A local properties problem for difference sets

Sanjana Das

Massachusetts Institute of Technology

JMM 2024

Motivation

Question

Suppose we have a big set of numbers, and we know that every small subset is ‘arithmetically unstructured.’ How arithmetically unstructured does the *entire* set have to be?

Motivation

Question

Suppose we have a big set of numbers, and we know that every small subset is ‘arithmetically unstructured.’ How arithmetically unstructured does the *entire* set have to be?

- In other words, can we go from ‘*local* arithmetic unstructuredness’ to ‘*global* arithmetic unstructuredness’?

Motivation

Question

Suppose we have a big set of numbers, and we know that every small subset is 'arithmetically unstructured.' How arithmetically unstructured does the *entire* set have to be?

- In other words, can we go from '*local* arithmetic unstructuredness' to '*global* arithmetic unstructuredness'?
- We can measure how 'arithmetically unstructured' a set is by how many distinct differences it contains — we can think of sets with lots of distinct differences as unstructured.

Motivation

Question

Suppose we have a big set of numbers, and we know that every small subset is 'arithmetically unstructured.' How arithmetically unstructured does the *entire* set have to be?

- In other words, can we go from '*local* arithmetic unstructuredness' to '*global* arithmetic unstructuredness'?
- We can measure how 'arithmetically unstructured' a set is by how many distinct differences it contains — we can think of sets with lots of distinct differences as unstructured.

Question

Suppose we have a big set of numbers, and we know that every small subset has lots of distinct differences. How many distinct differences does this force the *entire* set to have?

The Question

Definition

For a set $A \subseteq \mathbb{R}$, its difference set is $A - A = \{a - b \mid a, b \in A,\ a > b\}$.

The Question

Definition

For a set $A \subseteq \mathbb{R}$, its difference set is $A - A = \{a - b \mid a, b \in A,\ a > b\}$.

Definition

We define $g(n, k, \ell)$ to be $\min |A - A|$ over all n-element sets $A \subseteq \mathbb{R}$ with the 'local property' that every k-element subset $A' \subseteq A$ has $|A' - A'| \geq \ell$.

The Question

Definition

For a set $A \subseteq \mathbb{R}$, its difference set is $A - A = \{a - b \mid a, b \in A,\ a > b\}$.

Definition

We define $g(n, k, \ell)$ to be $\min |A - A|$ over all n-element sets $A \subseteq \mathbb{R}$ with the 'local property' that every k-element subset $A' \subseteq A$ has $|A' - A'| \geq \ell$.

Question

For fixed k and ℓ, how does $g(n, k, \ell)$ grow asymptotically with n?

The Question

Definition

For a set $A \subseteq \mathbb{R}$, its difference set is $A - A = \{a - b \mid a, b \in A,\ a > b\}$.

Definition

We define $g(n, k, \ell)$ to be $\min |A - A|$ over all n-element sets $A \subseteq \mathbb{R}$ with the 'local property' that every k-element subset $A' \subseteq A$ has $|A' - A'| \geq \ell$.

Question

For fixed k and ℓ, how does $g(n, k, \ell)$ grow asymptotically with n?

Observation

Any n-element set A satisfies $n - 1 \leq |A - A| \leq \binom{n}{2}$. So we consider ℓ with $k - 1 \leq \ell \leq \binom{k}{2}$; then $g(n, k, \ell)$ is always at least linear in n, and at most quadratic in n.

Thresholds

Question

As we increase ℓ from $k-1$ to $\binom{k}{2}$, at what point does $g(n,k,\ell)$ begin to behave in a certain way?

Thresholds

Question

As we increase ℓ from $k-1$ to $\binom{k}{2}$, at what point does $g(n,k,\ell)$ begin to behave in a certain way?

Definition

- The superlinear threshold is the largest ℓ (as a function of k) for which $g(n,k,\ell) = O(n)$.
- The quadratic threshold is the smallest ℓ with $g(n,k,\ell) = \Omega(n^2)$.

Thresholds

Question

As we increase ℓ from $k-1$ to $\binom{k}{2}$, at what point does $g(n,k,\ell)$ begin to behave in a certain way?

Definition

- The superlinear threshold is the largest ℓ (as a function of k) for which $g(n,k,\ell) = O(n)$.
- The quadratic threshold is the smallest ℓ with $g(n,k,\ell) = \Omega(n^2)$.

Theorem (Li '22)

For each k, the superlinear threshold is $k-1$.

Theorem (Li '22)

For each k, the quadratic threshold is at most $\approx \frac{3}{8}k^2$.

Previous bounds

Several lower bounds are known.

- Fish, Pohoata, Sheffer (2020) proved a family of lower bounds for ℓ between $\approx \frac{7}{32}k^2$ and $\approx \frac{1}{4}k^2$ — e.g., when $4 \mid k$, we have

$$g\left(n, k, \frac{k^2}{4} + 1\right) = \Omega(n^{2-\frac{8}{k}}).$$

Previous bounds

Several lower bounds are known.

- Fish, Pohoata, Sheffer (2020) proved a family of lower bounds for ℓ between $\approx \frac{7}{32}k^2$ and $\approx \frac{1}{4}k^2$ — e.g., when $4 \mid k$, we have

$$g\left(n, k, \frac{k^2}{4} + 1\right) = \Omega(n^{2-\frac{8}{k}}).$$

- Li (2022) proved several others — e.g., when k is a power of 2,

$$g\left(n, k, \frac{k^{\log_2 3} + 1}{2}\right) = \Omega(n^{1+\frac{2}{k-2}}).$$

Previous bounds

Several lower bounds are known.

- Fish, Pohoata, Sheffer (2020) proved a family of lower bounds for ℓ between $\approx \frac{7}{32}k^2$ and $\approx \frac{1}{4}k^2$ — e.g., when $4 \mid k$, we have

$$g\left(n, k, \frac{k^2}{4} + 1\right) = \Omega(n^{2-\frac{8}{k}}).$$

- Li (2022) proved several others — e.g., when k is a power of 2,

$$g\left(n, k, \frac{k^{\log_2 3} + 1}{2}\right) = \Omega(n^{1+\frac{2}{k-2}}).$$

Some upper bounds are known for ‘small’ ℓ (compared to k^2), due to Fish–Lund–Sheffer (2019), Fish–Pohoata–Sheffer (2020), and Li (2022).

- Fish, Lund, Sheffer (2019) proved that

$$g\left(n, k, \frac{k^{\log_2 3} - 1}{2}\right) = O(n^{\log_2 3}).$$

The quadratic threshold

Theorem (D. '23+)

- *For k even, the quadratic threshold is* $\frac{k^2}{4} + 1$.

The quadratic threshold

Theorem (D. '23+)

- *For k even, the quadratic threshold is* $\frac{k^2}{4} + 1$.
- *For k odd, the quadratic threshold is between* $\frac{(k+1)^2}{4} - 3$ *and* $\frac{(k+1)^2}{4}$.

The quadratic threshold

Theorem (D. '23+)

- *For k even, the quadratic threshold is $\frac{k^2}{4}+1$.*
- *For k odd, the quadratic threshold is between $\frac{(k+1)^2}{4}-3$ and $\frac{(k+1)^2}{4}$.*

- To prove that $g(n,k,\frac{k^2}{4}+1)=\Omega(n^2)$, we show that any set A with $|A-A|\ll n^2$ must contain k elements with

$$a_1+a_2=a_3+a_4=\cdots=a_{k-1}+a_k.$$

Then $\{a_1,\ldots,a_k\}$ has at most $\frac{k^2}{4}$ distinct differences.

The quadratic threshold

Theorem (D. '23+)

- *For k even, the quadratic threshold is $\frac{k^2}{4} + 1$.*
- *For k odd, the quadratic threshold is between $\frac{(k+1)^2}{4} - 3$ and $\frac{(k+1)^2}{4}$.*

- To prove that $g(n, k, \frac{k^2}{4} + 1) = \Omega(n^2)$, we show that any set A with $|A - A| \ll n^2$ must contain k elements with

$$a_1 + a_2 = a_3 + a_4 = \cdots = a_{k-1} + a_k.$$

  Then $\{a_1, \ldots, a_k\}$ has at most $\frac{k^2}{4}$ distinct differences.
- To prove that $g(n, k, \frac{k^2}{4}) = o(n^2)$, we use a random construction. We analyze which k-element 'configurations' are expected to appear in it, and show that all of them have at least $\frac{k^2}{4}$ distinct differences (i.e., the configuration $a_1 + a_2 = \cdots = a_{k-1} + a_k$ is the 'worst').

Intermediate bounds

Theorem (D. '23+)

For all $1 < c \leq 2$, *we have* $g(n,k,\ell) = o(n^c)$ *for* $\ell \approx \left(\frac{c-1}{c}\right)^2 k^2$.

Intermediate bounds

Theorem (D. '23+)

For all $1 < c \leq 2$*, we have* $g(n,k,\ell) = o(n^c)$ *for* $\ell \approx \left(\frac{c-1}{c}\right)^2 k^2$.

Theorem (D. '23+)

For all $t \in \mathbb{N}$*, we have* $g(n,k,\ell) = \Omega(n^{1+\frac{1}{2^t-1}})$ *for* $\ell \approx \frac{1}{3} \cdot \left(\frac{3}{4}\right)^t k^2$.

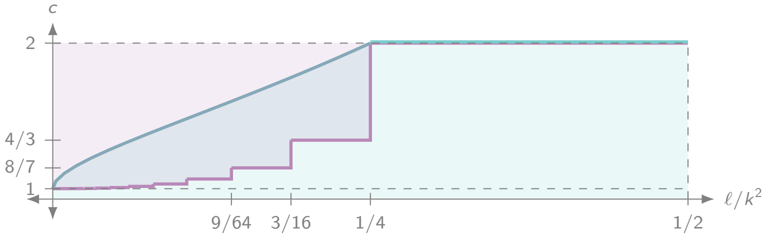
Intermediate bounds

Theorem (D. '23+)

For all $1 < c \leq 2$, *we have* $g(n,k,\ell) = o(n^c)$ *for* $\ell \approx \left(\frac{c-1}{c}\right)^2 k^2$.

Theorem (D. '23+)

For all $t \in \mathbb{N}$, *we have* $g(n,k,\ell) = \Omega(n^{1+\frac{1}{2^t-1}})$ *for* $\ell \approx \frac{1}{3} \cdot \left(\frac{3}{4}\right)^t k^2$.



Thresholds for $\Omega(n^c)$

Definition

For each $1 < c \leq 2$, we define the threshold for $\Omega(n^c)$ as the smallest ℓ (as a function of k) for which $g(n, k, \ell) = \Omega(n^c)$.

Thresholds for $\Omega(n^c)$

Definition

For each $1 < c \leq 2$, we define the threshold for $\Omega(n^c)$ as the smallest ℓ (as a function of k) for which $g(n, k, \ell) = \Omega(n^c)$.

Corollary (D. '23+)

For each $1 < c \leq 2$, the threshold for $\Omega(n^c)$ is quadratic in k (i.e., between $\approx a_1 k^2$ and $a_2 k^2$ for some a_1 and a_2).

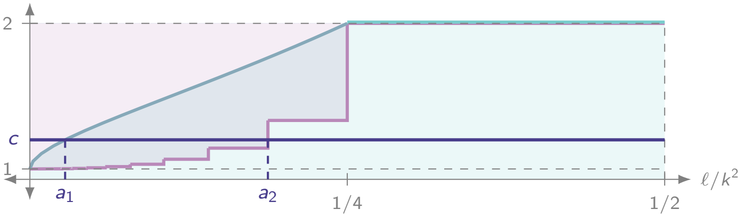
Thresholds for $\Omega(n^c)$

Definition

For each $1 < c \leq 2$, we define the threshold for $\Omega(n^c)$ as the smallest ℓ (as a function of k) for which $g(n,k,\ell) = \Omega(n^c)$.

Corollary (D. '23+)

For each $1 < c \leq 2$, the threshold for $\Omega(n^c)$ is quadratic in k (i.e., between $\approx a_1 k^2$ and $a_2 k^2$ for some a_1 and a_2).



The number of possible exponents

Definition

We define S_k as the set of 'possible exponents' of n in $g(n,k,\ell)$, i.e.,

$$S_k = \left\{ \liminf_{n\to\infty} \frac{\log g(n,k,\ell)}{\log n} \;\middle|\; k-1 \le \ell \le \binom{k}{2} \right\}.$$

The number of possible exponents

Definition

We define S_k as the set of 'possible exponents' of n in $g(n,k,\ell)$, i.e.,

$$S_k = \left\{ \liminf_{n\to\infty} \frac{\log g(n,k,\ell)}{\log n} \;\middle|\; k-1 \le \ell \le \binom{k}{2} \right\}.$$

Corollary (D. '23+)

For some constant $a > 0$, we have $|S_k| \ge a \log\log k$ for all k.

The number of possible exponents

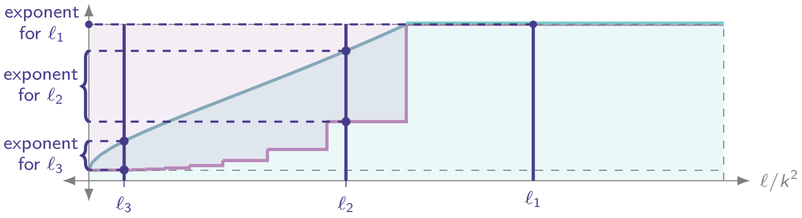
Definition

We define S_k as the set of 'possible exponents' of n in $g(n,k,\ell)$, i.e.,

$$S_k = \left\{ \liminf_{n\to\infty} \frac{\log g(n,k,\ell)}{\log n} \;\middle|\; k-1 \le \ell \le \binom{k}{2} \right\}.$$

Corollary (D. '23+)

For some constant $a > 0$, we have $|S_k| \ge a \log\log k$ for all k.



Acknowledgements

This research was conducted at the University of Minnesota Duluth REU and supported by the generosity of Jane Street Capital, the National Security Agency, and the CYAN Undergraduate Mathematics Fund. I would like to thank Joe Gallian and Colin Defant for organizing the REU, and Noah Kravitz, Maya Sankar, and Yelena Mandelshtam for helpful guidance.

Thanks for listening!