An Approximate Counting Version of the Multidimensional Szemerédi Theorem*

Natalie Behague†1, Joseph Hyde‡2, Natasha Morrison§2, Jonathan A. Noel¶2, and Ashna Wright‖2

[1]Mathematics Institute, University of Warwick, UK
[2]Department of Mathematics and Statistics, University of Victoria, Canada

Abstract

Given a finite set $X \subseteq \mathbb{N}^d$, a *non-trivial copy* of X is a set obtained by scaling X by a positive factor and translating it. The Multidimensional Szemerédi Theorem of Furstenberg and Katznelson [6] asserts that the largest cardinality of a subset of $[n]^d$ without a non-trivial copy of X, denoted $r_X(n)$, is $o(n^d)$. We prove that, for any X with $|X| \geq 3$, there exists $C_X > 1$ such that the number of subsets of $[n]^d$ without a non-trivial copy of X is at most $2^{C_X \cdot r_X(n)}$ for infinitely many n.

1 Introduction

Roth's Theorem [10] famously states that every subset of $[n] := \{1, \ldots, n\}$ without three elements in arithmetic progression has cardinality at most $o(n)$. This was extended to arithmetic progressions of arbitrary length in the groundbreaking work of Szemerédi [12]. Szemerédi's Theorem can be seen as a very strong "density version" of the elementary van der Waerden Theorem [13] which says that every colouring of the natural numbers with finitely many colours contains monochromatic arithmetic progressions of arbitrary length.

A few years later, Furstenberg [5] reproved Szemerédi's Theorem using tools from ergodic theory. This new perspective turned out to be widely applicable, yielding several remarkably general results. One such example is the so-called "Multidimensional Szemerédi Theorem" of Furstenberg and Katznelson [6], which we explain next. Given a set $X \subseteq \mathbb{N}^d$, a *copy* of X is a set of the form

$$\vec{b} + r \cdot X = \{\vec{b} + r\vec{x} : \vec{x} \in X\}$$

where $\vec{b} \in \mathbb{R}^d$ and $r \geq 0$. It is said to be a *non-trivial* copy if $r > 0$. As an example, a 3-term arithmetic progression is nothing more than a copy of $X = \{1, 2, 3\}$. A set $A \subseteq \mathbb{N}^d$ is X-*free* if it does not contain a copy of X and $r_X(n)$ denotes the cardinality of the largest X-free subset of $[n]^d$.

Theorem 1 (Multidimensional Szemerédi Theorem [6]). *For any finite set $X \subseteq \mathbb{N}^d$, $r_X(n) = o(n^d)$.*

We focus on the closely related problem of counting the number of X-free subsets of $[n]^d$. An obvious lower bound is $2^{r_X(n)}$, which one can get by taking any subset of the largest X-free set. Our main result says that the exponent $r_X(n)$ is within a constant factor of being tight for infinitely many $n \in \mathbb{N}$.

*The full version of this work can be found in [3] and will be published elsewhere.

†Email: natalie.behague@warwick.ac.uk. Research of N. Behague supported by PIMS
‡Email: josephhyde@uvic.ca.
§Email: nmorrison@uvic.ca. Research of N. Morrison supported by NSERC and a university start-up grant.
¶Email: noelj@uvic.ca. Research of J. A. Noel supported by NSERC and a university start-up grant.
‖Email: ashnawright@uvic.ca. Research of A. Wright supported by NSERC.

Theorem 2. *For any finite set $X \subseteq \mathbb{N}^d$ with $|X| \geq 3$ there exists $C_X > 1$ such that the number of X-free subsets of $[n]^d$ is at most $2^{C_X \cdot r_X(n)}$ for infinitely many n.*

We note that Theorem 2 extends the work of Balogh, Liu and Sharifzadeh [1] who proved it for arithmetic progressions and Kim [9] who focused on the case that $X = \{\vec{0}, \vec{e}_1, \ldots, \vec{e}_d\}$ where $\vec{e}_i$ is the ith standard basis vector of $\mathbb{R}^d$.

2 Key Ideas and Challenges

To anyone who has followed recent developments on obtaining "counting versions" of important theorems from extremal combinatorics, it should be no surprise that our proof is an application of the container method developed by Saxton and Thomason [11] and Balogh, Morris and Samotij [2]. The container method provides a widely applicable "recipe" for bounding the number of independent sets in a "well-behaved" hypergraph from above.

In our setting, the choice of the hypergraph is straightforward; we let $\mathcal{H}$ be the hypergraph with vertex set $[n]^d$ in which every non-trivial copy of X forms a hyperedge. An independent set in $\mathcal{H}$ clearly corresponds to an X-free set, and so our goal is to bound the number of independent sets in $\mathcal{H}$. Note that every vertex of $\mathcal{H}$ is contained within $\Theta(n)$ hyperedges and that any pair of distinct vertices of $\mathcal{H}$ are only contained within $O(1)$ hyperedges together. In other words, the hypergraph $\mathcal{H}$ has very small "co-degrees." As it turns out, this implies that it is sufficiently "well behaved" to apply the method.

The key ingredient in most applications of the container method is a so-called "supersaturation bound." In the context of our problem, a supersaturation theorem is a lower bound on the number of copies of X within a subset of $[n]^d$ of cardinality greater than $r_X(n)$. Roughly speaking, our key supersaturation bound is as follows.

Lemma 3 (Supersaturation Bound, Roughly)**.** *Let $X \subseteq \mathbb{N}^d$ with $|X| \geq 3$. There exists $C_X > 1$ such that, for infinitely many $n \in \mathbb{N}$, every set $A \subseteq [n]^d$ with $|A| \geq C_X \cdot r_X(n)$ contains a "large" number of copies of X.*

Given a sufficiently strong supersaturation bound of the type described in Lemma 3, Theorem 2 can be deduced from many different forms of the Hypergraph Container Lemma; in particular, we apply a version from Saxton and Thomason [11]. Most of the work in the paper is devoted to proving Lemma 3.

One of the key challenges in establishing a supersaturation bound that applies to all sets of cardinality $C_X \cdot r_X(n)$ for some constant C_X is that the function $r_X(n)$ itself is not well-understood. Letting $r_k(n) := r_{\{1,\ldots,k\}}(n)$, the best known bounds on $r_3(n)$ (i.e. the case of 3-term arithmetic progressions) are currently

$$\frac{n}{2^{c\sqrt{\log_2(n)}}} \leq r_3(n) \leq \frac{n}{2^{\log(n)^\beta}} \tag{1}$$

where $0 < c < 2\sqrt{2}$ and $\beta > 0$ are constants. Both inequalities were proved recently by Hunter [7] (lower bound) and Kelly and Meka [8] (upper bound) and represent substantial breakthroughs in the field. In spite of these achievements, we still do not know $r_3(n)$ to within a constant factor, nor do we have precise asymptotics for $r_X(n)$ for other sets X of interest. Thus, we will need to find a way of proving a supersaturation result for sets of cardinality $C_X \cdot r_X(n)$ which does not require us to know any detailed information about the growth rate of $r_X(n)$.

To obtain the desired supersaturation result, a key idea from [1] is to apply a "crude" supersaturation bound which we "amplify" via a double-counting trick. For the crude bound, if we let $\Gamma_X(A)$ denote the number of copies of X in a set $A \subseteq [n]^d$, then

$$\Gamma_X(A) \geq |A| - r_X(n).$$

This is trivial to see by simply deleting an element of A within a copy of X (if one exists) and applying induction. This trivial idea becomes more powerful if one applies it within a random "subcube" of $[n]^d$.

To be a bit less vague, suppose that $M \ll n$ and take a random copy of $[M]^d$ within $[n]^d$ by scaling $[M]^d$ by a random prime p and translating it by a random vector $\vec{b}$ (where p and $\vec{b}$ are chosen with some constraints to make this copy a subset of $[n]^d$). Now, given any set $A \subseteq [n]^d$, we can use the crude bound to get that the number of copies of X contained within this copy of $[M]^d$ is at least the number of points of A within this subcube minus $r_X(M)$. We can also bound this quantity above in terms of $\Gamma_X(A)$; putting these two bounds together yields a lower bound on $\Gamma_X(A)$.

Unfortunately, as it turns out, the argument in the previous paragraph provides a lower bound on $\Gamma_X(A)$ in terms of $r_X(M)$, where $M = M(n)$ is a sublinear function of n. In fact, the bound obtained is a multiple of

$$\frac{|A|}{2n^d} - \frac{r_X(M)}{M^d}$$

which is clearly useless in cases where $\frac{|A|}{2n^d} \leq \frac{r_X(M)}{M^d}$. Recall that we need the supersaturation bound to work for any set A such that $|A| \geq C_X \cdot r_X(n)$ for some constant $C_X > 1$. Thus, in order for the bound to be useful, we need that $\frac{C_X \cdot r_X(n)}{2n^d}$ is significantly larger than $\frac{r_X(M)}{M^d}$. Again, we run into the same problem: we do not understand the growth rate of $r_X(n)$. In particular, if the ratio $\frac{r_X(n)}{n^d}$ fluctuates wildly as n tends to infinity, then we have no hope of obtaining a bound of this form that holds for all n. However, by combining the lower bound in (1) (in fact, a weaker bound of Behrend [4] from 1946 suffices) generalized to $r_X(n)$ with a simple limit argument based on the fact that $\sqrt{N + \sqrt{N}} - \sqrt{N}$ converges, we can get that such a bound holds for infinitely many n, allowing us to obtain Lemma 3.

References

[1] J. Balogh, H. Liu, and M. Sharifzadeh, *The number of subsets of integers with no k-term arithmetic progression*, Int. Math. Res. Not. IMRN (2017), no. 20, 6168–6186.

[2] J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, J. Amer. Math. Soc. **28** (2015), no. 3, 669–709.

[3] N. Behague, J. Hyde, N. Morrison, J. A. Noel, and A. Wright, *An approximate counting version of the multidimensional Szemerédi theorem*, E-print arXiv:2311.13709v1, 2023.

[4] F. A. Behrend, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U.S.A. **32** (1946), 331–332.

[5] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.

[6] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291 (1979).

[7] Z. Hunter, *New lower bounds for $r_3(N)$*, E-print arXiv:2401.16106v2, 2024.

[8] Z. Kelley and R. Meka, *Strong bounds for 3-progressions*, E-print arXiv:2302.05537v4, 2023.

[9] Y. Kim, *The number of k-dimensional corner-free subsets of grids*, Electron. J. Combin. **29** (2022), no. 2, Paper No. 2.53, 19.

[10] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109.

[11] D. Saxton and A. Thomason, *Hypergraph containers*, Invent. Math. **201** (2015), no. 3, 925–992.

[12] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

[13] B. van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw Arch. Wisk. **19** (1927), 212–216.