The Typical Structure of Small Sumsets

M. Campos  M. Coulson  G. Perarnau  O. Serra  M. Wötzel

1Universitat Politècnica de Catalunya & Barcelona Graduate School of Mathematics

Additive Combinatorics in Marseille 2020

September 9, 2020

*Thanks to support from MDM-2014-0445.
Classical Inverse Results and an Observation

**Theorem (Folklore)**

Suppose \( A, B \subset \mathbb{Z} \) are finite. Then \( |A + B| = |A| + |B| - 1 \), if and only if \( A \) and \( B \) are arithmetic progressions with the same common difference.

**Theorem (Freiman)**

If \( A \subset \mathbb{Z} \) is finite such that \( |2A| \leq K|A| \), then \( A \) is contained in a generalized arithmetic progression of dimension \( d(K) \) and size \( f(K)|A| \).
Theorem (Folklore)

Suppose \( A, B \subset \mathbb{Z} \) are finite. Then \(|A + B| = |A| + |B| - 1\), if and only if \( A \) and \( B \) are arithmetic progressions with the same common difference.

Theorem (Freiman)

If \( A \subset \mathbb{Z} \) is finite such that \(|2A| \leq K|A|\), then \( A \) is contained in a generalized arithmetic progression of dimension \( d(K) \) and size \( f(K)|A| \).

- Suppose that \( P \subset \mathbb{Z} \) is an arithmetic progression of size \( Ks/2 \).
- If \( A \subset P \) is an arbitrary subset of size \( s \), then we clearly have \( 2A \subset 2P \) and hence \(|2A| \leq |2P| \approx Ks \).
- Question: Can we get an inverse result in this direction by going to the random setting?
Question: Can we get an inverse result in this direction by going to the random setting?

Answer: Yes, Campos proved the following rough structural result.

Theorem (Campos)

Let $s = \Omega((\log n)^3)$ and $K = O(s/(\log n)^3)$. 

Note: Campos also proved a counting result about the number of $s$-sets with doubling constant $K$. 

M. Wötzel (UPC & BGSMath)
The Typical Case - Rough Structure

- Question: Can we get an inverse result in this direction by going to the random setting?
- Answer: Yes, Campos proved the following rough structural result.

**Theorem (Campos)**

Let $s = \Omega((\log n)^3)$ and $K = O(s/(\log n)^3)$. Then for almost all sets $A \subset [n]$ with $|A| = s$ and $|2A| \leq Ks$ there exists an arithmetic progression $P$ of size

\[
|P| \leq \frac{1 + o(1)}{2} Ks
\]

such that at most $o(s)$ points of $A$ are not contained in $P$.

Note: Campos also proved a counting result about the number of $s$-sets with doubling constant $K$. 
The Typical Case - Precise Structure

This was then used to get the following more precise structural result in the case \( K = O(1) \).

**Theorem (Campos, Collares, Morris, Morrison, Souza)**

*Fix \( K \geq 3 \) and \( \epsilon > 0 \). For \( n \) sufficiently large, let \( s \geq (\log n)^4 \).*
The Typical Case - Precise Structure

This was then used to get the following more precise structural result in the case $K = O(1)$.

**Theorem (Campos, Collares, Morris, Morrison, Souza)**

*Fix $K \geq 3$ and $\epsilon > 0$. For $n$ sufficiently large, let $s \geq (\log n)^4$. Then for all but an $\epsilon$ proportion of sets $A \subset [n]$ with $|A| = s$ and $|2A| \leq Ks$, it holds that $A$ is contained in an arithmetic progression $P$ of size

$$|P| \leq \frac{Ks}{2} + c(K, \epsilon),$$

where $c(K, \epsilon) = O(K^2 \log K \log(1/\epsilon))$.***
The Typical Case - Precise Structure

This was then used to get the following more precise structural result in the case $K = O(1)$.

**Theorem (Campos, Collares, Morris, Morrison, Souza)**

> Fix $K \geq 3$ and $\epsilon > 0$. For $n$ sufficiently large, let $s \geq (\log n)^4$. Then for all but an $\epsilon$ proportion of sets $A \subset [n]$ with $|A| = s$ and $|2A| \leq Ks$, it holds that $A$ is contained in an arithmetic progression $P$ of size

$$|P| \leq \frac{Ks}{2} + c(K, \epsilon),$$

where $c(K, \epsilon) = O(K^2 \log K \log(1/\epsilon))$.

Note that the value of $c$ is close to optimal, as there is also a lower bound of the form $\Omega(K^2 \log(1/\epsilon))$. 
Our Result

We generalize Campos’ original result to distinct sets.

**Theorem (Campos, Coulson, Perarnau, Serra, W.)**

Let $s = \Omega((\log n)^3)$ and $K = O(s/(\log n)^3)$.

$\text{Typical Small Sumsets}$

CAM2020 5/17
Our Result

We generalize Campos’ original result to distinct sets.

**Theorem (Campos, Coulson, Perarnau, Serra, W.)**

Let \( s = \Omega((\log n)^3) \) and \( K = O(s/(\log n)^3) \). Then for almost all sets \( A, B \subset [n] \) with \( |A| = |B| = s \) and \( |A + B| \leq Ks \) there exist arithmetic progressions \( P, Q \) with the same common difference of size

\[
|P|, |Q| \leq \frac{1 + o(1)}{2} Ks
\]

such that \( |A \setminus P|, |B \setminus Q| = o(s) \).

More precise bounds for the \( o \) terms are obtained, similar but slightly weaker to those of Campos.

Main tool in the proof: recent version of the method of hypergraph containers.
The Method of Hypergraph Containers

- Introduced explicitly in 2015 by Balogh, Morris and Samotij, and independently by Saxton and Thomason, used to count the number of combinatorial objects avoiding some specific substructure.
The Method of Hypergraph Containers

- Introduced explicitly in 2015 by Balogh, Morris and Samotij, and independently by Saxton and Thomason, used to count the number of combinatorial objects avoiding some specific substructure.

- General idea: If $\mathcal{H}$ is a hypergraph satisfying some specific degree conditions, then there exists a relatively small family $C \subset 2^{V(\mathcal{H})}$ of containers such that:
  1. For each independent set $I \subseteq V(\mathcal{H})$, there exists a container $C$ with $I \subseteq C$.
  2. Each container $C$ is smaller than the original set $V(\mathcal{H})$ by some constant factor.

As long as the degree conditions are met, one can now reapply this result to the induced hypergraph on each container $C$. 
The Method of Hypergraph Containers

- Introduced explicitly in 2015 by Balogh, Morris and Samotij, and independently by Saxton and Thomason, used to count the number of combinatorial objects avoiding some specific substructure.

- General idea: If $\mathcal{H}$ is a hypergraph satisfying some specific degree conditions, then there exists a relatively small family $C \subset 2^{V(\mathcal{H})}$ of containers such that:
  
  1. For each independent set $I \subset V(\mathcal{H})$, there exists a container $C \in C$ with $I \subset C$.
  2. Each $C$ is smaller than $V(\mathcal{H})$ by some constant factor.
The Method of Hypergraph Containers

- Introduced explicitly in 2015 by Balogh, Morris and Samotij, and independently by Saxton and Thomason, used to count the number of combinatorial objects avoiding some specific substructure.

- General idea: If $\mathcal{H}$ is a hypergraph satisfying some specific degree conditions, then there exists a relatively small family $C \subset 2^{V(\mathcal{H})}$ of containers such that:
  1. For each independent set $I \subset V(\mathcal{H})$, there exists a container $C \in C$ with $I \subset C$.
  2. Each $C$ is smaller than $V(\mathcal{H})$ by some constant factor.

- As long as the degree conditions are met, one can now reapply this result to the induced hypergraph on each container $C \in C$. 

After iterating, end up with a still small-ish collection of containers, each of which is very small, and it still holds that every independent set $I$ of $\mathcal{H}$ is contained in one of these containers.
The Method of Hypergraph Containers

- After iterating, end up with a still small-ish collection of containers, each of which is very small, and it still holds that every independent set $I$ of $\mathcal{H}$ is contained in one of these containers.

- Consider now a specific $\mathcal{H}$ whose hyperedges encode some structure (e.g. triangles in $K_n$).
After iterating, end up with a still small-ish collection of containers, each of which is very small, and it still holds that every independent set $I$ of $\mathcal{H}$ is contained in one of these containers.

Consider now a specific $\mathcal{H}$ whose hyperedges encode some structure (e.g. triangles in $K_n$).

The iteration stopped, hence the induced hypergraph on each container has few edges (i.e., few triangles).
The Method of Hypergraph Containers

- After iterating, end up with a still small-ish collection of containers, each of which is very small, and it still holds that every independent set $I$ of $\mathcal{H}$ is contained in one of these containers.

- Consider now a specific $\mathcal{H}$ whose hyperedges encode some structure (e.g. triangles in $K_n$).

- The iteration stopped, hence the induced hypergraph on each container has few edges (i.e., few triangles).

- Can now use classical stability results to get structural statements about the containers, which translate down to the independent sets (i.e. triangle free graphs).
Original method works well when e.g. counting the number or determining structure of $H$-free graphs, for some fixed graph $H$. 

Problem: When trying to get results about induced subgraphs, one wants to consider both edges and non-edges. In order to give a structure theorem about induced-$C_4$-free graphs, Morris, Samotij and Saxton recently developed an asymmetric (bipartite) version of the container method. Campos slightly modified this so that the two components can shrink at different rates and allowed non-uniform hypergraphs.
An Asymmetric Version of the Container Lemma

- Original method works well when e.g. counting the number or determining structure of $H$-free graphs, for some fixed graph $H$.
- Problem: When trying to get results about induced subgraphs, one wants to consider both edges and non-edges.
Original method works well when e.g. counting the number or determining structure of $H$-free graphs, for some fixed graph $H$.

Problem: When trying to get results about induced subgraphs, one wants to consider both edges and non-edges.

In order to give a structure theorem about induced-$C_4$-free graphs, Morris, Samotij and Saxton recently developed an asymmetric (bipartite) version of the container method.
Original method works well when e.g. counting the number or determining structure of $H$-free graphs, for some fixed graph $H$.

Problem: When trying to get results about induced subgraphs, one wants to consider both edges and non-edges.

In order to give a structure theorem about induced-$C_4$-free graphs, Morris, Samotij and Saxton recently developed an asymmetric (bipartite) version of the container method.

Campos slightly modified this so that the two components can shrink at different rates and allowed non-uniform hypergraphs.
Proof Outline – Generalizing the Asymmetric Container Lemma

- We generalize the bipartite container lemma to an $r$-partite, $r$-bounded version and apply it to the following hypergraph.
We generalize the bipartite container lemma to an $r$-partite, $r$-bounded version and apply it to the following hypergraph.

For $(V_1, V_2, V_3) \in 2^{[n]} \times 2^{[n]} \times 2^{[2n]}$, we consider the 3-partite and 3-uniform hypergraph $\mathcal{H}(V_1, V_2, V_3)$ with vertex set $V_1 \cup V_2 \cup V_3$ and hyperedges

$$\{(x, y, z) \in V_1 \times V_2 \times V_3 : x + y = z\}.$$
Proof Outline – Generalizing the Asymmetric Container Lemma

- We generalize the bipartite container lemma to an $r$-partite, $r$-bounded version and apply it to the following hypergraph.
- For $(V_1, V_2, V_3) \in 2^n \times 2^n \times 2^{2n}$, we consider the 3-partite and 3-uniform hypergraph $\mathcal{H}(V_1, V_2, V_3)$ with vertex set $V_1 \cup V_2 \cup V_3$ and hyperedges

  $$\{(x, y, z) \in V_1 \times V_2 \times V_3 : x + y = z\}.$$ 

- Note that the independent sets $I$ we are interested in will satisfy the additional condition that they have a large intersection with the third component, more precisely $|I \cap V_3| \geq |V_3| - Ks$. 
Proof Outline – Generalizing the Asymmetric Container Lemma

- We generalize the bipartite container lemma to an $r$-partite, $r$-bounded version and apply it to the following hypergraph.

- For $(V_1, V_2, V_3) \in 2^n \times 2^n \times 2^{2n}$, we consider the 3-partite and 3-uniform hypergraph $H(V_1, V_2, V_3)$ with vertex set $V_1 \cup V_2 \cup V_3$ and hyperedges
  \[\{(x, y, z) \in V_1 \times V_2 \times V_3 : x + y = z\}\].

- Note that the independent sets $I$ we are interested in will satisfy the additional condition that they have a large intersection with the third component, more precisely $|I \cap V_3| \geq |V_3| - Ks$.

- Think: We want to count triples $(A, B, C)$ such that $|A| = |B| = s$, $A + B \subset C$ and $|C| \leq Ks$. 
Proof Outline – The Container Theorem

We get the following theorem.

**Theorem (CCPOW)**

Let $n, m, s$ be integers such that $\log n \leq s \leq m \leq s^2$ and $\epsilon > 0$. 

*Proof Outline*
Proof Outline – The Container Theorem

We get the following theorem.

**Theorem (CCPOW)**

Let $n, m, s$ be integers such that $\log n \leq s \leq m \leq s^2$ and $\epsilon > 0$. There is a family $\mathcal{A} \subset 2^{[n]} \times 2^{[n]} \times 2^{[2n]}$ of triples of sets $(X, Y, Z)$ of size

$$|\mathcal{A}| \leq \exp(O(\sqrt{m\epsilon^{-2}(\log n)^{3/2}))),$$

such that:

1. For all $A \subseteq B \subseteq n$, $C \subseteq 2n$ with $|A| = |B| = s$, $A \subseteq X$, $B \subseteq Y$ and $Z \subseteq C$.
2. For every $(X, Y, Z) \in \mathcal{A}$, $|Z| \leq m$ and either $\max(|X| - |Y|) \geq m \log n$ or there are at most $\epsilon 2^{-j} (j)$ pairs $(x, y) \subseteq X \cup Y$ such that $x \not\subseteq Z$.
Proof Outline – The Container Theorem

We get the following theorem.

**Theorem (CCPOW)**

Let $n, m, s$ be integers such that $\log n \leq s \leq m \leq s^2$ and $\epsilon > 0$. There is a family $\mathcal{A} \subset 2^{[n]} \times 2^{[n]} \times 2^{[2n]}$ of triples of sets $(X, Y, Z)$ of size

$$|\mathcal{A}| \leq \exp(O(\sqrt{m\epsilon^{-2}(\log n)^{3/2}})),$$

such that:

1. For all $A, B \subset [n], C \subset [2n]$ with $|A| = |B| = s$, $A + B \subset C$ and $|C| \leq m$, there is a triple $(X, Y, Z) \in \mathcal{A}$ such that $A \subset X$, $B \subset Y$ and $Z \subset C$. 

We get the following theorem.

**Theorem (CCPOW)**

Let $n, m, s$ be integers such that $\log n \leq s \leq m \leq s^2$ and $\epsilon > 0$. There is a family $\mathcal{A} \subset 2^{[n]} \times 2^{[n]} \times 2^{[2n]}$ of triples of sets $(X, Y, Z)$ of size $|\mathcal{A}| \leq \exp(O(\sqrt{m}\epsilon^{-2}(\log n)^{3/2}))$, such that:

1. For all $A, B \subset [n], C \subset [2n]$ with $|A| = |B| = s$, $A + B \subset C$ and $|C| \leq m$, there is a triple $(X, Y, Z) \in \mathcal{A}$ such that $A \subset X$, $B \subset Y$ and $Z \subset C$.

2. For every $(X, Y, Z) \in \mathcal{A}, |Z| \leq m$ and either $\max\{|X|, |Y|\} < m/\log n$ or there are at most $\epsilon^2|X||Y|$ pairs $(x, y) \in X \times Y$ such that $x + y \notin Z$. 
Proof Outline – A Stability Result for Distinct Sets of Different Sizes

- Note that while the sets $A, B$ in the end will have the same cardinality, the containers $X$ and $Y$ might not, so we need a stability result for this more general case.

Following arguments by Lev and Shao and Xu, we prove the following.

**Lemma**

Let $\epsilon$ be such that

$$\epsilon > \frac{1}{9}.$$ 

Let $X - Y - Z$ with

$$1 - \epsilon \leq \frac{|Z|}{|X|^2}.$$

and

$$\max f|_X - |_Y > 1 - 4\sqrt{\epsilon |Z|}.$$

then one of the following holds:

1. There are at least $\epsilon |X|^2$ pairs $x, y$ such that $x \not \in Z$.
2. There are arithmetic progressions $P - Q$ of length at most $\epsilon |Z|^2$ with the same common difference such that $P$ contains all but at most $\epsilon |X|$ points of $X$, and similarly for $Q$ and $Y$.

If the second property holds, we say that $X$ is $\epsilon$-close to $P$. 

M. Wötzel (UPC & BGSMath)
Proof Outline – A Stability Result for Distinct Sets of Different Sizes

- Note that while the sets $A$, $B$ in the end will have the same cardinality, the containers $X$ and $Y$ might not, so we need a stability result for this more general case.
- Following arguments by Lev and Shao and Xu, we prove the following.

**Lemma**

Let $\epsilon \leq 2^{-9}$. Let $X, Y, Z \subset \mathbb{Z}$ such that $(1 - \epsilon)|Z| \leq |X| + |Y|$ and $\max\{|X|, |Y|\} \leq (1 + 4\sqrt{\epsilon})|Z|/2$, then one of the following holds:
Proof Outline – A Stability Result for Distinct Sets of Different Sizes

- Note that while the sets $A$, $B$ in the end will have the same cardinality, the containers $X$ and $Y$ might not, so we need a stability result for this more general case.

- Following arguments by Lev and Shao and Xu, we prove the following.

**Lemma**

Let $\epsilon \leq 2^{-9}$. Let $X, Y, Z \subset \mathbb{Z}$ such that $(1 - \epsilon)|Z| \leq |X| + |Y|$ and $\max\{|X|, |Y|\} \leq (1 + 4\sqrt{\epsilon})|Z|/2$, then one of the following holds:

1. There are at least $\epsilon^2|X||Y|$ pairs $(x, y) \in X \times Y$ such that $x + y \notin Z$.  

M. Wötzel (UPC & BGSMath)
Proof Outline – A Stability Result for Distinct Sets of Different Sizes

- Note that while the sets $A$, $B$ in the end will have the same cardinality, the containers $X$ and $Y$ might not, so we need a stability result for this more general case.

- Following arguments by Lev and Shao and Xu, we prove the following.

**Lemma**

Let $\epsilon \leq 2^{-9}$. Let $X, Y, Z \subset \mathbb{Z}$ such that $(1 - \epsilon)|Z| \leq |X| + |Y|$ and $\max\{|X|, |Y|\} \leq (1 + 4\sqrt{\epsilon})|Z|/2$, then one of the following holds:

1. There are at least $\epsilon^2|X||Y|$ pairs $(x, y) \in X \times Y$ such that $x + y \notin Z$.
2. There are arithmetic progressions $P, Q$ of length at most $|Z|/2 + 3\sqrt{\epsilon}|Z|$ with the same common difference such that $P$ contains all but at most $\epsilon|X|$ points of $X$, and similarly for $Q$ and $Y$.

If the second property holds, we say that $X$ is $(\epsilon, |Z|)$-close to $P$. 
Applying this stability result to the family $\mathcal{A}$ from our container theorem, we see that every container triple $(X, Y, Z) \in \mathcal{A}$ satisfies one of the following properties.
Applying this stability result to the family $\mathcal{A}$ from our container theorem, we see that every container triple $(X, Y, Z) \in \mathcal{A}$ satisfies one of the following properties.

1. $|X| + |Y| \leq (1 - \varepsilon)Ks$,

2. $\max\{|X|, |Y|\} > (1 + 4\sqrt{\varepsilon})Ks/2$, or

3. $X$ and $Y$ are $(\varepsilon, Ks)$-close to arithmetic progressions with the same common difference.
Proof Outline – Counting the Exceptions (i)

Suppose the container triple \((X, Y, Z) \in \mathcal{A}\) satisfies either

1. \(|X| + |Y| \leq (1 - \epsilon)Ks\), or
2. \(\max\{|X|, |Y|\} > (1 + 4\sqrt{\epsilon})Ks/2\).
Proof Outline – Counting the Exceptions (i)

Suppose the container triple \((X, Y, Z) \in \mathcal{A}\) satisfies either

1. \(|X| + |Y| \leq (1 - \epsilon) Ks\), or
2. \(\max\{|X|, |Y|\} > (1 + 4\sqrt{\epsilon}) Ks / 2\).

Then one can use standard identities and bounds for the binomial coefficient to show that at most

\[ o(1) \binom{Ks/2}{s}^2 \]

pairs of \(s\)-sets \(A \subset X, B \subset Y\) exist.
Finally, suppose the container triple $\langle X, Y, Z \rangle \in \mathcal{A}$ satisfies

- $X$ and $Y$ are $(\epsilon, Ks)$-close to arithmetic progressions $P, Q$ with the same common difference.
Finally, suppose the container triple $(X, Y, Z) \in A$ satisfies

- $X$ and $Y$ are $(\epsilon, Ks)$-close to arithmetic progressions $P, Q$ with the same common difference.

For this case, we exploit the structure of $X$ and $Y$, as well as a general upper bound on the size of their sum to show that there are only $o(1)(Ks^2/2)^2$ pairs of sets $A \subset X$ and $B \subset Y$, both of cardinality $s$, such that one of them is not $(\Omega(\epsilon), Ks)$-close to $P$ or $Q$. 
The Theorem Once More

Using these counting results and the fact that one can take a single pair of disjoint arithmetic progressions $P, Q$ with the same common difference of size $Ks/2$, we arrive again at our main result.
The Theorem Once More

Using these counting results and the fact that one can take a single pair of disjoint arithmetic progressions $P, Q$ with the same common difference of size $Ks/2$, we arrive again at our main result.

**Theorem (Campos, Coulson, Perarnau, Serra, W.)**

Let $s = \Omega((\log n)^3)$ and $K = O(s/(\log)^3)$. Then for almost all sets $A, B \subset [n]$ with $|A| = |B| = s$ and $|A + B| \leq Ks$ there exist arithmetic progressions $P, Q$ with the same common difference of size

$$|P|, |Q| \leq \frac{1 + o(1)}{2} Ks$$

such that $|A \setminus P|, |B \setminus Q| = o(s)$. 

M. Wötzel (UPC & BGSMath)
Open Questions

- What if $A$ and $B$ are not of the same size?
- More than two distinct summands?
- Groups other than $\mathbb{Z}$?
- What about the precise structural version of CCMMS?
Thank you for your attention!