Proceedings of the 12th European Conference on Combinatorics, Graph Theory and Applications
EUROCOMB'23
Prague, August 28 - September 1, 2023

Locality in Sumsets

(Extended abstract)

Peter van Hintum* Peter Keevash †

Abstract

Motivated by the Polynomial Freiman-Ruzsa (PFR) Conjecture, we develop a theory of locality in sumsets, with several applications to John-type approximation and stability of sets with small doubling. One highlight shows that if $A \subset \mathbb{Z}$ with $|A+A| \leq (1-\epsilon)2^d|A|$ is non-degenerate then A is covered by $O(2^d)$ translates of a d-dimensional generalised arithmetic progression (d-GAP) P with $|P| \leq O_{d,\epsilon}(|A|)$; thus we obtain one of the polynomial bounds required by PFR, under the non-degeneracy assumption that A is not efficiently covered by $O_{d,\epsilon}(1)$ translates of a $(d-1)$-GAP.

We also prove a stability result showing for any $\epsilon, \alpha > 0$ that if $A \subset \mathbb{Z}$ with $|A+A| \leq (2-\epsilon)2^d|A|$ is non-degenerate then some $A' \subset A$ with $|A'| > (1-\alpha)|A|$ is efficiently covered by either a $(d+1)$-GAP or $O_\alpha(1)$ translates of a d-GAP. This 'dimension-free' bound for approximate covering makes for a surprising contrast with exact covering, where the required number of translates not only grows with d, but does so exponentially. Another highlight shows that if $A \subset \mathbb{Z}$ is non-degenerate with $|A+A| \leq (2^d+\ell)|A|$ and $\ell \leq 0.1 \cdot 2^d$ then A is covered by $\ell+1$ translates of a d-GAP P with $|P| \leq O_d(|A|)$; this is tight, in that $\ell+1$ cannot be replaced by any smaller number.

The above results also hold for $A \subset \mathbb{R}^d$, replacing GAPs by a suitable common generalisation of GAPs and convex bodies, which we call generalised convex progressions. In this setting the non-degeneracy condition holds automatically, so we obtain essentially optimal bounds with no additional assumption on A. Here we show that if $A \subset \mathbb{R}^k$ satisfies $|\frac{A+A}{2}| \leq (1+\delta)|A|$ with $\delta \in (0,1)$, then $\exists A' \subset A$ with $|A'| \geq (1-\delta)|A|$ so that $|\operatorname{co}(A')| \leq O_{k,1-\delta}(|A|)$. This is a dimensionally independent sharp stability result for the Brunn-Minkowski inequality for equal sets, which hints towards a possible analogue for the Prékopa-Leindler inequality.

*New College, University of Oxford, UK. E-mail: peter.vanhintum@new.ox.ac.uk
†Mathematical Institute, University of Oxford, UK. Supported by ERC Advanced Grant 883810.

These results are all deduced from a unifying theory, in which we introduce a new intrinsic structural approximation of any set, which we call the 'additive hull', and develop its theory via a refinement of Freiman's theorem with additional separation properties. A further application that will be published separately is a proof of Ruzsa's Discrete Brunn-Minkowski Conjecture [vHKT23].

DOI: `https://doi.org/10.5817/CZ.MUNI.EUROCOMB23-079`

1 Introduction

A foundational result in Additive Combinatorics is *Freiman's Theorem* [Fre59] that any subset A of integers with bounded doubling is a dense subset of a generalised arithmetic progression (GAP) P of bounded dimension (see the book of Tao and Vu [TV06] for definitions and background). This gives a satisfactory qualitative description of A: it can be approximated by some P belonging to a simple class of sets with bounded doubling. However, the doubling of P may be much larger than that of A, so the quest for a more quantitative version of Freiman's Theorem has been a major driving force in the development of Additive Combinatorics. A little thought reveals that one must allow P to come from a broader class of sets than just GAPs. One natural attempt is to approximate A by sets $X + P$ with $|X|$ bounded, i.e. the union of a bounded number of translates of P. One might hope to find a such an approximation with polynomial bounds, i.e. if $|A + A| \leq e^{O(d)}|A|$ then we could find such X, P with $\dim(P) = O(d)$, $|X| \leq e^{O(d)}$ and $|X + P| \leq e^{O(d)}|A|$. However, an example of Lovett and Regev [LR17] shows that this is not always possible. Improving the bounds in Freiman's theorem has been the subject of a rich body of research [Ruz94, Bil99, Cha02, GT06, San08, Sch11]. The best known bounds, due to Sanders [San12], have $\dim(P)$ and $\log(|X + P|/|A|)$ about $O(d^6)$.

The *Polynomial Freiman-Ruzsa (PFR) Conjecture* (see [Gre07]) attempts to approximate by translates of a *convex progression*, i.e. a set P of the form $\phi(C \cap \mathbb{Z}^k)$ for some convex set $C \subset \mathbb{R}^k$ and linear map $\phi : \mathbb{Z}^k \to \mathbb{Z}$, for which polynomial bounds may be true. The conjecture states that if $|A + A| \leq e^{O(d)}|A|$ then one can find such P with $k = O(d)$ and $|P| \leq e^{O(d)}|A|$ such that $A \subset X + P$ for some X with $|X| \leq e^{O(d)}$. Below we will describe three perspectives on the PFR Conjecture that provide a thematic overview of our results; these are (1) John-type approximation, (2) Stability, (3) Locality. The third theme of locality is our primary focus, i.e. most of the technical work goes into developing the theory of locality, which is then used to deduce the results discussed within the first two themes. Our results hold both for the discrete setting $A \subset \mathbb{Z}$ considered in PFR and the continuous setting $A \subset \mathbb{R}^k$. For now we will continue to focus on the discrete setting (in some sense the hardest case; we achieve better bounds in the continuous setting).

Our first perspective interprets PFR as a *John-type approximation.* In general terms, a John-type theorem says that any object in some class is approximated efficiently (i.e. up to some constant factor) by some object from some simpler class. Some examples are John's Theorem approximating convex bodies by ellipsoids, Freiman's Theorem approximating sets of small doubling by GAPs, and a theorem of Tao and Vu [TV08] approximating convex progressions by GAPs.

Theme 1: John-type approximation. One question that we address is the existence of John-type approximations P for sets of bounded doubling A. E.g. if $A \subset \mathbb{Z}$ is non-degenerate with $|A+A| \leq (2^d + \ell)|A|$ and $\ell \leq 0.1 \cdot 2^d$ we show that A is covered by $\ell + 1$ translates of a d-GAP P with $|P| \leq O_d(|A|)$. Here $\ell + 1$ translates is optimal (as shown by adding ℓ scattered points to a d-GAP), so in the sense of John-type approximation we have a precise characterisation of such sets A. We also show (see Theorem 1.5) that if $A \subset \mathbb{Z}$ with $|A+A| \leq (1-\epsilon)2^d|A|$ is non-degenerate then A is contained in $O(2^d)$ translates of a d-dimensional convex progression P with $|P| \leq O_{d,\epsilon}(|A|)$; thus we obtain one of the polynomial bounds required by PFR.

Our second perspective sees PFR as a *stability* statement. In general terms, if an object in some class is close to maximising some function on the class, then it must be structurally close to some extremal example. The possible meanings of 'structurally close' are nicely expressed by terminology of Tao: we speak of 1%, 99% or 100% stability according to whether we approximate some constant fraction (1%), all bar some constant fraction (99%), or everything (100%). For sets of small doubling, stability results are only known when the doubling is quite close to the minimum possible, such as the celebrated Freiman $3k-4$ Theorem (see [Fre59]) and various results (described below) for 'non-degenerate' A in $\mathbb{R}^k$ or $\mathbb{Z}^k$ with $|A+A| \leq (2^k + \delta)|A|$ for small δ. Ruzsa's Covering Lemma (see e.g. [Ruz99]) converts any 1% stability theorem into a 100% stability theorem. However, this argument is quantitatively weak, so we require an alternative approach for optimal bounds.

Theme 2: Stability. We approach 100% stability via 99% stability, i.e. we first seek a structural description for almost all of A, and then use it to deduce the remaining structure. This approach is well-known in Extremal Combinatorics (the 'stability method'), but we are not aware of applications to Freiman's Theorem. Our 99% stability result (see Theorem 1.6) shows for any $\epsilon, \alpha > 0$ that if $A \subset \mathbb{Z}$ with $|A+A| \leq (2-\epsilon)2^d|A|$ is non-degenerate then some $A' \subset A$ with $|A'| > (1-\alpha)|A|$ is efficiently covered by either a $(d+1)$-GAP or $O_{\epsilon,\alpha}(1)$ translates of a d-GAP. This 'dimension-free' bound for approximate covering makes for a surprising contrast with exact covering, where the required number of translates not only grows with d, but does so exponentially.

Our third perspective on PFR sees it as describing the *locality* of A. We think of $|X|$ as the number of locations for A, taking the view that elements of the same convex progression are close additively, even though they need not be close metrically. This perspective is particularly clarifying for the continuous setting of $A \subset \mathbb{R}^k$. Here the classical Brunn-Minkowski inequality shows that[1] $|A+A| \geq 2^k|A|$, with equality if and only if A is convex up to a null set. There is a substantial literature on $A \subset \mathbb{R}^k$ with $|A+A| \leq (2^k+\delta)|A|$ for small $\delta > 0$. For such A, Christ [Chr12a] showed that the convex hull $\mathrm{co}(A)$ satisfies $|\mathrm{co}(A)| \leq (1+\epsilon(\delta))|A|$, where $\epsilon(\delta) \to 0$ as $\delta \to 0$. Improvements were obtained by Figalli and Jerison [FJ15, FJ21], recently culminating in a sharp stability result with optimal

[1] For now we use $|\cdot|$ notation for both (discrete) cardinality and (continuous) measure, but for clarity later we use $|\cdot|$ for measure and $\#(\cdot)$ for cardinality.

parameters by van Hintum, Spink and Tiba [vHST22]. Thus for small δ the locality of A is simply its convex hull, but for larger δ the picture becomes more complicated.

Theme 3: Locality. A natural starting point is to consider $A \subset \mathbb{R}^k$ with $|A+A| \leq (2^k+\delta)|A|$ and ask how large $\delta > 0$ may be for us to still efficiently cover A by a convex set. This is clearly impossible for $\delta \geq 1$: consider a convex set and add one faraway point. Our general theory shows that the threshold is exactly at 1, i.e. if $\delta < 1$ then $|\operatorname{co}(A)| < O_k(|A|)$. We have a simple proof for this case, which will be published separately in [vHK23b]. As δ increases we can efficiently cover A by translates of a convex set; moreover, while $\delta < 0.1 \cdot 2^k$ we can do so with $\delta+1$ translates, which is optimal (similarly to the result in $\mathbb{Z}$ mentioned above; both are subsumed in a more general picture). However, when δ reaches 2^k, any fixed number of translates of a convex set will not suffice, as A may be of the form $X+P$ where P is convex and X is an AP.

Generalised convex progressions. Hence, we introduce a common generalisation of convex progressions and GAPs: a *convex (k,d)-progression* is a set $P = \phi(C \cap (\mathbb{R}^k \times \mathbb{Z}^d))$, where $C \subset \mathbb{R}^{k+d}$ is convex and $\phi : \mathbb{R}^{k+d} \to \mathbb{R}^k$ is a linear map. In Theorem 1.4 we consider non-degenerate $A \subset \mathbb{R}^k$ with $|A+A| \leq (2^{k+d}+\ell)|A|$ for $d, \ell \geq 0$ and find an efficient covering of A by $X+P$, where P is a convex (k,d)-progression and $|X|$ is tightly controlled in terms of ℓ; in particular, for $\ell \leq 0.1 \cdot 2^{k+d}$ we obtain the optimal bound $|X| \leq \ell+1$. Setting $d=0$ recovers our results discussed above for $A \subset \mathbb{R}^k$. We will also see that the results in $\mathbb{Z}$ follow from those in $\mathbb{R}$. Thus in our general setting we can still think of $|X|$ as measuring locality, provided that we think of elements of the same generalised convex progression as being close additively, if not metrically.

In the next subsection we introduce notation that is used to formally develop the above concepts and state our precise results during the remainder of this introduction. These are organised by subsection according to our main theme of locality, covering the results discussed above, and the following further results.

- For non-degenerate $A \subset \mathbb{R}^k$ with $|A+A| \leq (2^{k+d}+\delta)|A|$ where $\delta \in (0,1)$ we find a convex (k,d)-progression P with $|P \setminus A| < O_{k,d}(\delta|A|)$ (see Theorem 1.9). Setting $d=0$ recovers the previously mentioned sharp stability result of [vHST22] for $A \subset \mathbb{R}^k$, whereas setting $k=0$ recovers a sharp stability result for non-degenerate $A \subset \mathbb{Z}^d$ by the same authors [vHST23a].
- We obtain a very precise structural description of sets $A \subset \mathbb{R}$ in the line with doubling less than 4 (see Theorem 1.8).

1.1 Overview and notation

The following generalised notion of convex hull will play a crucial role throughout the paper. For $A \subset \mathbb{R}^k$, we write $\operatorname{co}_t^{\mathbb{R}^k,d}(A) = X+P$, where P is a proper convex (k,d)-progression and $\#X \leq t$, choosing X and P so that $A \subset X+P$ and $|X+P|$ is minimal; we fix an arbitrary choice if $X+P$ is not unique. In some cases we will omit k,d if $d=0$ and t if $t=1$, e.g. $\operatorname{co}(X)$ should be understood as $\operatorname{co}_1^{\mathbb{R}^k,0}(X)$, where k is the dimension of the ambient space for X, so that it coincides with the common notion of the convex hull.

We also require the closely related notion $\operatorname{gap}_t^{\mathbb{R}^k,d}(A)$, defined as a minimum volume set $X+P+Q$ containing A such that $\#X \le t$, P is a proper d-GAP and Q is a parallelotope. This is roughly equivalent to the variant $\operatorname{sco}_t^{\mathbb{R}^k,d}(A)$ defined exactly as $\operatorname{co}_t^{\mathbb{R}^k,d}(A)$ but imposing the symmetry requirement $P=-P$. Indeed, $|\operatorname{co}_t^{\mathbb{R}^k,d}(A)| \le |\operatorname{sco}_t^{\mathbb{R}^k,d}(A)| \le |\operatorname{gap}_t^{\mathbb{R}^k,d}(A)|$ is clear, and we will show that $|\operatorname{gap}_t^{\mathbb{R}^k,d}(A)| = O_{d,k}(|\operatorname{sco}_t^{\mathbb{R}^k,d}(A)|)$. Most results in this paper will be stated using $\operatorname{gap}_t^{\mathbb{R}^k,d}$, are equivalent to the corresponding statement using $\operatorname{sco}_t^{\mathbb{R}^k,d}$, and imply the corresponding statement using $\operatorname{co}_t^{\mathbb{R}^k,d}$. However, for some very precise statements we require $\operatorname{co}_t^{\mathbb{R}^k,d}$.

To state our results in $\mathbb{Z}$, let $\operatorname{co}_t^{\mathbb{Z},d}, \operatorname{sco}_t^{\mathbb{Z},d}$, and $\operatorname{gap}_t^{\mathbb{Z},d}$ be the corresponding functions for subsets of $\mathbb{Z}$, replacing 'convex (k,d)-progression' by 'convex d-progression'. To be precise, $\operatorname{co}_t^{\mathbb{Z},d}(A) = X+P$ where P is a convex d-progression and $\#X \le t$, choosing X and P so that $A \subset X+P$ and $\#(X+P)$ minimal. Analogously define $\operatorname{sco}_t^{\mathbb{Z},d}(A)$ and $\operatorname{gap}_t^{\mathbb{Z},d}(A)$ with the additional requirement that P is origin symmetric and a generalised arithmetic progression, respectively. Intuitively, throughout the paper we think of k as 'continuous dimension' and d as 'discrete dimension'. To stress this connection we write $\operatorname{co}_t^{k,d}$ for $\operatorname{co}_t^{\mathbb{R}^k,d}$ and $\operatorname{co}_t^{0,d}$ for $\operatorname{co}_t^{\mathbb{Z},d}$.

As further illustrations of this notation we can restate Freiman's Theorem and PFR as follows.

$$\begin{aligned}
\text{Freiman's Theorem.} \qquad & \text{If } A \subset \mathbb{Z} \text{ with } \#(A+A) \le K\#A, \\
& \qquad \text{then } \#\left(\operatorname{co}_{O_K(1)}^{0,O_K(1)}(A)\right) \le O_K(\#A). \\
\text{PFR.} \qquad & \text{If } A \subset \mathbb{Z} \text{ with } \#(A+A) \le e^{O(d)}\#A, \\
& \qquad \text{then } \#\left(\operatorname{co}_{e^{O(d)}}^{0,O(d)}(A)\right) \le e^{O(d)}\#A.
\end{aligned}$$

Many of our theorems include a non-degeneracy condition which should be interpreted as follows

$$\#\left(\operatorname{gap}_{O_{d,\xi}(1)}^{0,d-1}(A)\right) \ge \Omega_{d,\xi}(\#A) \iff \forall d,\xi, \exists C,c : \#\left(\operatorname{gap}_c^{0,d-1}(A)\right) \ge C\#A, \tag{1}$$

i.e. there is no collection of few (c depending only on d, and ξ) translates of a $d-1$ dimensional generalized arithmetic progression covering A efficiently (exceeding the size of A by at most a factor C depending only on d and ξ).

We will state our results in the following subsections according to the theme of locality, i.e. with respect to the bounds on the parameter t in $\operatorname{co}_t^{k,d}(A)$. A rough summary of their contents is as follows:

- A big part of the set is in one place.
- The entire set is in few places.
- Almost all of the set is in a constant number of places.

- Sets in the line with doubling less than 4 are almost convex.
- A sharp doubling condition for almost convexity.

The theory of sumsets has been developed in several groups. Particular attention has been given to the continuous setting of $\mathbb{R}^k$ (e.g. [FMP09, FMP10, Chr12a, FJ15, FJ17, FJ21, vHST23c, vHST22]) and to the discrete setting of $\mathbb{Z}$ (e.g. [Fre59, Ruz94, Bil99, Cha02, GT06, San08, Sch11]). In the context of this paper, the setting does not make much difference to our results, so although we state some results below for $\mathbb{Z}$, for the proofs we will generally prefer to work in $\mathbb{R}^k$. This is justified as (a) the proofs are the same modulo the theory of the additive hull, and (b) the results in $\mathbb{Z}$ follow from the results in $\mathbb{R}$ via the following proposition.

Proposition 1.1. *For any $A \subset \mathbb{Z}$ and $d, t \in \mathbb{N}$ there is $\epsilon = \epsilon(A, d, t) > 0$ so that $B := A + [-\epsilon, \epsilon] \subset \mathbb{R}$ has $\left|\mathrm{gap}_t^{1,d}(B)\right| = \Theta_{d,t}\left(\epsilon \# \left(\mathrm{gap}_t^{0,d}(A)\right)\right)$.*

1.2 A big part of the set is in one place

We start with the continuous setting of $\mathbb{R}^k$, where we obtain a clean unified formulation of the 1% stability phenomenon to be discussed in this subsection (a big part in one place) and as $\delta \to 0$ the 99% stability phenomenon (for which we will describe sharper results below). Combining 1% stability with Ruzsa covering one obtains 100% stability (for which we will also describe sharper results below). Previous stability results for the Brunn-Minkowski inequality only applied for much smaller δ; in particular, we are not aware of any previous results of this kind where δ does not decrease with the dimension.

Theorem 1.2. *Let $A \subset \mathbb{R}^k$ with $\left|\frac{A+A}{2}\right| \leq (1+\delta)|A|$, where $\delta \in (0,1)$. Then there exists $A' \subset A$ with $|A'| \geq (1 - \min\{\delta, \delta^2(1+O(\delta))\})|A|$ and $|\mathrm{co}(A')| \leq O_{1-\delta,k}(|A|)$.*

Since the size of A' is also independent of the dimension, this result is closely related to the stability question of the Prékopa-Leindler inequality for equal functions, which can be seen as a dimensionally independent version of Brunn-Minkowski. We expand on this connection and formulate a conjectured extension of Theorem 1.2 in [vHK23a, Section 11.1].

Now we consider the integer setting, where the Freiman-Bilu theorem [Bil99] shows that for sets with small doubling a large part of the set is contained in a small GAP of low dimension. A quantitative version of Green and Tao [GT06] shows that for $A \subset \mathbb{Z}$ with $\#(A+A) \leq 2^d(2-\epsilon)\#A$ there exists some $A' \subset A$ with $\#\left(\mathrm{gap}^{0,d}(A')\right) \leq \#A$ and

$$\#A' \geq \exp\left(-O(8^d d^3)\right) \epsilon^{O(2^d)} \#A.$$

With the following result we establish a bound on $\#A'$ that is optimal up to lower order terms, at the cost of a non-degeneracy assumption and relaxing the bound on $\#\,\mathrm{gap}^{0,d}(A')$ in the spirit of our John-type theme. We emphasise that our bound on $\#A'$ does not depend on the dimension.

Theorem 1.3. *Fix $\beta > 0$. Suppose $A \subset \mathbb{Z}$ with $\#(A+A) \leq 2^d(1+\delta)\#A$, where $\delta \in (0,1)$. If $\#\left(\mathrm{gap}^{0,d-1}_{O_{d,\delta,\beta}(1)}(A)\right) \geq \Omega_{d,\delta,\beta}(\#A)$ then there exists $A' \subset A$ with $\#\left(\mathrm{gap}^{0,d}(A')\right) \leq O_{\delta,d}(\#A)$ and*

$$\frac{\#\left(A \setminus A'\right)}{\#A} \leq \min\left\{(1+\beta)\delta, \delta^2 + 60\delta^3\right\}.$$

The above results are sharp up to lower order terms both as $\delta \to 0$ for large k and as $\delta \to 1$. For $\delta \to 0$, consider the union of two homothetic convex sets of volumes δ^2 and $1-\delta^2$; for $\delta \to 1$ consider an arithmetic progression of $\frac{1}{1-\delta}$ equal convex sets. The second example suggests a $k+1$ dimensional convex structure, which we indeed establish in Section 1.4.

1.3 The entire set is in few places

Now we consider the 100% stability problem: what is the maximum 'locality' for given doubling? Our fundamental example is a GAP together with some scattered points, i.e. $A = P \cup S \subset \mathbb{Z}$, where P is a proper d-dimensional GAP and $\#S = \ell$. Then $\#(A+A) \leq (2^d + \ell)\#A$ and A has locality $\ell + 1$.

The following result shows for non-degenerate A that this example is exactly sharp for a large range of ℓ and asymptotically sharp when $2^d - \ell \gg 2^{d/2}$. We remark that even the much weaker bound of $O(2^d)$ for the locality is already sufficient to cover A by $X + P$ with doubling $O(2^{2d})$, i.e. our John-type approximation only loses a square in the doubling, whereas the PFR setting allows any polynomial loss.

Theorem 1.4. *Let $\ell \in (0, 2^d)$ and $A \subset \mathbb{Z}$ with $\#(A+A) \leq (2^d+\ell)\#A$ and $\#\left(\mathrm{gap}^{d-1}_{O_d(1)}(A)\right) \geq \Omega_d(\#A)$. Then $\#\left(\mathrm{gap}^{0,d}_{\ell'}(A)\right) \leq O_{d,\ell}(\#A)$, where*

$$\ell' \leq \begin{cases} \ell + 1 & \text{for } \ell \in \mathbb{N} \text{ if } \ell \leq 0.1 \cdot 2^d, \text{ or if } \ell \leq 0.315 \cdot 2^d \text{ and } d \geq 13, \\ \ell\left(1 + O\left(\sqrt[3]{\frac{2^d}{(2^d-\ell)^2}}\right)\right) & \text{if } 0.1 \cdot 2^d \leq \ell \leq \left(1 - \frac{1}{\sqrt{2^d}}\right)2^d, \\ (1+o(1))\frac{d+1}{2\epsilon} & \text{where } \epsilon = \frac{2^d-\ell}{2^d} \text{ and } o(1) \to 0 \text{ as } \epsilon \to 0. \end{cases}$$

The final bound in Theorem 1.4 gives an asymptotically sharp result for the limiting case $\epsilon = \frac{2^d-\ell}{2^d} \to 0$, i.e. as the doubling approaches 2^{d+1}. Here the above fundamental example breaks down and a new example takes over, which can be thought of as a cone over a GAP; intuitively, this describes the 'most d-dimensional' $(d+1)$-dimensional construction.

The above results are very sharp for non-degenerate sets, but to make further progress towards PFR we need to weaken the non-degeneracy condition. Our next result takes a step in this direction, but its applicability is limited by the double-exponential dependence on d'.

Theorem 1.5. *Let $A \subset \mathbb{Z}$ with $\#(A+A) < 2^d\#A$ and $d' < d$. If $\#\left(\mathrm{gap}^{0,d-d'}_{O_d(1)}(A)\right) \geq \Omega_d(\#A)$ then*

$$\#\left(\mathrm{gap}^{0,d}_{2^d \exp\exp(O(d'))}(A)\right) = O_d(\#A).$$

1.4 Almost all of the set is in a constant number of places

Now we consider 99% stability. Here we find that the number of locations can be bounded by a constant independently of the doubling. This is a remarkable contrast with the 100% stability problem, for which we needed an exponential number of locations to cover the set unless it has close to the minimum possible doubling. Our fundamental example had many scattered points but essentially all of the mass of the set in one location, which hints that one should be able to do much better if one can discard a small part of the set. Furthermore, the second example that takes over as the doubling approaches 2^{d+1} is highly structured so that $\mathrm{co}^{0,d+1}(A)$ is small. The following shows that any non-degenerate set is approximately described by one of these two configurations: it is concentrated in a single $(d+1)$-progression or few d-progressions, where 'few' depends only on the approximation accuracy, not on d.

Theorem 1.6. *For any $\alpha, \epsilon > 0$ and $A \subset \mathbb{Z}$ with $\#(A+A) \leq 2^d(2-\epsilon)\#A$ there is $A' \subset A$ with $\#A' \geq (1-\alpha)\#A$ and*

$$\min\left\{\#\left(\mathrm{gap}^{0,d-1}_{O_d(1)}(A)\right), \#\left(\mathrm{gap}^{0,d}_{O(\alpha^{-2})}(A')\right), \#\left(\mathrm{gap}^{0,d+1}_{1}(A')\right)\right\} \leq O_{d,\epsilon,\alpha}(\#A).$$

Hence, we can always find arithmetic structure in an absolute constant fraction of the set.

Corollary 1.7. *Suppose $A \subset \mathbb{Z}$ with $\#(A+A) \leq 2^d(2-\epsilon)\#A$ and $\#\,\mathrm{gap}^{0,d-1}_{O_d(1)}(A) > \Omega_{d,\epsilon}(\#A)$. Then there is $A' \subset A$ with $\#A' \geq \frac{1}{50000}\#A$ and $\#\left(\mathrm{gap}^{0,d+1}(A')\right) \leq O_{d,\epsilon}(\#A')$.*

1.5 Linear stability results

We additionally prove the following results. For more details and background refer to [vHK23a].

We prove the following extension of Freiman's $3k-4$ theorem, characterizing sets $A \subset \mathbb{R}$ with $|A+A| < 4|A|$. Let $\mathrm{ap}_t(A)$ be a minimum size set containing A that is an AP of t intervals whose lengths are in arithmetic progression.

Theorem 1.8. *There is an absolute constant $C > 0$ such that the following holds. Suppose $A \subset \mathbb{R}$ with $|A+A| < 4|A|$ and $|\mathrm{co}(A)| \geq C|A|$. Let t be minimal so that $|\mathrm{co}_t(A)| < 2|A|$, and let $\delta := \frac{|A+A|}{|A|} - (4-2/t)$. Then $|\mathrm{co}^{1,1}(A) \setminus A| \leq \max\{150, 4t^2\}\delta|A|$. Moreover, if $\delta \leq (2t)^{-2}$ then $|\mathrm{ap}_t(A) \setminus A| \leq 100t\delta|A|$.*

We extend the main results from [vHST22, vHST23a] to establish the optimal bound on the doubling for which a set needs to be approximately convex. An independent proof of the corollary is published separately in [vHK23b].

Theorem 1.9. *For any $d \in \mathbb{N}$, $\gamma, \epsilon > 0$, $\delta \in (0, 1-\epsilon)$, if $A \subset \mathbb{Z}$ with $\#(A+A) \leq (2^d+\delta)\#A$ and $\#\left(\mathrm{gap}^{0,d-1}_{O_{d,\gamma,\epsilon}(1)}(A)\right) = \Omega_d(\#A)$ then $\#(\mathrm{co}^{0,d}(A) \setminus A)/\#(A) \leq O_d(\gamma+\delta)$.*

Corollary 1.10. *If $A \subset \mathbb{R}^k$ satisfies $|A+A| = (2^k+\delta)|A|$ with $\delta < 1$ then $|\mathrm{co}(A) \setminus A| \leq O_k(\delta)|A|$.*

References

[BEKR23] Ben Barber, Joshua Erde, Peter Keevash, and Alexander Roberts. Isoperimetric stability in lattices. *Forthcoming in the Proceedings of the American Mathematical Society*, 2023+.

[Bil98] Yuri Bilu. The $(\alpha + 2\beta)$-inequality on a torus. *Journal of the London Mathematical Society*, 57(3):513–528, 1998.

[Bil99] Yuri Bilu. Structure of sets with small sumset. *Astérisque*, 258(xi):77–108, 1999.

[CDR19] Pablo Candela and Anne De Roton. On sets with small sumset in the circle. *The Quarterly Journal of Mathematics*, 70(1):49–69, 2019.

[CGSG19] Pablo Candela, Diego González-Sánchez, and David Joseph Grynkiewicz. A step towards the $3k-4$ conjecture in $\mathbb{Z}/p\mathbb{Z}$ and an application to m-sum-free sets. *Acta Mathematica Universitatis Comenianae*, 88(3):521–525, 2019.

[Cha02] Mei-Chu Chang. A polynomial bound in Freiman's theorem. *Duke mathematical journal*, 113(3):399–419, 2002.

[Chr12a] Michael Christ. Near equality in the Brunn–Minkowski inequality. *arXiv preprint arXiv:1207.5062*, 2012.

[Chr12b] Michael Christ. Near equality in the two-dimensional Brunn–Minkowski inequality. *arXiv preprint arXiv:1206.1965*, 2012.

[CSS20] Pablo Candela, Oriol Serra, and Christoph Spiegel. A step beyond Freiman's theorem for set addition modulo a prime. *Journal de Théorie des Nombres de Bordeaux*, 32(1):275–289, 2020.

[EGM14] Sean Eberhard, Ben Green, and Freddie Manners. Sets of integers with no large sum-free subset. *Annals of Mathematics*, pages 621–652, 2014.

[FJ15] Alessio Figalli and David Jerison. Quantitative stability for sumsets in $\mathbb{R}^n$. *Journal of the European Mathematical Society*, 17(5):1079–1106, 2015.

[FJ17] Alessio Figalli and David Jerison. Quantitative stability for the Brunn–Minkowski inequality. *Advances in Mathematics*, 314:1–47, 2017.

[FJ21] Alessio Figalli and David Jerison. A sharp Freiman type estimate for semisums in two and three dimensional euclidean spaces. *Annales Scientifiques de l'École Normale Supérieure*, 54(4):235–257, 2021.

[FMP09] Alessio Figalli, Francesco Maggi, and Aldo Pratelli. A refined Brunn–Minkowski inequality for convex sets. In *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, volume 26, pages 2511–2519. Elsevier, 2009.

[FMP10] Alessio Figalli, Francesco Maggi, and Aldo Pratelli. A mass transportation approach to quantitative isoperimetric inequalities. *Inventiones mathematicae*, 182(1):167–211, 2010.

[Fre59] Gregory Freiman. The addition of finite sets. i. *Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika*, (6):202–213, 1959.

[GR06] Ben Green and Imre Ruzsa. Sets with small sumset and rectification. *Bulletin of the London Mathematical Society*, 38(1):43–52, 2006.

[Gre07] Ben Green. The polynomial Freıman–Ruzsa conjecture. *Terence Tao's blog*, 2007.

[GT06] Ben Green and Terence Tao. Compressions, convex geometry and the Freiman–Bilu theorem. *Quarterly Journal of Mathematics*, 57(4):495–504, 2006.

[vHK23a] Peter van Hintum and Peter Keevash. Locality in sumsets. *arXiv preprint arXiv:2304.01189*, 2023.

[vHK23b] Peter van Hintum and Peter Keevash. The sharp doubling threshold for approximate convexity. *arXiv preprint arXiv:2304.01176*, 2023.

[vHKT23] Peter van Hintum, Peter Keevash, and Marius Tiba. On Ruzsa's discrete Brunn-Minkowski conjecture. *Manuscript*, 2023.

[vHST22] Peter van Hintum, Hunter Spink, and Marius Tiba. Sharp stability of Brunn–Minkowski for homothetic regions. *Journal of the European Mathematical Society*, 24(12):4207–4223, 2022.

[vHST23a] Peter van Hintum, Hunter Spink, and Marius Tiba. Sets in $\mathbb{Z}^k$ with doubling $2^k + \delta$ are near convex progressions. *Advances in Mathematics*, 413:108830, 2023.

[vHST23b] Peter van Hintum, Hunter Spink, and Marius Tiba. Sharp l^1 inequalities for sup-convolution. *Forthcoming in Discrete Analysis*, 2023+.

[vHST23c] Peter van Hintum, Hunter Spink, and Marius Tiba. Sharp quantitative stability of the planar Brunn–Minkowski inequality. *Forthcoming in Journal of the European Mathematical Society*, 2023+.

[Jin07] Renling Jin. Freiman's inverse problem with small doubling property. *Advances in Mathematics*, 216(2):711–752, 2007.

[Kee11] Peter Keevash. Hypergraph turan problems. *Surveys in combinatorics*, 392:83–140, 2011.

[LR17] Shachar Lovett and Oded Regev. A counterexample to a strong variant of the polynomial Freiman-Ruzsa conjecture in euclidean space. *Discrete Analysis*, 8:379–388, 2017.

[LS23] Vsevolod F Lev and Oriol Serra. Towards $3n-4$ in groups of prime order. *arXiv preprint arXiv:2302.08465*, 2023.

[Mag08] Francesco Maggi. Some methods for studying stability in isoperimetric type problems. *Bulletin of the American Mathematical Society*, 45(3):367–408, 2008.

[MFY74] DA Moskvin, Gregory Freiman, and Aleksandr Aleksandrovich Yudin. Structure theory of set addition and local limit theorems for independent lattice random variables. *Theory of Probability & Its Applications*, 19(1):53–64, 1974.

[Ruz94] Imre Ruzsa. Generalized arithmetical progressions and sumsets. *Acta Mathematica Hungarica*, 8(4):1–6, 1994.

[Ruz99] Imre Ruzsa. An analog of Freiman's theorem in groups. *Astérisque*, 258(199):323–326, 1999.

[Ruz09] Imre Ruzsa. Sumsets and structure. *Combinatorial number theory and additive group theory*, pages 87–210, 2009.

[San08] Tom Sanders. Appendix to 'Roth's theorem on progressions revisited' by J. Bourgain. *Journal d'Analyse Mathématique*, 1(104):193–206, 2008.

[San12] Tom Sanders. On the Bogolyubov–Ruzsa lemma. *Analysis & PDE*, 5(3):627–655, 2012.

[Sch11] Tomasz Schoen. Near optimal bounds in Freiman's theorem. *Duke Mathematical Journal*, 158(1):1–12, 2011.

[Tao18] Terence Tao. An inverse theorem for an inequality of Kneser. *Proceedings of the Steklov Institute of Mathematics*, 303(1):193–219, 2018.

[TV06] Terence Tao and Van Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.

[TV08] Terence Tao and Van Vu. John-type theorems for generalized arithmetic progressions and iterated sumsets. *Advances in Mathematics*, 219(2):428–449, 2008.