A canonical van der Waerden theorem in random sets

José D. Alvarado*[1], Yoshiharu Kohayakawa†[1], Patrick Morris‡[2], Guilherme O. Mota§[1], and Miquel Ortega¶[2]

[1]Instituto de Matemática e Estatística , Universidade de São Paulo, Rua do Matão 1010, 05508–090 São Paulo, Brazil
[2]Departament de Matemàtiques, Universitat Politècnica de Catalunya (UPC), Carrer de Pau Gargallo 14, 08028 Barcelona, Spain

Abstract

The canonical van der Waerden theorem states that, for large enough n, any colouring of $[n]$ gives rise to monochromatic or rainbow k-APs. In this work, we are interested in sparse random versions of this result. More concretely, we determine the threshold at which the binomial random set $[n]_p$ inherits the canonical van der Waerden properties of $[n]$.

1 Introduction

Arithmetic Ramsey theory is a branch of combinatorics that studies what sort of arithmetic structure must appear in all possible – although frequently restricted to finite – colourings of the integers. One of the first and most celebrated results in the field is van der Waerden's theorem [23], which states that any finite colouring of the integers contains monochromatic arithmetic progressions of arbitrary length.

There are several paths to generalizing van der Waerden's theorem, which also provide a better understanding of the underlying phenomenon responsible for the appearance of such arithmetic structure. A first option consists in studying when can one guarantee the existence of other objects besides arithmetic progressions. An example of such a generalization is the work of Rado [15], who studied the case of general linear structures. In fact, he was able to characterize precisely those homogeneous linear systems of equations that admit monochromatic solutions in any finite colouring of the integers (for more details see, for example, [9]).

A second possibility for generalizing van der Waerden's theorem consists in dropping the restriction on the number of colours, and studying what kind of structure one may still find for all possible colourings, using a finite numbers of colours or not. It turns out that, in the case of arithmetic progressions, the needed piece of structure to complete the puzzle is that of *rainbow* progressions, namely, arithmetic progressions where every element has a different colour. This gives rise to the canonical van der Waerden theorem, first proved by Erdős and Graham [6], where an arithmetic progression which is monochromatic or rainbow is called *canonical*.

Theorem 1 (Canonical van der Waerden)**.** *For any integer $k \geq 1$, there exists large enough n such that the following holds. Any colouring of $[n] = \{1, \dots, n\}$ contains a canonical arithmetic progression of length k.*

*Email: josealvarado.mat17@ime.usp.br. Partially supported by FAPESP (2020/10796-0)
†Email: yoshi@ime.usp.br. Partially supported by CNPq (406248/2021-4, 407970/2023-1, 315258/2023-3).
‡Email: pmorrismaths@gmail.com. Partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Walter Benjamin program - project number 504502205.
§Email: mota@ime.usp.br. Partially supported by (315916/2023-0, 406248/2021-4) and FAPESP (2018/04876-1).
¶Email: miquel.ortega.sanchez-colomer@upc.edu. Partially supported by FPI PRE2021-099120

Note that Theorem 1 clearly implies that any colouring of the integers (possibly with an infinite numbers of colours) contains arbitrarily long canonical arithmetic progressions. A well-known compactness argument, with a nice additional idea, can be applied to prove that this "infinite" version in fact does imply the "finite" version in the theorem above (see, e.g., [7, p. 29]).

A final path to generalizing van der Waerden's theorem is changing the ambient set where the theorem holds. Instead of colouring the integers, one might consider looking for monochromatic arithmetic progressions in colourings of some particular subset of the integers. For example, a common way to do this is proving an analogue of the theorem in random sets (see, for example, [19]). The theorem also holds when the ambient set is substituted by a set that is pseudorandom enough, meaning, in very informal terms, that it has statistical properties similar to those of a random set (see [5] and the references therein for an in-depth discussion).

In fact, the focus of this work will be a sparse random version of Theorem 1. Consider the random set $[n]_p$, where every element of $[n]$ is sampled independently with probability p. We study how large must p be for $[n]_p$ to satisfy an analogue of the canonical van der Waerden theorem. More formally, a threshold for a monotone property $\mathcal{P}$ is a function $p^* = p^*(n)$ such that

$$\lim_{n\to\infty} P([n]_p \in \mathcal{P}) = \begin{cases} 0 \text{ if } p \leq cp^* \\ 1 \text{ if } p \geq Cp^* \end{cases}$$

for constants $C, c > 0$. The case of the statement that guarantees $[n]_p \notin \mathcal{P}$ with high probability is referred to as the 0-*statement* and the other one as the 1-*statement*.

For example, Rödl and Ruciński [18,19] established such a threshold for the van der Waerden property (and, more generally, a Rado type theorem). To state it precisely, let us say a set is (r,k)-*van der Waerden* if any r-colouring contains a monochromatic arithmetic progression of length k. Their result, for the case of arithmetic progressions, reads as follows.

Theorem 2 (Sparse van der Waerden). *Given integers $k \geq 3$ and $r \geq 2$, there exist constants $C, c > 0$ such that:*

- *For $p > Cn^{-1/(k-1)}$, the random set $[n]_p$ is a.a.s. (r,k)-van der Waerden.*
- *For $p < cn^{-1/(k-1)}$, the random set $[n]_p$ is a.a.s. not (r,k)-van der Waerden.*

See also [8] for further discussion of Theorem 2, where sharpness of the threshold is proved.

The main contribution of this work consists in proving the same kind of statement for the canonical van der Waerden theorem. Indeed, let us say a set is *canonically k-van der Waerden* if every colouring contains a canonical arithmetic progression of length k. We prove the following.

Theorem 3 (Sparse canonical van der Waerden). *Given a natural number $k \geq 3$, there exist constants $C, c > 0$ such such that:*

- *For $p > Cn^{-1/(k-1)}$, the random set $[n]_p$ is a.a.s. canonically k-van der Waerden.*
- *For $p < cn^{-1/(k-1)}$, the random set $[n]_p$ is a.a.s. not canonically k-van der Waerden.*

Note that the 0-statement follows from the corresponding 0-statement in Theorem 2 with $r = 2$, since a rainbow arithmetic progression of length k cannot be formed with only two colours.

Results such as this one, where a known theorem over discrete ambient sets is translated to sparser settings, and particularly to sparse random subsets, have become a common theme in modern combinatorics in the last decades. In the case of Ramsey's theorem, this was first carried out in a seminal series of papers by Rödl and Ruciński [16–18], and both the 0-statement and the 1-statement had delicate and involved proofs. The 1-statement was later reproved with a short and elegant argument by Nenadov and Steger [14], using the method of hypergraph containers [3]. The ideas of Nenadov and Steger have also been used in arithmetic Ramsey theory, for example, to give short proofs of the 1-statement in sparse random versions of Rado's theorem (see [10] or [22]). Our proof uses the ideas of [14] and the method of hypergraph containers, although a straightforward application of their methods is not possible because their argument relies on the boundedness of the number of colours.

Sparse random versions of canonical Ramsey theorems are much more recent. In these, one must somehow overcome the difficulty of having a possibly unbounded number of colours in a given colouring. In a breakthrough result, Kamčev and Schacht [11] have proven a sparse analogue of the canonical Ramsey theorem for cliques, using the transference principle of Conlon and Gowers [4]. In independent work, a subset of the authors [1] prove a canonical Ramsey theorem when the colourings are constrained by some prefixed lists of colours. This is then one of of the ingredients to establish a canonical Ramsey theorem for even cycles [2]. These results use a combination of ideas from the method of containers and the work of Rödl and Ruciński [19].

The current work is inspired in the previous work of [1, 2] and [11], and aims to prove an analogous result in the arithmetic setting. It turns out that, when looking for arithmetic progressions, the situation for canonical theorems is different from theirs, since we only look for monochromatic or rainbow copies of arithmetic progressions, whereas when dealing with graphs, one might also find *lexicographic* copies of graphs (see [1] or [11] for more details). In the course of proving our result, we use a new set of ideas which allow for a streamlined proof in the arithmetic setting.

2 Sketch of the proof

In this section we give a rough sketch of the proof of Theorem 3 and some of the ideas involved in it. As we noted before, we concentrate only on the proof of the 1-statement.

The proof of the 1-statement starts out with a basic dichotomy over a given colouring of $[n]_p$. If such a colouring has a colour with positive density, we are able to apply the sparse version of Szemerédi's theorem (see Theorem 5 below) to find a monochromatic arithmetic progression of length k or k-AP for short). If, on the other hand, all colours are sparse, it turns out that we may find a rainbow k-AP. In order to split according to this criterion, we introduce bounded colourings.

Definition 4. *An r-colouring $\chi\colon A \to [r]$ of a set $A \subset \mathbb{N}$ is α-bounded for $\alpha > 0$ if $|\chi^{-1}(i)| \leq \alpha|A|$ for all $i \in [r]$, that is, all colours have density at most α in A.*

2.1 The dense colour case

We say a set A is (δ, k)-*Szemerédi* if every subset of A of size greater than $\delta|A|$ contains a k-AP. The sparse random version of Szemerédi's theorem, proven by Conlon and Gowers [4] and Schacht [21] independently, establishes the threshold where $[n]_p$ satisfies this condition.

Theorem 5 (Szemerédi's theorem for sparse random sets)**.** *Given $\delta > 0$ and a natural number $k \geq 3$, there exists a constant C such that, for $p > Cn^{-1/(k-1)}$, the random set $[n]_p$ is a.a.s. (δ, k)-Szemerédi.*

For our purposes, this can be rephrased in terms of bounded colourings.

Corollary 6. *Given $\alpha > 0$ and a natural number $k \geq 3$, there exists $C = C(\alpha, k)$ such that the set $[n]_p$ a.a.s. satisfies the following property for $p > Cn^{-1/(k-1)}$. Every colouring of $[n]_p$ that is not α-bounded contains a monochromatic k-AP.*

2.2 Searching for rainbows

On account of the previous observation, it suffices to establish the following to prove Theorem 3.

Proposition 7. *Given $k > 0$, there exist $C = C(k)$ and $\alpha = \alpha(k)$ such that the set $[n]_p$ a.a.s. satisfies the following property for $p > Cn^{-1/(k-1)}$. Every α-bounded colouring of $[n]_p$ contains a rainbow k-AP.*

Once picking a suitable constant α, successively merging the smallest colours, we may reduce the proof of Proposition 7 to the case of α-bounded colourings with at most $r = \lceil 2/\alpha \rceil$ colours. This leaves us in a better position, since now the number of colours is bounded in terms of α and we may use container type arguments.

2.2.1 A container theorem

In very loose terms, the hypergraph container theorem [3, 20] is a way to cluster independent sets of a sufficiently regular hypergraph. This allows one to control the probability of lying in one of these clusters when a simple union bound over all possible independent sets would be too large. Still in somewhat vague terms, it gives, for every sufficiently regular hypergraph $\mathcal{H}$, a collection of *containers* $\mathcal{C} \subset \mathcal{P}(V(\mathcal{H}))$ that satisfy the following properties:

- The containers are almost independent. They contain few edges of $\mathcal{H}$, usually in the sense that $e(\mathcal{H}[C]) < \varepsilon e(\mathcal{H})$ for $\varepsilon > 0$ as small as needed and every $C \in \mathcal{C}$.
- Every independent set in $\mathcal{H}$ is contained in one of the containers.
- There are few containers. More precisely, every independent set has a small subset (its *fingerprint*) that is uniquely associated to a container. The number of containers may be bounded by the total amount of small subsets.

In order to prove Proposition 7 we just look for rainbow k-APs, so we apply the container theorem to the rainbow copy hypergraph $\mathcal{H} = \mathcal{H}(n, k, r)$ with vertex set consisting of r copies of $[n]$, one for every possible colour, and edge set formed by all possible rainbow k-APs. More formally, $\mathcal{H}$ is the k-uniform hypergraph with vertex set $V(\mathcal{H}) = [n] \times [r]$, and edge set

$$E(\mathcal{H}) = \left\{ \{(n_1, c_1), \ldots, (n_k, c_k)\} \in \binom{V(\mathcal{H})}{k} : (n_1, \ldots, n_k) \text{ forms a } k\text{-AP and } c_i \neq c_j \, \forall i \neq j \right\}.$$

It is useful to identify subsets of $V(\mathcal{H})$ with the following notion of an $[r]$-coloured set.

Definition 8. *An $[r]$-colouring of a set $A \subset [n]$ is a function $\chi\colon A \to \mathcal{P}([r])$. Such a pair (A, χ) forms an $[r]$-coloured set. A subcolouring (A', χ') of (A, χ) is a colouring such that $A' \subset A$ and $\chi'(i) \subset \chi(i)$ for all $i \in A'$.*

Indeed, an $[r]$-coloured subset (A, χ) with $A \subset [n]$ can also be thought of as a subset of $[n] \times [r] = V(\mathcal{H})$ in a natural way, and we write $A_\chi \subset V(\mathcal{H})$ for such a set. An application of the hypergraph container theorem then gives the following (for a very similar application of the container method, see [12, 13]).

Theorem 9. *For any $k \in \mathbb{N}$ and $\varepsilon > 0$, there exists a constant $C = C(k, \varepsilon) > 0$, a collection $\mathcal{A}$ of $[r]$-coloured sets, and a function $f\colon \mathcal{P}([n])^r \to \mathcal{A}$ satisfying:*

- *Every $A_\chi \in \mathcal{A}$ satisfies $e(\mathcal{H}[A_\chi]) < \varepsilon n^2$, that is, there are less than εn^2 rainbow k-APs compatible with the $[r]$-coloured set A_χ.*
- *For every $[r]$-coloured set I_χ with no rainbow k-AP, there exists a subcolouring $S_\psi \subset I_\chi$ such that*

$$|S| \leq C n^{1-1/(k-1)} \quad \text{and} \quad I_\chi \subset f(S_\psi).$$

2.2.2 A supersaturation result

In order to use Theorem 9, we must study what we can deduce about our containers from the fact that they admit few rainbow k-APs. In most applications of the hypergraph container theorem this is stated as a supersaturation result, which gives conditions that guarantee the existence of many solutions. Here, we state it in the contrapositive form, which turns out to be slightly more comfortable for this application. We obtain the following.

Proposition 10. *Given k, there exists $M = M(k)$ such that the following holds. For every r, there is an $\varepsilon = \varepsilon(k, r) > 0$ such that any $[r]$-coloured subset (χ, A) satisfying $|A| \geq n/2$ and $e(\mathcal{H}[A_\chi]) < \varepsilon n^2$ admits a subset of colours $L \subset [r]$ with $|L| \leq M$ and a subset $B \subset A$ with*

$$|B| \geq n/8 \quad \textit{and} \quad \chi(B) \subset L, \tag{1}$$

that is, one can find a large number of values in A which only use colours in L, and the size of L is bounded independently of the number of colours r.

We remark that the crucial part of Proposition 10 is that the bound M on $|L|$ does not depend on the total number of colours r, but only on k. The proof of this result starts out by noticing that the set

$$C = \{x \in A \colon |\chi(x)| \geq k\}$$

is small, say $|C| < n/4$. Otherwise, by a supersaturated version of Szemerédi's theorem, we find many k-APs in C, and, since every element of C has at least k different colours to choose from, every k-AP in C gives rise to at least one rainbow k-AP.

Let $A' = A \setminus C$. By somewhat more delicate counting arguments, one can prove that, for most values of $x \in A'$, the list $\chi(x)$ is made up uniquely of colours that appear at least in βn other values of A', for some $\beta = \beta(k) > 0$. These colours make up our list L, and the bound on its size follows from the fact that there are at most $|A'|k$ pairs of value and colour, so that

$$|L| \leq \frac{|A'|k}{\beta n} \leq \frac{k}{\beta}.$$

2.2.3 Putting it together

Proposition 10 implies, for ε small enough, that each container obtained in Theorem 9 is either small ($|A| \leq n/2$), or there exists $B \subset A$ satisfying (1). From Theorem 9 it follows that, if the set $[n]_p$ can be coloured with an α-bounded colouring with no rainbow k-AP, then $[n]_p \subset A$ for some $A_\chi \in \mathcal{A}$. If $|A| \leq n/2$, it is exponentially unlikely that $[n]_p \subset A$, so we focus on the latter case. In fact, we expect $[n]_p$ to have size close to np and intersect the corresponding B in about $|B|p \geq np/8$ positions. Assuming such estimates would leave us in a very good position, since any colouring of $[n]_p$ compatible with B would have a colour of density at least

$$\frac{|B|p}{npM} \geq \frac{1}{8M},$$

which, setting $\alpha \ll 1/8M$, would contradict the α-boundedness of the colouring. Thus, the conclusion of Proposition 7 fails only when $[n]_p$ or its intersection with B have large deviations from their expected value. Using Chernoff bounds to obtain exponential concentration and some standard estimates involving a union bound over all possible containers gives Proposition 7.

References

[1] J. D. Alvarado, Y. Kohayakawa, P. Morris, and G. O. Mota, *A canonical Ramsey theorem with list constraints in random graphs*, Procedia Computer Science **223** (2023), 13–19, XII Latin-American Algorithms, Graphs and Optimization Symposium (LAGOS 2023).

[2] ———, *A canonical Ramsey theorem for even cycles in random graphs*, manuscript (2024).

[3] J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, J. Amer. Math. Soc. **28** (2015), no. 3, 669–709. MR 3327533

[4] D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, Ann. of Math. (2) **184** (2016), no. 2, 367–454. MR 3548529

[5] D. Conlon, J. Fox, and Y. Zhao, *Extremal results in sparse pseudorandom graphs*, Adv. Math. **256** (2014), 206–290. MR 3177293

[6] P. Erdős and R. L. Graham, *Old and new problems and results in combinatorial number theory*, Monographies de L'Enseignement Mathématique [Monographs of L'Enseignement Mathématique], vol. 28, Université de Genève, L'Enseignement Mathématique, Geneva, 1980. MR 592420

[7] S. Farhangi, *On refinements of van der Waerden's theorem*, Master's thesis, Virginia Polytechnic Institute and State University, 2016.

[8] E. Friedgut, H. Hàn, Y. Person, and M. Schacht, *A sharp threshold for van der Waerden's theorem in random subsets*, Discrete Anal. (2016), Paper No. 7, 20pp. MR 3533306

[9] R. L. Graham, B. L. Rothschild, and J. H. Spencer, *Ramsey theory*, Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons, Inc., New York, 1980, A Wiley-Interscience Publication. MR 591457

[10] R. Hancock, K. Staden, and A. Treglown, *Independent sets in hypergraphs and Ramsey properties of graphs and the integers*, SIAM J. Discrete Math. **33** (2019), no. 1, 153–188. MR 3899160

[11] N. Kamčev and M. Schacht, *Canonical colourings in random graphs*, 2023. Available as arXiv:2303.11206.

[12] X. Li, H. Broersma, and L. Wang, *Integer colorings with no rainbow* 3*-term arithmetic progression*, Electron. J. Combin. **29** (2022), no. 2, Paper No. 2.28, 15pp. MR 4417188

[13] H. Lin, G. Wang, and W. Zhou, *Integer colorings with no rainbow k-term arithmetic progression*, European J. Combin. **104** (2022), Paper No. 103547, 12pp. MR 4414806

[14] R. Nenadov and A. Steger, *A short proof of the random Ramsey theorem*, Combin. Probab. Comput. **25** (2016), no. 1, 130–144. MR 3438289

[15] R. Rado, *Studien zur Kombinatorik*, Math. Z. **36** (1934), 424–480 (German).

[16] V. Rödl and A. Ruciński, *Lower bounds on probability thresholds for Ramsey properties*, Combinatorics, Paul Erdős is eighty, Vol. 1, Bolyai Soc. Math. Stud., János Bolyai Math. Soc., Budapest, 1993, pp. 317–346. MR 1249720

[17] ______, *Random graphs with monochromatic triangles in every edge coloring*, Random Structures Algorithms **5** (1994), no. 2, 253–270. MR 1262978

[18] ______, *Threshold functions for Ramsey properties*, J. Amer. Math. Soc. **8** (1995), no. 4, 917–942. MR 1276825

[19] ______, *Rado partition theorem for random subsets of integers*, Proc. London Math. Soc. (3) **74** (1997), no. 3, 481–502. MR 1434440

[20] D. Saxton and A. Thomason, *Hypergraph containers*, Invent. Math. **201** (2015), no. 3, 925–992. MR 3385638

[21] M. Schacht, *Extremal results for random discrete structures*, Ann. of Math. (2) **184** (2016), no. 2, 333–365. MR 3548528

[22] C. Spiegel, *A note on sparse supersaturation and extremal results for linear homogeneous systems*, Electron. J. Combin. **24** (2017), no. 3, Paper No. 3.38, 19pp. MR 3691555

[23] B. L. van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw Arch. Wiskd., II. Ser. **15** (1927), 212–216 (German).