arxiv.org/abs/1108.4500
- O. Lazarev, S. J. Miller and K. O'Bryant, *Distribution of Missing Sums in Sumsets*, to appear in Experimental Mathematics. http://arxiv.org/abs/1109.4700

Bibliography (cont)

- 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*. To appear in : Proceedings of CRM-Clay Conference on Additive Combinatorics, Montréal 2006. http://arxiv.org/abs/math/0608131
- S. J. Miller, B. Orosz and D. Scheinerman, *Explicit constructions of infinite families of MSTD sets*, Journal of Number Theory **130** (2010), 1221–1233. http://arxiv.org/abs/0809.4621
- S. J. Miller, S. Pegado and S. Robinson, *Explicit Constructions of Large Families of Generalized More Sums Than Differences Sets*, Integers **12** (2012), #A30. http://arxiv.org/abs/1303.0605
- M. B. Nathanson, *Problems in additive number theory, 1*. To appear in : Proceedings of CRM-Clay Conference on Additive Combinatorics, Montréal 2006. http://arxiv.org/abs/math/0604340
- M. B. Nathanson, *Sets with more sums than differences*, Integers : Electronic Journal of Combinatorial Number Theory **7** (2007), Paper A5 (24pp). http://arxiv.org/abs/math/0608148

Bibliography (cont)

- M. B. Nathanson, K. O'Bryant, B. Orosz, I. Ruzsa and M. Silva, *Binary linear forms over finite sets of integers* (2007). To appear in Acta Arithmetica. http://arxiv.org/abs/math/0701001
- I. Z. Ruzsa, *On the cardinality of $A + A$ and $A - A$*, Combinatorics year (Keszthely, 1976), vol. 18, Coll. Math. Soc. J. Bolyai, North-HollandBolyai Tarsulat, 1978, 933–938.
- V. H. Vu, *New bounds on nearly perfect matchings of hypergraphs: Higher codegrees do help*, Random Structures and Algorithms **17** (2000), 29–63.
- V. H. Vu, *Concentration of non-Lipschitz functions and Applications*, Random Structures and Algorithms **20** (2002), no. 3, 262-316.
- Y. Zhao, *Constructing MSTD Sets Using Bidirectional Ballot Sequences*, Journal of Number Theory **130** (2010), no. 5, 1212–1220. http://arxiv.org/abs/0908.4442
- Y. Zhao, *Sets Characterized by Missing Sums and Differences*, Journal of Number Theory **131** (2011), 2107–2134. http://arxiv.org/abs/0911.2292