Ramsey Theory

– Lecture Notes –
Version: April 10, 2014

Mathias Schacht

Fachbereich Mathematik, Universität Hamburg
E-mail address: schacht@math.uni-hamburg.de

Preface

These are the notes based on the course on *Ramsey Theory* taught at Universität Hamburg in Summer 2011 and 2014. The lectures were based on the textbook "Ramsey theory" of Graham, Rothschild, and Spencer [44]. In fact, large part of the material is taken from that book.

Hamburg, Spring 2014 Mathias Schacht

Contents

CHAPTER 1

Introduction

Ramsey theory is a branch of Discrete Mathematics, which was named after the seminal result of Ramsey [75]. Roughly speaking, Ramsey theory concerns the study of finite partitions (sometimes called colourings) of discrete structures, such as graphs, hypergraphs, integers, discrete functions, finite dimensional vector spaces over finite fields, posets etc. A typical result in Ramsey theory asserts that a given configuration will be completely contained in one of the partition classes for any finite partition of some sufficiently large or "rich" structure. We start with a brief overview and state some of the main results of that type.

1.1. A few cornerstones in Ramsey theory

1.1.1. Ramsey's theorem. Ramsey's theorem concerns partitions of the edge set of hypergraphs or set systems and we discuss it in detail in Chapter 2.

Theorem 1.1 (Ramsey 1930). *For all integers r, $k \geq 1$, and $\ell \geq k$ and every (countably) infinite set X the following holds. For any partition $E_1\dot\cup\dots\dot\cup E_r = \binom{X}{k}$ of the k-element subsets of X, there exists an index $j \in [r]$ and an ℓ-element subset $Y \subseteq X$ such that $\binom{Y}{k} \subseteq E_j$.*

The finite version of Ramsey's theorem (see Theorem 2.1) was reproved by Skolem [81] and by Erdős and Szekeres [30]. In fact, Erdős and Szekeres were interested in a question of Esther Klein. She proved that among 5 points in general position in the plane (i.e., no three points are colinear), there always exist 4 points which span a convex quadrilateral and asked for the following generalisation: *Does there exist for every integer $k \geq 3$ some integer n such that among every set of n points in general position in the plane there always exist k points, which span a convex k-gon.* In [30] Erdős and Szekeres showed that the affirmative answer of that question follows from the finite version of Ramsey's theorem and gave a new proof of Ramsey's theorem. Moreover, it is more than fair to say the Erdős popularised Ramsey's theorem a lot. Erdős was certainly one of the main contributors to Ramsey theory and we have to agree with the authors from [44, p. 26] who write "it is difficult to overestimate the effect of this paper" with reference to [30].

1.1.2. Ramsey-type results for the integers. The first Ramsey-type results predate Theorem 2.7 and concern partitions of the integers. One of the oldest results can be traced back to the work of Hilbert in [46].

Theorem 1.2 (Hilbert's cube lemma – 1892). *For all integers $r \geq 1$ and $k \geq 0$ the following holds. For any partition $E_1\dot\cup\dots\dot\cup E_r = \mathbb{N}$ of the natural numbers there exists some $j \in [r]$ and there exist natural numbers a and $\lambda_1 < \dots < \lambda_k \in \mathbb{N}$ such that the k-cube spanned by $a, \lambda_1, \dots, \lambda_k$ is contained in E_j, i.e., for every 0-1-vector $(\delta_1, \dots, \delta_k) \in \{0,1\}^k$ we have $a + \sum_{i=1}^{k} \delta_i\lambda_i \in E_j$.*

In fact, several other Ramsey-type results on the integers predate the work of Ramsey from 1930. One of them is due to Schur [79], who used the following "combinatorial lemma" to give a simpler proof of Dickson's theorem [13, 14], which asserts that Fermat's last theorem does not hold in modular arithmetic.

THEOREM 1.3 (Schur 1916). *For every integer $r \geq 1$, every integer $n \geq r!e$ and every partition $E_1 \dot\cup \ldots \dot\cup E_r = [n]$ of the first n positive integers there exists some $j \in [r]$ and x, y, $z \in E_j$ such that $x + y = z$.*

The next notable contribution is due to van der Waerden [89]. This result was proved in Hamburg while van der Waerden was visiting Artin.

THEOREM 1.4 (van der Waerden 1927). *For all integers $r \geq 1$ and $k \geq 1$ the following holds. For any partition $E_1 \dot\cup \ldots \dot\cup E_r = \mathbb{N}$ of the natural numbers there exists some $j \in [r]$ such that E_j contains an arithmetic progression of length k, i.e., there exists an $a \in \mathbb{N}$ and $\lambda > 0$ such that $a + i\lambda \in E_j$ for every $i = 0, \ldots, k-1$.*

In 1933 Rado (in his PhD-thesis [71] supervised by Schur) found a beautiful generalisation of Theorems 1.2-1.4. In fact he characterised all systems of linear equations with the property that any finite colouring of the integers yields a monochromatic solution. It is easy to check, that cubes of finite dimension, triples of the form $x + y = z$, and arithmetic progressions of finite size can be described as solutions of homogeneous systems of linear equations.

For an $\ell \times k$ matrix $A = (a_{ij}) \in \mathbb{Z}^{\ell \times k}$ of integers consider the system $\mathcal{L}(A)$ of homogeneous linear equations

$$\sum_{j=1}^{k} a_{ij} x_j = 0 \quad \text{for } 1 \leq i \leq \ell\,.$$

We say that a matrix A is *partition regular* if for any integer $r \geq 1$ and any partition $E_1 \dot\cup \ldots \dot\cup E_r$ of $\mathbb{N}$ there exists a solution $(x_1, \ldots, x_k)$ of $\mathcal{L}(A)$ with all x_i having the same colour. Rado obtained the following characterisation of partition regular matrices. For an $\ell \times k$ matrix A and $j \in [k]$ we denote by $\vec{a}^j$ the j-th column vector and for a set of integer vectors $\mathcal{V}$ we we denote by $\mathrm{Span}_{\mathbb{Q}}(\mathcal{V})$ the linear vector space spanned by $\mathcal{V}$ over $\mathbb{Q}$.

THEOREM 1.5 (Rado 1933). *The integer matrix $A \in \mathbb{Z}^{\ell \times k}$ is partition regular if and only if there exists a partition of $J_0 \dot\cup \ldots \dot\cup J_p = [k]$ of the indices of the columns such that*

(*i*) $\sum_{j \in J_0} \vec{a}^j = \vec{0}$ *and*
(*ii*) *for every $i = 1, \ldots, p$ we have* $\sum_{j \in J_i} \vec{a}^j \in \mathrm{Span}_{\mathbb{Q}}(\{\vec{a}^j \colon j \in J_0 \dot\cup \ldots \dot\cup J_{i-1}\})$.

Conditions (*i*) and (*ii*) are referred to as the *column condition*. If $\ell = 1$, then the column condition reduces (for non-trivial $A = (a_{11}, \ldots, a_{1k})$) to the simple condition that there exists a non-empty subset $J_0 \subseteq [k]$ such that $\sum_{j \in J_0} a_{1j} = 0$ and at least one of the a_{1j} for $j \in J_0$ is non-zero.

It is easy to check that the first non-trivial case of Theorem 1.2 ($k = 2$) follows from Theorem 1.5, since the solutions of the homogeneous system $\mathcal{L}(A)$ for $A = \begin{pmatrix}1 & -1 & -1 & 1\end{pmatrix}$ form 2-cubes in the sense of Theorem 1.2. Moreover, the qualitative statement of Theorem 1.3 follows from this characterisation, as Schur's theorem asserts monochromatic solutions of $\mathcal{L}(A)$ for the matrix $A = \begin{pmatrix}1 & 1 & -1\end{pmatrix}$. Similarly, it is easy to check that the special case of Theorem 1.4 for $k = 3$ follows from Theorem 1.5 by considering $A = \begin{pmatrix}1 & 1 & -2\end{pmatrix}$.

On the other hand, it follows from Theorem 1.5 that $\begin{pmatrix}1 & 1 & -3\end{pmatrix}$ is not partition regular, which might be a bit surprising in view of the fact that $\begin{pmatrix}1 & 1 & -1\end{pmatrix}$ and $\begin{pmatrix}1 & 1 & -2\end{pmatrix}$ are partition regular.

Another beautiful strengthening of Theorem 1.4 is due to Szemerédi [85]. It follows from Szemerédi's theorem that in fact the *largest* partition class E_j, i.e., the one which maximises

$$\limsup_{n \to \infty} \frac{|E_j \cap [n]|}{n}\,,$$

always contains an arithmetic progression of length k. Note that a similar generalisation of Schur's theorem obviously fails to be true, since for $r \geq 3$ we could choose all odd integers to be included in the largest partition class.

THEOREM 1.6 (Szemerédi 1975). *Let $k \geq 1$ be an integer. If a set $E \subseteq \mathbb{N}$ satisfies $\limsup_{n\to\infty} |E \cap [n]|/n > 0$, then E contains an arithmetic progression of length k.*

Theorem 1.6 was first conjectured by Erdős and Turán in 1936 [18] and the first non-trivial case ($k = 3$) was solved by Roth [77, 78]. Szemerédi's theorem initiated a lot of research in quite different areas of mathematics, including graph theory, ergodic theory, and additive number theory (see, e.g., [86]).

We discuss van der Waerden's theorem (Theorem 1.4) and Roth' theorem (Theorem 1.6 for $k = 3$) in detail in Chapter 3. Moreover, for one of the proofs of Roth' theorem we will consider a strengthened version of Theorem 1.2.

1.1.3. Ramsey-type results in higher dimensional spaces. All results in Section 1.1.2 concerned partitions of the integers and one may seek for natural generalisations for the d-dimensional lattice of integers $\mathbb{N}^d$.

We say $F' \subseteq \mathbb{N}^d$ is a *homothetic* copy of some a configuration $F \subseteq \mathbb{N}^d$ if there exists some $\vec{v}_0 \in \mathbb{N}^d$ and $\lambda > 0$ such that

$$F' = \vec{v}_0 + \lambda F = \{\vec{v}_0 + \lambda\vec{u}\colon\ \vec{u} \in F\}\,.$$

For example, Theorem 1.4 asserts that in any finite partition of $\mathbb{N}$ one of the partition classes contains a homothetic copy of $[k]$. The following generalisation for $d \geq 2$ was first obtained by Grünwald (also known as Gallai, see [72, page 123]) and, independently, for $d = 2$ by Witt [92] (former professor at Universiät Hamburg).

THEOREM 1.7 (Gallai 1942 and Witt 1952). *For all integers r, $d \geq 1$, and every finite configuration $F \subseteq \mathbb{N}^d$ the following holds. For any partition $E_1\dot\cup\dots\dot\cup E_r = \mathbb{N}^d$ of the d-dimensional integer lattice there exists some $j \in [r]$ such that E_j contains a homothetic copy of F.*

Theorem 1.7 asserts the existence of a "small" d-dimensional object in a "large" space of the same dimension d. The following theorem of Graham, Leeb and Rothschild [41], which was first conjectured by Rota, is in some sense complementary to Theorem 1.7. It asserts monochromatic subspaces in a sufficiently high dimensional vector space over a finite field.

THEOREM 1.8 (Graham, Leeb & Rothschild 1972). *For all integers $r \geq 1$ and $\ell \geq k \geq 0$ and every finite field $\mathbb{F}$ there exists some integer n_0 such that for every $n \geq n_0$ the following holds. For any partition $E_1\dot\cup\dots\dot\cup E_r$ of the k-dimensional subspaces of $\mathbb{F}^n$ there exists some $j \in [r]$ and some ℓ-dimensional subspace with all its k-dimensional subspaces belonging to E_j.*

The following affine version of Theorem 1.8 is equivalent to it (see Section 4.1.3).

THEOREM 1.9 (affine Ramsey theorem). *For all integers $r \geq 1$ and $\ell \geq k \geq 0$ and every finite field $\mathbb{F}$ there exists some integer n_0 such that for every $n \geq n_0$ the following holds. For any partition $E_1\dot\cup\dots\dot\cup E_r$ of the k-dimensional affine subspaces of $\mathbb{F}^n$ there exists some $j \in [r]$ and some ℓ-dimensional affine subspace with all its k-dimensional affine subspaces belonging to E_j.*

Already the case $k = 0$ and $\ell = 1$ in Theorem 1.9 is interesting. It asserts a monochromatic line (i.e., 1-dimensional affine subspaces) in any finite partition of the points of a sufficiently high dimensional space $\mathbb{F}^n$. In fact, for this case more is true. It can be shown that already for a very special set of lines the conclusion holds, i.e., one of those special lines is completely contained in one of the partition classes. In fact, this result is a direct consequence of a more abstract result due to Hales and Jewett [45].

DEFINITION 1.10. For a finite set A (also called *alphabet*) of cardinality k and an integer $n \geq 1$ we denote by A^n the set of all functions from $[n]$ to A.

We say a k-element subset $\mathcal{L} = \{f_a\colon\ a \in A\} \subseteq A^n$ is a *combinatorial line in* A^n if there exist a non-empty set $X \subseteq [n]$ and a function $g\colon\ [n] \setminus X \to A$ such that for every $a \in A$ we have

(*i*) $f_a(x) = g(x)$ for every $x \in [n] \setminus X$ and
(*ii*) $f_a(x) = a$ for every $x \in X$.

If $\mathbb{F} = \mathrm{GF}(p)$ is a prime field for some prime p, then it is easy to see that every combinatorial line in $\mathbb{F}^n$ is indeed a line in $\mathbb{F}^n$. On the other hand, in this case $\mathbb{F}^n$ contains only $(p+1)^n - p^n$ combinatorial lines, but the number of lines in $\mathbb{F}^n$ is

$$\frac{\binom{p^n}{2}}{\binom{p}{2}} = \frac{p^{n-1}(p^n-1)}{p-1} \gg (p+1)^n\,.$$

The Hales-Jewett theorem asserts that for any finite set A every finite partition of A^n for sufficiently large n yields a monochromatic combinatorial line. Consequently, this is clearly a strengthening of Theorem 1.9 for the case $k = 0$ and $\ell = 1$. In fact, this result is more general and Theorem 1.9 can be deduced from it (as was shown by Spencer in [83]) and also Theorems 1.4 and 1.7 are consequences of it.

THEOREM 1.11 (Hales & Jewett 1963). *For all integers $r \geq 1$ and every finite alphabet A there exists some integer n_0 such that for every $n \geq n_0$ the following holds. For any partition $E_1 \dot\cup \ldots \dot\cup E_r$ of A^n there exists some $j \in [r]$ such that E_j contains a combinatorial line.*

We discuss the Hales-Jewett theorem and some of its consequences in Chapter 4.

1.2. A unifying framework

All statements mentioned so far and many other results in Ramsey theory can interpreted as statements regarding the *chromatic number* of a suitable chosen *hypergraph*.

DEFINITION 1.12 (hypergraphs). A *hypergraph* $H = (V, E)$ is a pair, where the *vertex set* V is some set and the set of *hyperedges* $E \subseteq 2^V$ is a subset of the power set. For a hypergraph H we denote by $V(H)$ the set of *vertices* and by $E(H)$ the set of its *hyperedges*, i.e., $H = (V(H), E(H))$. For a subset $W \subseteq V$ we denote by $H[W] = (W, E \cap 2^W)$ the subhypergraph of H *induced* on W.

A hypergraph $H = (V, E)$ is *k-uniform* for some integer $k \geq 1$, if $E \subseteq \binom{V}{k}$ and 2-uniform hypergraphs are called *graphs*.

In these notes the vertex set V will be almost always just finite or countable. Now we can model the Ramsey-type theorems from Section 1.1 in the following way: Let V be the set which will be partitioned and let E correspond to the configurations, which are asserted to appear in at least one of the partition classes.

More concretely, for example in the context of van der Waerden's theorem for some fixed k, the vertex set is $V = \mathbb{N}$ and an edge e is a k-element subset of $\mathbb{N}$, which forms an arithmetic progression, i.e., the corresponding hypergraph is k-uniform. For a fixed integer k, Theorem 1.4 asserts that for any partition of V some partition classes induces at least one edge. Similarly, in the context of Ramsey's theorem $V = \binom{X}{k}$ and for every ℓ-element set $Y \subseteq X$ the family $e_Y = \binom{Y}{k}$ forms an edge in the corresponding hypergraph H. Then Ramsey's theorem asserts that for every finite partition of the vertices of H at least one of the partition classes contains one edge. We can reword this assertion for every partition of the vertex set of the corresponding hypergraph, in terms of the chromatic number of the hypergraph.

DEFINITION 1.13 (chromatic number). The *chromatic number* $\chi(H)$ of a hypergraph $H = (V, E)$ is the smallest integer $r \geq 1$ such that there exists a partition $V_1 \dot\cup \dots \dot\cup V_r = V$ of the vertex set such that for every $j \in [r]$ the partition class V_j contains no edge of H, i.e., $E \cap 2^{V_j} = \emptyset$ for every $j \in [r]$. If no such integer r exists, then we set $\chi(H) = \infty$.

With this definition at hand Theorems 1.1–1.11 can be phrased as follows: *the chromatic number of the corresponding hypergraph is bigger than r for every $r \geq 1$.* Note, however, that in Theorems 1.1, 1.2, 1.4, 1.7–1.11 we have additional parameters (like integers k and ℓ, matrix A, configuration F, and alphabet A) and the corresponding hypergraphs would be "customised" for one particular choice of those additional parameters.

1.3. The compactness principle

In this introduction we chose to present the infinite version of several of the theorems (Theorems 1.1, 1.2, 1.4–1.7). However, also finite versions of those theorems hold, in which the infinite set X, or the set $\mathbb{N}$, or $\mathbb{N}^d$ will be replaced by an n-element set X, or $[n]$, or $[n]^d$ for some sufficiently large n (depending on the other parameters of those theorems). In fact, for several results in Ramsey theory the finite and infinite version are (qualitatively) equivalent. This correspondence is often established by the *compactness principle*, which first appeared in the work of Rado [**73**] (see also [**39, 9**]). We state the compactness principle in the setting of hypergraphs from the last section.

THEOREM 1.14 (compactness principle). *Let $r \geq 1$ be an integer and let $H = (V, E)$ be a hypergraph such that every edge $e \in E$ is finite. If $\chi(H[W]) \leq r$ for every finite subset $W \subseteq V$, then $\chi(H) \leq r$.*

Theorem 1.14 is trivial when V is finite and can be easily proved for countable sets V. In the countable case it has a simple elementary proof or it can be deduced from Kőnig's infinity lemma [**52**] (see, e.g., [**15**, Section 8]). For uncountable sets V the proof of the compactness principle relies on the axiom of choice and can be deduced from Tychonoff's theorem [**88**] (Theorem A.10). For completeness we include this proof here (a short overview over the necessary notions from topology can be found in Appendix A.1). The proof based on Tychonoff's theorem goes back to [**39**].

PROOF. Let $H = (V, E)$ and $r \geq 1$ be given. We consider $X = [r]^V$, the family of all functions from $f\colon V \to [r]$. We view X as a topological space. In fact, consider $[r]$ to be the topological space endowed with the discrete topology, and let X be the product space $\prod_{v \in V}[r]$, endowed with the product topology. Since $[r]$ is a compact topological space, it follows from Theorem A.10 that X is compact.

For a function $f \in X$, we consider the corresponding partition $V_1^f \dot\cup \dots \dot\cup V_r^f = V$, where $V_j^f = f^{-1}(j)$ for every $j \in [r]$. For a finite subset $W \subseteq V$ we denote by $X_W \subseteq X$ the set of all functions $f \in X$ with the property that the corresponding partition $V_1^f \dot\cup \dots \dot\cup V_r^f = V$ induces no monochromatic edge in $H[W]$, i.e., there exists no edge $e \in E \cap 2^W$ such that $e \subseteq V_j^f \cap W$ for some $j \in [r]$.

Owing to the assumption on H in Theorem 1.14 we have $X_W \neq \emptyset$ for every finite subset $W \subseteq V$.

Below we check that for every finite set $W \subseteq V$ the family X_W is closed (and open) in X. Indeed, since W and $[r]$ are finite, the set X_W is the finite union of sets X_W^g, where for some colouring $g\colon W \to [r]$ we set

$$X_W^g = \prod_{v\in V} Z_v^g \quad \text{where} \quad Z_v^g = \begin{cases} [r] & \text{if } v \in V \setminus W, \\ g(v) & \text{if } v \in W\,. \end{cases}$$

Obviously, we have $X_W = \bigcup_g X_W^g$, where the union runs over all colourings g of W which do not induce a monochromatic edge in W. Equivalently, $X_W = X \setminus \bigcup_h X_W^h$ where the union runs over all colourings h of W which induce a monochromatic edge in W. Since, the set X_W^h is a basic open set in the product topology of X for every function $h\colon W \to [r]$, we infer that X_W is a closed set in X.

Summarising, we obtain that

$$\mathscr{C} = \{X_W \colon W \subseteq V \text{ and } W \text{ is finite}\}$$

is a collection of closed sets in the compact space X. The collection $\mathscr{C}$ has the finite intersection property, i.e., the joint intersection of finitely many members of $\mathscr{C}$ is non-empty. Indeed, let $W_1, \dots, W_\ell \subseteq V$ be a finite collection of finite subsets from V. Hence, the set $W = \bigcup_{i\in[\ell]} W_i$ is finite and

$$X_{W_1} \cap \dots \cap X_{W_\ell} \supseteq X_W \neq \emptyset\,.$$

Since X is compact and $\mathscr{C}$ is a collection of closed sets in X with the finite intersection property, it follows from Proposition A.8 that

$$\bigcap_{\substack{W\subseteq V \\ W \text{ finite}}} X_W \neq \emptyset\,.$$

In other words, there exists a function $f\colon V \to [r]$ such that $f \in X_W$ for every finite set $W \subseteq V$. We claim that the partition $V_1^f \dot\cup \dots \dot\cup V_r^f = V$ shows that $\chi(H) \leq r$. Suppose for a contradiction that there is some $e \in E$ and $j \in [r]$ such that $e \subseteq V_j^f$. In particular, $f \notin X_e$, but by assumption on H the edge e is finite, which yields a contradiction to the fact that $f \in X_W$ for every finite set $W \subseteq V$. $\square$

1.4. Other topics in Ramsey theory

Owing to the fact that these are the lecture notes for a short course in Ramsey theory, there are many topics in Ramsey theory, which we will not here. Below we briefly mention some of them and include a few references for further studies.

Ramsey numbers: Already Ramsey discussed his estimate for the smallest integer n_0 (see definition of the Ramsey function $R^{(k)}(\ell_1, \dots, \ell_r)$ in Section 2.1) for which the finite version of Ramsey's theorem holds (see Theorem 2.1). Improving the upper bound was one of the main motivations of Skolem [81] and this bound was further improved by Erdős and Szekeres [30]. The first exponential lower bounds for graphs was given by a probabilistic argument by Erdős [19]. (In fact [19] marks one of the first appearances of the probabilistic method in combinatorics, which grew into an important brach of combinatorics itself.) The bounds obtained in [30] and [19] yield

$$2^{\ell/2} \leq R^{(2)}(\ell, \ell) \leq 2^{2\ell}\,.$$

Despite a lot of effort these bounds were only slightly improved within the last 60 years (see [82] and [7]).

For hypergraphs the gap between the upper and the lower bound is much larger. It follows from the work of Erdős and Rado [28] that $R^{(k)}(\ell, \ell)$ is bound from above by a $(k-1)$-times iterated exponential function, while the lower bound of Erdős, Hajnal, and Rado [25] is a $(k-2)$-times iterated exponential. For example, for $k = 3$ we have

$$2^{c\ell^2} \leq R^{(3)}(\ell,\ell) \leq 2^{2^{c'\ell}}.$$

for some positive constants c and c' independent of ℓ. Erdős, Hajnal, and Rado conjectured that the upper bound is asymptotically correct (up to the constant c') and Erdős offered \$500 for a proof (see, e.g., [20]). We briefly discuss asymptotic bounds for $R^{(k)}(\ell,\ell)$ in Section 2.4. More details can be found in [42] and [44, Chapter 4] (see also [8] for more recent developments).

In some cases the precise value of the Ramsey function is known and we refer the interested reader to the dynamic survey [74].

Ramsey classes: We stated some classical results in Ramsey theory and discussed a unifying framework in Section 1.2. Another abstract generalisation, which was considered by Leeb [55] and others (see, e.g., [10, 62]), is the following. Suppose $\mathscr{C}$ is a class of finite discrete structures endowed with a notion of isomorphism ("$\equiv$") between elements of $\mathscr{C}$ and a transitive notion of substructure ("$\sqsubseteq$"), which is compatible with the notion of isomorphism (i.e., if B and $B' \in \mathscr{C}$ are isomorphic, then there is a one-to-one correspondence φ between the substructures of B and B' such that every substructure A of B is isomorphic to the substructure $\varphi(A)$ of B'). This abstract formalism can be naturally described in terms of *category theory*, where the substructure relation corresponds to the class of *morphisms*.

For A and $B \in \mathscr{C}$ we denote by $\binom{B}{A}_{\mathscr{C}}$ the family of all copies of A in B, i.e., the set of all substructures A' of B which are isomorphic to A

$$\binom{B}{A}_{\mathscr{C}} = \{A' \in \mathscr{C}\colon\ A' \sqsubseteq B \text{ and } A' \equiv A\}.$$

We say a class $\mathscr{C}$ has the *Ramsey property for* $A \in \mathscr{C}$ (or $\mathscr{C}$ *is* A*-Ramsey*) if for every $r \in \mathbb{N}$ and every $B \in \mathscr{C}$ there exists some $C \in \mathscr{C}$ such that

$$C \longrightarrow (B)^A_r$$

by which we mean that for every partition $E_1 \dot\cup \ldots \dot\cup E_r$ of $\binom{C}{A}_{\mathscr{C}}$ there exist $j \in [r]$ and $B' \in \binom{C}{B}_{\mathscr{C}}$ such that $\binom{B'}{A}_{\mathscr{C}} \subseteq E_j$. Moreover, we say $\mathscr{C}$ is a *Ramsey class*, if it has the Ramsey property for every $A \in \mathscr{C}$.

For example, let $\mathscr{C}$ be the class of all finite sets, where two sets of the same cardinality are considered isomorphic and the substructure relation coincides with the subset relation. Then the finite version of Ramsey's theorem asserts that for every $k \in \mathbb{N}$ the class $\mathscr{C}$ is A-Ramsey for every k-element set A. Consequently, the finite version of Ramsey's theorem (see Theorem 2.1) is equivalent to the statement that $\mathscr{C}$ is a Ramsey class.

Another example is given by Theorem 1.8, which asserts that the set of all finite dimensional vector spaces over a given finite field is a Ramsey class (where subobjects are subspaces and all spaces of the same dimension are isomorphic).

The notion of a Ramsey class is very restrictive and only few other Ramsey classes were found [2, 11, 55, 56, 58, 60, 61, 63, 64, 65, 66, 67] (see also the survey of Nešetřil [59]). A deep connection between Ramsey classes, Fraïssé limits of ultrahomogeneous structures, and topological dynamics of extremely amenable automorphism groups of countable structures was established by Kechris, Pestov and Todorcevic [51].

Canonical Ramsey theory: In canonical versions of Ramsey-type theorems the number of colours (or partition classes) is not fixed. In particular, we allow colourings where every element in the underlying space may get a different colour and clearly in this case no interesting monochromatic configuration can be guaranteed. However, it turns out that for many instances one can guarantee the existence of a configuration with a *canonical* colouring. Canonical colourings usually include monochromatic copies, rainbow copies (i.e., every element in the configuration has a different colour) and other "obvious" colour patterns (which depend on the context). A canonical version of Ramsey's theorem was first studied by Erdős and Rado [27] and the survey of Deuber [12] is a good starting point for further studies.

Transfinite Ramsey theory: Roughly speaking, transfinite Ramsey theory concerns extensions of Ramsey's theorem to cardinal numbers. One of the first results of this sort is the Dushnik–Miller–Erdős theorem [16]. Many more results can be found in Erdős and Rado [28, 29] and for more details we refer to the work Erdős, Hajnal, and Rado [25] and to the monograph [24].

Euclidean Ramsey theory: The main questions in this area concern extensions of Theorem 1.7 in $\mathbb{R}^d$ (instead of $\mathbb{N}^d$) with the following twists: on the one hand we insist on real copies of F and do not allow homothetic copies, i.e., one of the partition class must contain a translate of F; on the other hand, we are satisfied, if such copies can be only guaranteed for finite partitions of $\mathbb{R}^n$ for some sufficiently large n (depending on the number of partition classes and on the configuration F). Questions of this type were first considered by Erdős et al. in [21, 22, 23]. In particular, it was shown that every configuration $F \subseteq \mathbb{R}^d$ for which such a Ramsey-type statement holds, must be contained on some $(d-1)$-dimensional the sphere in $\mathbb{R}^d$. Graham conjectures that this condition is also sufficient for being Ramsey and offers \$ 1000 for a proof (see, e.g., [40]). So far this conjecture is known to be true only for a few point sets, including the corners of rectangular parallelepipeds [21], of simplices [33], and of Platonic solids [53].

CHAPTER 2

Ramsey's theorem

2.1. Statement of Ramsey's theorem and notation

We first state the finite version Ramsey's theorem from [75]. Theorem 2.1 asserts that for any r-colouring of the edges of the complete k-uniform hypergraph on some sufficiently large (depending on r, k, and ℓ) vertex sets X there exists a monochromatic clique on at least ℓ vertices.

THEOREM 2.1 (Ramsey's theorem – finite version). *For all integers $r \geq 1$, $k \geq 1$, and $\ell \geq k$ there exists an n_0 such that for every n-element set X with $n \geq n_0$ the following holds. For every partition of $E_1 \dot\cup \dots \dot\cup E_r = \binom{X}{k}$ of the k-element subsets of X there exists an ℓ-element subset $Y \subseteq X$ such that $\binom{Y}{k} \subseteq E_j$ for some $j \in [r]$.*

Clearly, Theorem 2.1 implies Theorem 1.1. In fact, the compactness principle (Theorem 1.14) yields the equivalence of these two versions of Ramsey's theorem.

We will use the following convenient notation. For a set X satisfying the conclusion of Theorem 2.1 we write

$$X \longrightarrow (\ell)^k_r$$

and for $X = [n] = \{1, \dots, n\}$ we simply write

$$n \longrightarrow (\ell)^k_r\,.$$

The smallest integer n having the property $n \longrightarrow (\ell)^k_r$ is the *Ramsey number*

$$R^{(k)}(\ell; r)$$

and Ramsey's theorem asserts that the Ramsey function $R^{(k)}(\ell; r)$ is well defined.

For the proof of Theorem 2.1 it will be convenient to consider an asymmetric version of the Ramsey function. For a set X and integers $r \geq 1$, $k \geq 1$, and $\ell_1, \dots, \ell_r \geq k$ we write

$$X \longrightarrow (\ell_1, \dots, \ell_r)^k$$

if for any partition $E_1 \dot\cup \dots \dot\cup E_r$ of $\binom{X}{r}$ there exists some $j \in [r]$ and an ℓ_j-element subset $Y \subseteq X$ such that $\binom{Y}{k} \subseteq E_j$. Similarly, as above we write $n \longrightarrow (\ell_1, \dots, \ell_r)^k$ for $[n] \longrightarrow (\ell_1, \dots, \ell_r)^k$ we denote by

$$R^{(k)}(\ell_1, \dots, \ell_r)$$

the smallest integer n with the property $n \longrightarrow (\ell_1, \dots, \ell_r)^k$.

THEOREM 2.2 (Ramsey's theorem). *For all integers r, $k \geq 1$, and $\ell_1, \dots, \ell_r \geq k$ there exists an integer n such that $n \longrightarrow (\ell_1, \dots, \ell_r)^k$.*

Clearly, Theorems 2.1 and 2.2 are equivalent, since

$$R^{(k)}(\ell; r) = R^{(k)}(\underbrace{\ell, \dots, \ell}_{r\text{-times}}) \quad \text{and} \quad R^{(k)}(\ell_1, \dots, \ell_r) \leq R^{(k)}(\max_{j \in [r]} \ell_j; r)\,.$$

2.2. Proof of Ramsey's theorem for $k = 1$ and 2

2.2.1. The pigeonhole principle. Ramsey's theorem for $k = 1$ is equivalent to the pigeonhole principle (or Dirichlet's Schubfachprinzip). In fact, it is easy to see that

$$R^{(1)}(\ell_1, \ldots, \ell_r) = 1 + \sum_{j=1}^{r} (\ell_j - 1)\,. \tag{2.1}$$

Note, that for the proof of equality (2.1) we are required to show two assertions. Firstly, we have to show that the 1-element subsets of an $(\sum_{j=1}^{r}(\ell_j - 1))$-element set X can be partitioned into r classes $E_1 \dot\cup \ldots \dot\cup E_r = \binom{X}{1}$ such that $|E_j| < l_j$ for every $j \in [r]$, which is obviously possible. Secondly, we have to show that such a partition does not exist for a $(1 + \sum_{j=1}^{r}(\ell_j - 1))$-element set X. But also this is rather obvious, since supposing the contrary and assuming the existence of a partition $E_1 \dot\cup \ldots \dot\cup E_r = \binom{X}{1}$ such that $|E_j| < \ell_j$ gives rise to the following contradiction

$$|X| = \left|\binom{X}{1}\right| = \sum_{j=1}^{r} |E_j| \leq \sum_{j=1}^{r} (\ell_j - 1) < 1 + \sum_{j=1}^{r} (\ell_j - 1) = |X|\,.$$

2.2.2. Ramsey's theorem for graphs. Since 2-uniform hypergraphs are (simple, undirected) graphs, we will use the language from graph theory here. Below, we give two proofs of Ramsey's theorem for graphs. First we show Theorem 2.2 for $k = 2$ by induction on $\sum_j \ell_j$ (see Section 2.2.2.1). Then in Section 2.2.2.2 we present a proof, which reduces the problem more directly to the $k = 1$ case.

2.2.2.1. *Induction on* $\sum_j \ell_j$. Let us warm up with the first non-trivial cases. Clearly, $R^{(2)}(2, 2) = 2$ and it is not hard to show that $R^{(2)}(2, 3) = R^{(2)}(3, 2) = 3$. This leaves $R^{(2)}(3, 3)$ as the first open case. Below we give a simple argument, which shows that

$$R^{(2)}(3, 3) \leq 2 + (R^{(2)}(2, 3) - 1) + (R^{(2)}(3, 2) - 1) = 6\,.$$

For that we have to show that every colouring of the edges of K_6 (i.e., for every partition $E_1 \dot\cup E_2 = E(K_6)$) there exists a monochromatic copy of K_3 (i.e., $K_3 \subseteq E_1$ or $K_3 \subseteq E_2$). For that let $E_1 \dot\cup E_2$ be an arbitrary partition of $E(K_6)$ and let $v \in V(K_6)$ be arbitrary. Let $N_1(v)$ be the neighbours of v, which are connected to v by an edge from E_1 and let $N_2(v)$ be defined analogously. Since

$$|N_1(v)| + |N_2(v)| = 5 > (R^{(2)}(2, 3) - 1) + (R^{(2)}(3, 2) - 1)$$

either $|N_1(v)| \geq R^{(2)}(2, 3)$ or $|N_2(v)| \geq R^{(2)}(3, 2)$. Without loss of generality we may assume that $|N_1(v)| \geq R^{(2)}(2, 3)$ and in this case the definition of $R^{(2)}(2, 3)$ asserts: either $E_1 \cap \binom{N_1(v)}{2}$ contains a copy of K_2, which together with v extends to a copy of K_3 in E_1 or $E_2 \cap \binom{N_1(v)}{2}$ contains a copy of K_3. In either of the cases we found a monochromatic copy and we are done.

In fact, one can also show that $R^{(2)}(3, 3) > 5$ by considering an edge partition of K_5 given by two edge-disjoint cycles of length 5. This way we obtain $R^{(2)}(3, 3) = 6$. In general it is extremely hard to determine the Ramsey function precisely and only a very few precise results are known. However, the proof of the upper bound on $R^{(2)}(3, 3)$ easily extends to the general case. This proof of Ramsey's theorem appeared already in [30] and we give the details of this proof below.

Proof of Theorem 2.2 for $k = 2$. We proceed by induction on $\sum_{j=1}^{r} \ell_j$. Clearly, for $r = 1$ we have $R^{(2)}(\ell; 1) = \ell$ for every integer $\ell \geq 2$. Moreover, it follows directly from the definition of the Ramsey function that

$$R^{(2)}(\ell_1, \ldots, \ell_{r-1}, 2) = R^{(2)}(\ell_1, \ldots, \ell_{r-1})$$

and, more generally, we have

$$R^{(2)}(\ell_1,\dots,\ell_{j-1},2,\ell_{j+1},\dots,\ell_r) = R^{(2)}(\ell_1,\dots,\ell_{j-1},\ell_{j+1}\ell_r)\,. \tag{2.2}$$

In particular, $R^{(2)}(2;r) = 2$ for every $r \geq 1$, which establishes the induction start.

For the induction step let $r > 1$ and $\ell_1,\dots,\ell_r \geq 2$. If $\ell_j = 2$ for some $j \in [r]$, then we can appeal to (2.2) and obtain $R^{(2)}(\ell_1,\dots,\ell_r) = R^{(2)}(\ell_1,\dots,\ell_{j-1},\ell_{j+1}\ell_r)$, which is ensured by induction assumption. So let $\ell_1,\dots,\ell_r \geq 3$. Similar as in the proof of $R^{(2)}(3,3) \leq 6$ we set

$$n = 2 + \sum_{j=1}^{r}\big(R^{(2)}(\ell_1,\dots,\ell_{j-1},\ell_j - 1,\ell_{j+1},\dots,\ell_r) - 1\big) \tag{2.3}$$

and we will show that

$$R^{(2)}(\ell_1,\dots,\ell_r) \leq n\,.$$

For that let $E_1\dot\cup\dots\dot\cup E_r$ be an arbitrary partition of $E(K_n)$, let v be some vertex of K_n and for $j \in [r]$ we denote by $N_j(v)$ those neighbours of v which are joint to v by an edge from E_j. Since v has $n-1$ neighbours, the pigeonhole principle and the choice of n ensures the existence of some index $j \in [r]$ such that

$$|N_j(v)| \geq R^{(2)}(\ell_1,\dots,\ell_{j-1},\ell_j - 1,\ell_{j+1},\dots,\ell_r)\,.$$

Consequently, there either exists an $i \neq j$ such that E_i (restricted to $N_j(v)$) contains a copy of K_{ℓ_i} or E_j restricted to $N_j(v)$ (i.e., $E_j \cap \binom{N_j(v)}{2}$) contains a copy of $K_{\ell_j - 1}$, which together with v spans a copy of K_{ℓ_j}. In either case, there exists an $i \in [r]$ such that E_i contains a copy of K_{ℓ_i}. Since $E_1\dot\cup\dots\dot\cup E_r = E(K_n)$ was arbitrary, this shows $R^{(2)}(\ell_1,\dots,\ell_r) \leq n$. □

One may argue that the proof given above relied on the pigeon hole principle and, hence, in some sense we reduced the $k = 2$ case of Theorem 2.2 to the case $k = 1$. In fact, in the proof we chose n to be $n = 1 + R^{(1)}(L_1,\dots,L_r)$ where

$$L_j = R^{(2)}(\ell_1,\dots,\ell_{j-1},\ell_j - 1,\ell_{j+1},\dots,\ell_r)\,,$$

and this view leads to a generalisation of the proof for arbitrary k. Another key idea in this proof was the reduction of $\sum_j \ell_j$, by applying the induction assumption too a suitably large monochromatic neighbourhood of some fixed vertex. In the proof presented in the next section we also consider monochromatic neighbourhoods. However, the difference between the two approaches becomes more apparent, when we extend them to arbitrary k.

2.2.2.2. *Reduction to* $k = 1$. Again we outline the main idea for the case $R^{(2)}(3,3)$. This time we only show that

$$R^{(2)}(3,3) \leq 2^{R^{(1)}(2,2)} = 8\,.$$

Let $E_1\dot\cup E_2 = E(K_8)$ be arbitrary and fix some vertex x_1 from $V(K_8)$. Note that either at least 4 of the 7 neighbours of x_1 are connected to x_1 by edges from E_1 or the same statement holds when E_1 is replaced by E_2. Let X_1 be the set of those neighbours and let $i_1 \in \{1,2\}$ indicate the matching index of the edge set. We repeat this choice within X_1. So let $x_2 \in X_1$ be arbitrary and note that at least two of the three neighbours of x_2 in $X_1 - x_1$ are either joined to x_2 only by edges form E_1 or only by edges from E_2. Let X_2 be the set of those neighbours and let $i_2 \in \{1,2\}$ be the index of the corresponding set from the partition. Finally, select one vertex x_3 from the two remaining vertices in X_2, set $X_3 = X_2 - x_3$ and let i_3 be the index of the edge set, which contains the edge consisting of x_3 and the vertex from X_3. By the pigeonhole principle, there exist two vertices x_{j_1} and x_{j_2} with $j_1 < j_2$ in the set $\{x_1,x_2,x_3\}$ for which $i_{j_1} = i_{j_2}$. Let y denote the unique vertex from X_3. We claim that x_{j_1}, x_{j_2}, and y span a K_3 contained in $E_{i_{j_1}}$. In fact

x_{j_2} and y are contained in X_{j_1} and, hence, the edges $x_{j_1}x_{j_2}$ and $x_{j_1}y$ are contained in $E_{i_{j_1}}$. Moreover, $y \in X_{j_2}$ and, hence, $x_{j_2}y \in E_{i_{j_2}}$ and the claim follows from $i_{j_1} = i_{j_2}$.

Roughly speaking, the proof above can be generalised for $R^{(2)}(\ell_1, \dots, \ell_r)$ by the following reasoning. Suppose $n = r^t$ for $t = R^{(1)}(\ell_1 - 1, \dots, \ell_r - 1)$ then we can select vertices $x_1, \dots, x_t$ and y such that all edges from $x_j x_{j'}$ with $j' > j$ and $x_j y$ come from the same partition class, say E_{i_j}. In fact, in every selection step we will ensure that the set of remaining vertices shrinks roughly by the factor $1/r$ and, hence, our choice of n is supposed to ensure that after t selection steps at least one vertex y is left.

Owing to the special property that the partition class from $E_1, \dots, E_r$ which contains every edge $x_j x_{j'}$ with $j' > j$ and $x_j y$ only depends on the vertex x_j, we naturally obtain a partition of $\binom{X'}{1}$ for $X' = \{x_1, \dots, x_t\}$ into r classes. Then Ramsey's theorem for $k = 1$ asserts the existence of some $j \in [r]$ and a subset $Y' = \{y_1, \dots, y_{\ell_j - 1}\} \subseteq X'$ such that $\binom{Y'}{1}$ is from the same partition class. In fact, this implies $\binom{Y'+y}{2} \subseteq E_j$ for some $j \in [r]$ and we conclude Theorem 2.2 for $k = 2$. Below we give the details of this outline. This proof of Ramsey's theorem goes back to the work of Erdős and Rado [28].

Proof of Theorem 2.2 for $k = 2$. Since $r = 1$ is obvious, let $r \geq 2$ and let $\ell_1, \dots, \ell_r \geq 2$ be given. Appealing to Theorem 2.2 for $k = 1$, we set

$$t = R^{(1)}(\ell_1 - 1, \dots, \ell_r - 1) \qquad \text{and} \qquad n = r^t$$

and we will show that

$$R^{(2)}(\ell_1, \dots, \ell_r) \leq n\,.$$

So let $E_1 \dot\cup \dots \dot\cup E_r$ be an arbitrary partition of $E(K_n)$. We set $X_0 = V(K_n)$ and inductively we show that for every $i = 1, \dots, t$ there exists a vertex $x_i \in X_{i-1}$ and a set $X_i \subseteq X_{i-1} - x_i$ with

- (i) $|X_i| \geq n/r^i = r^{t-i}$ and
- (ii) there exists some $j_i \in [r]$ such that $x_i x \in E_{j_i}$ for every $x \in X_i$.

Let $i \geq 1$ and suppose X_{i-1} was chosen already and $|X_{i-1}| \geq n/r^{i-1}$. (Note that X_0 satisfies this condition.) We fix some vertex $x_i \in X_{i-1}$. Since $r \geq 2$ and since $n > r^{i-1}$ we have

$$\left\lceil \frac{n/r^{i-1} - 1}{r} \right\rceil = \frac{n}{r^i}$$

and, consequently, there exists a set $X_i \subseteq X_{i-1} - x_1$ of size at least n/r^i and an index $j_i \in [r]$ such that all neighbours of x_i in X_i are connected by an edge from E_{j_i}.

Having verified the existence of vertices x_i and sets X_i satisfying (i) and (ii) for every $i = 1, \dots, t$. In particular X_t is not empty and finally we fix a vertex $y \in X_t$.

Set $X' = \{x_1, \dots, x_t\}$ and consider the following partition of $E'_1 \dot\cup \dots \dot\cup E'_r$ of $\binom{X'}{1}$ given by $\{x_i\} \in E'_j$ if $j_i = j$. Since $|X'| = R^{(1)}(\ell_1 - 1, \dots, \ell_r - 1)$ Theorem 2.2 for $k = 1$ yields an index $j \in [r]$ and a subset $Y' \subseteq X'$ with $|Y'| \geq \ell_j - 1$ such that $\binom{Y'}{1} \subseteq E'_j$, i.e., for every $x_i \in Y'$ we have $j_i = j$. Consequently, for every $x_i \in Y'$ we have $x_i y \in E_j$ and for every $i' \geq i$ we have $x_i x_{i'} \in E_j$, since such $x_{i'}$ and y are contained in X_i. Therefore, $\binom{Y'+y}{2} \subseteq E_j$ and we conclude the proof. $\square$

2.3. Proof of Ramsey's theorem for general k

We shall extend the proofs of Ramsey's theorem for graphs from Section 2.2.2. In both arguments we selected vertices and monochromatic neighbourhoods. Hence extending those arguments to k-uniform hypergraphs requires us to decide if a vertex was a 1-element subset or $(k-1)$-element subset or something completely different. Sometimes we can extend those arguments either way, giving different looking proofs of Ramsey's theorem and we will give three proofs of Ramsey's theorem. All proofs proceed by induction on k and we shall always assume that Theorem 2.2 holds for all $k' < k$.

First proof of Theorem 2.2. This proof is due to Erdős and Szekeres [30]. We use induction on $\sum_j \ell_j$. Similarly as in the corresponding proof for graphs, we may assume that $r \geq 2$ and $\ell_1, \dots, \ell_r \geq k+1$, since

$$R^{(k)}(\ell; 1) = \ell$$

and

$$R^{(k)}(\ell_1, \dots, \ell_{j-1}, k, \ell_{j+1}, \dots, \ell_r) = R^{(k)}(\ell_1, \dots, \ell_{j-1}, \ell_{j+1}\ell_r)\,.$$

For $j \in [r]$ we set

$$L_j = R^{(k)}(\ell_1, \dots, \ell_{j-1}, \ell_j - 1, \ell_{j+1}, \dots, \ell_r)\,,$$

which exist due to our induction assumption on $\sum_j \ell_j$. Next we appeal to the induction assumption for $k-1$ and we set

$$n = 1 + R^{(k-1)}(L_1, \dots, L_r)$$

and we shall show that

$$R^{(k)}(\ell_1, \dots, \ell_r) \leq n\,.$$

Let X be an n-element set and let $E_1 \dot\cup \dots \dot\cup E_r = \binom{X}{k}$ be an arbitrary partition of the k-element subsets of X. Fix some $x \in X$ and consider the "induced" partition $E_1' \dot\cup \dots \dot\cup E_r' = \binom{X-x}{k-1}$ of the $(k-1)$-element subsets of $X - x$. More precisely, a $(k-1)$-element subsets $K' \in \binom{X-x}{k-1}$ will be contained in E_j' if $(K'+x) \in E_j$.

Owing to the choice of n and to Ramsey's theorem for $k-1$, there exists an index $j \in [r]$ and a subset $Y \subseteq X - x$ of size $|Y| \geq L_j$ such that $\binom{Y}{k-1} \subseteq E_j'$, i.e., $(K'+x) \in E_j$ for every $K' \in \binom{Y}{k-1}$. Owing to the choice of L_j, we infer that either there exists some index $i \neq j$ and a subset $Z_i \subseteq Y$ of size ℓ_i such that $\binom{Z_i}{k} \subseteq E_i$ or there exists a subset $Z_j \subseteq Y$ of size $\ell_j - 1$ such that $\binom{Z_j}{k} \subseteq E_j$. In the former case we are done. For the latter case the set Z_j is one element short of the desired size. However, we recall that $Z_j \subseteq Y$ and, therefore, we also have $(K'+x) \in E_j$ for every $K' \in \binom{Z_j}{k-1}$. Consequently, $\binom{Z_j+x}{k} \subseteq E_j$ in this case, which concludes the proof of Theorem 2.2. □

Next we extend the proof from Section 2.2.2.2 in two ways.

Second proof of Theorem 2.2. Let $r \geq 1$ and $\ell_1, \dots, \ell_r \geq k$ be given. Set

$$t = R^{(1)}(\ell_1 - k + 1, \dots, \ell_r - k + 1)\,.$$

We consider the following recursively defined sequence of integers. For $i = t, \dots, 1$ we define

$$L_t = 1 + R^{(k-1)}(k-1; r) = k \quad \text{and} \quad L_i = 1 + R^{(k-1)}(L_{i+1}; r)$$

and we set

$$n = L_1\,.$$

It is easy to verify that this choice of n suffices, i.e., that $n \longrightarrow (\ell_1, \dots, \ell_r)^k$. In fact, for an n-element set X and a partition $E_1 \dot\cup \dots \dot\cup E_r = \binom{X}{k}$ we proceed as follows. Inductively we select vertices $x_1, \dots, x_t$ and subsets $X_0 = X \supseteq X_1 \supseteq \dots \supseteq X_t$ such that $x_i \in X_{i-1}$, $X_i \subseteq X_{i-1} - x_i$, and

(*i*) $|X_i| \geq L_i$ and
(*ii*) there exists some $j_i \in [r]$ such that $x_i + K' \in E_{j_i}$ for every $K' \in \binom{X_i}{k-1}$.

Suppose $x_1, \dots, x_i$ and $X_1, \dots, X_i$ were chosen this way. We select x_{i+1} from X_i in an arbitrary way and consider the induced partition of the $(k-1)$-element sets K' of $X_i - x_{i+1}$ given by $K' \in E'_j$ if $(x + K') \in E_j$. Owing to the choice of L_i, there exists a subset X_{i+1} satisfying *(ii)* by Ramsey's theorem for $k-1$.

Finally, we apply the pigeonhole principle to $X' = \{x_1, \dots, x_t\}$ and we obtain an index $j \in [r]$ and a subset $Y' \subseteq X'$ of size $\ell_j - k + 1$ such that $j_i = j$ for every $x_i \in Y'$. It is easy to check that $\binom{Y'+K'}{k} \subseteq E_j$ for every $(k-1)$-element subset $K' \in \binom{L_t}{k-1}$, which concludes the proof. $\square$

The last proof is somewhat wasteful by applying Ramsey's theorem for $k-1$ in a recursive manner. As a consequence this proof gives no reasonable upper bound on $R^{(k)}(\ell_1, \dots, \ell_r)$. In the next proof we will apply the induction assumption on k more carefully. In fact, also this proof (like the last proof) is an extension of the proof for graphs from Section 2.2.2.2. But this time we view the vertices from the $k = 2$ case in Section 2.2.2.2 as $(k-1)$-element sets and not as 1-element sets. This proof of Ramsey's theorem gives the best upper bound for general k (see Section 2.4.1). and goes back to Erdős and Rado [28].

Third proof of Theorem 2.2. We exclude the trivial case and let $r \geq 2$ and $\ell_1, \dots, \ell_r \geq k$. This time we set

$$t = R^{(k-1)}(\ell_1 - 1, \dots, \ell_r - 1) \qquad \text{and} \qquad n = r^{\binom{t}{k-1}} + k - 2\,. \tag{2.4}$$

We shall verify that this n satisfies $n \longrightarrow (\ell_1, \dots, \ell_r)^k$. So let X be an n-element set and consider an arbitrary partition $E_1 \dot\cup \dots \dot\cup E_r = \binom{X}{k}$. Similarly, as in the last proof we select inductively vertices $x_1, \dots, x_t$. In fact, we fix $k-2$ vertices $x_1, \dots, x_{k-2}$ in an arbitrary way and we set $X_{k-2} = X \setminus \{x_1, \dots, x_{k-2}\}$. Next we fix $x_{k-1}, \dots, x_t$ and subsets $X_{k-2} \supseteq X_{k-1} \supseteq \dots \supseteq X_t$ such that $x_i \in X_{i-1}$, $X_i \subseteq X_{i-1} - x_i$, and

(i) $|X_i| \geq (n-k+2)/r^{\binom{i}{k-1}} = r^{\binom{t}{k-1} - \binom{i}{k-1}}$ and

(ii) for every $K' \in \binom{\{x_1,\dots,x_i\}}{k-1}$ there exists a $j_{K'} \in [r]$ such that $K' + y \in E_{j_{K'}}$ for every $y \in X_i$.

Clearly, properties *(i)* and *(ii)* hold for $i = k-2$ and suppose $x_1, \dots, x_i$ and $X_1, \dots, X_i$ were chosen already. We select x_{i+1} from X_i in an arbitrary way. For a k-element set $K \in \binom{X}{k}$ we denote by j_K the index of that partition class from $E_1 \dot\cup \dots \dot\cup E_r$ which contains K. For a vertex $y \in X_i - x_{i+1}$ we consider the vector $\vec{J}_y = (j_{K''+x_{i+1}+y})_{K''}$, where K'' runs over all $(k-2)$-element subsets of $\{x_1, \dots, x_i\}$, i.e., $\vec{J}_y \in [r]^{\binom{\{x_1,\dots,x_i\}}{k-2}}$. Owing to *(i)* (which also ensures $|X_i| > 1$ for $i < t$), combined with $r \geq 2$ and $i \geq k-2$ we have

$$\left\lceil \frac{|X_i| - 1}{r^{\binom{i}{k-2}}} \right\rceil \geq r^{\binom{t}{k-1} - \binom{i}{k-1} - \binom{i}{k-2}} = r^{\binom{t}{k-1} - \binom{i+1}{k-1}}\,.$$

Hence, there exists a vector $\vec{J} \in [r]^{\binom{\{x_1,\dots,x_i\}}{k-2}}$ and a subset $X_{i+1} \subseteq X_i - x_{i+1}$ of size at least $r^{\binom{t}{k-1} - \binom{i+1}{k-1}}$ such that $\vec{J}_y = \vec{J}$ for every $y \in X_{i+1}$. Combining this with the property *(ii)* for all $K' \in \binom{\{x_1,\dots,x_i\}}{k-1}$ we infer that x_{i+1} and X_{i+1} satisfy properties *(i)* and *(ii)*.

Finally, we note that due to the choice of n the set X_t is non-empty and we fix some $x_t \in X_t$ and we consider the set $X' = \{x_1, \dots, x_t\}$. We consider the partition $E'_1 \dot\cup \dots \dot\cup E'_r = \binom{x'}{k-1}$ given by $K' \in E'_j$ if and only if $K' + y \in E_j$. Note that due to Property *(ii)* we also have $K' + x' \in E_j$ for every $x' \in X' \setminus \{x_1, x_2, \dots, x_{i_{k-1}}\}$, where $K' = \{x_{i_1}, \dots, x_{i_{k-1}}\}$ with $i_1 < \dots < i_{k-1}$. Since $|X'| = t$ the choice of t in (2.4) guarantees a subset $Y' \subseteq X'$ and of some index $j \in [r]$ such that $|Y'| \geq \ell_j - 1$ and

$\binom{Y'}{k-1} \subseteq E'_j$. We set $Y = Y' + y$ and claim that Y has the desired properties. In fact, $|Y| \geq \ell_j$. Moreover, let $K \in \binom{Y}{k}$ be arbitrary. Either $K = \{x_{i_1}, \ldots, x_{i_k}\}$ with $i_1 < \cdots < i_k$ or $K = \{x_{i_1}, \ldots, x_{i_{k-1}}, y\}$ for $i_1 < \cdots < i_{k-1}$. In either case, we have $x_{i_1}, \ldots, x_{i_{k-1}} \in Y'$ and, hence, $\{x_{i_1}, \ldots, x_{i_{k-1}}\} \in E'_j$, which yields $K \in E_j$ in both cases. $\square$

2.4. Bounds for the symmetric Ramsey function

Below we discuss lower and upper bounds for the symmetric Ramsey function $R^{(k)}(\ell, \ell)$. We begin with the upper bounds given by the proofs from the last section.

2.4.1. Upper bounds. Comparing the three proofs of Ramsey's theorem given in the last section shows that for $k \geq 3$ the last proof gives the best upper bounds on $R^{(k)}(\ell, \ell)$. For $k = 2$ the first proof gives a slightly better bound. (Recall that for $k = 1$ Ramsey's theorem "degenerates" to the pigeonhole principle, which yields (2.1)). The following notation will be useful. For numbers a and b we set

$$a \uparrow b = a^b\,.$$

Furthermore, for numbers $a_1, \ldots, a_n$ and $n \geq 3$ we define

$$a_1 \uparrow a_2 \uparrow \ldots \uparrow a_n = a_1 \uparrow (a_2 \uparrow (\ldots \uparrow (a_{n-1} \uparrow a_n) \ldots)) = a_1^{a_2^{\cdot^{\cdot^{\cdot^{a_n}}}}}\,.$$

We note that

$$(a_1 \uparrow a_2 \uparrow \ldots \uparrow a_n)^k = a_1^k \uparrow a_2 \uparrow \ldots \uparrow a_n\,. \tag{2.5}$$

Theorem 2.3 (Erdős & Szekeres 1935 and Erdős & Rado 1952).

(*i*) *For all ℓ_1, $\ell_2 \geq 2$ we have $R^{(2)}(\ell_1, \ell_2) \leq \binom{\ell_1+\ell_2-2}{\ell_1-1}$.*

(*ii*) *For all $r \geq 2$, $\ell > k \geq 2$ we have*

$$R^{(k)}(\ell; r) \leq r \uparrow r^{k-1} \uparrow r^{k-2} \uparrow \ldots \uparrow r^2 \uparrow (r(\ell - k) + 1)\,.$$

In particular, for every $k \geq 2$ there exists a constant $C > 0$ such that for every $\ell > k$ we have

$$\log_2^{(k-1)}\left(R^{(k)}(\ell, \ell)\right) \leq C\ell\,, \tag{2.6}$$

where $\log_2^{(k-1)}$ denotes the $(k-1)$-times iterated $\log_2$-function.

Proof. (*i*) We recall the proof of Ramsey's theorem for graphs presented in Section 2.2.2.1 (see also the first proof of Ramsey's theorem in Section 2.3). There it was shown that

$$R^{(2)}(\ell_1, \ell_2) \overset{(2.3)}{\leq} R^{(2)}(\ell_1 - 1, \ell_2) + R^{(2)}(\ell_1, \ell_2 - 1)\,.$$

A straightforward inductive argument yields

$$\begin{aligned} R^{(2)}(\ell_1, \ell_2) &\leq R^{(2)}(\ell_1 - 1, \ell_2) + R^{(2)}(\ell_1, \ell_2 - 1) \\ &\leq \binom{(\ell_1 - 1) + \ell_2 - 2}{(\ell_1 - 1) - 1} + \binom{\ell_1 + (\ell_2 - 1) - 2}{\ell_1 - 1} \\ &= \binom{\ell_1 + \ell_2 - 2}{\ell_1 - 1}, \end{aligned}$$

where for the last identity we used Pascal's rule $\binom{a-1}{j-2} + \binom{a-1}{j-1} = \binom{a}{j}$.

(*ii*) We consider the third proof of Ramsey's theorem from Section 2.3. In fact (2.4) combined with (2.1) yields for $k = 2$

$$R^{(2)}(\ell; r) \le r^{r(\ell-2)+1} + 2 - 2 = r \uparrow (r(\ell-2)+1)\,,$$

which establishes the induction start. A simple induction yields

$$\begin{aligned}
R^{(k)}(\ell; r) &\le \left(r \uparrow \binom{R^{(k-1)}(\ell-1; r)}{k-1}\right) + k - 2 \\
&\overset{(2.4)}{\le} \left(r \uparrow \binom{r \uparrow r^{k-2} \uparrow r^{k-3} \uparrow \ldots \uparrow r^2 \uparrow (r(\ell-1-(k-1))+1)}{k-1}\right) + k - 2 \\
&= \left(r \uparrow \binom{r \uparrow r^{k-2} \uparrow r^{k-3} \uparrow \ldots \uparrow r^2 \uparrow (r(\ell-k)+1)}{k-1}\right) + k - 2 \\
&\overset{(2.5)}{\le} \left(r \uparrow \frac{r^{k-1} \uparrow r^{k-2} \uparrow r^{k-3} \uparrow \ldots \uparrow r^2 \uparrow (r(\ell-k)+1)}{(k-1)!}\right) + k - 2 \\
&\le r \uparrow r^{k-1} \uparrow r^{k-2} \uparrow \ldots \uparrow r^2 \uparrow (r(\ell-k)+1)\,,
\end{aligned}$$

where we used $r \ge 2$ in the last inequality.

□

We remark that for $k = 2$ and $\ell_1 = \ell_2 = \ell$ the bound in (*i*) is better than the one given in (*ii*). Indeed, Stirling's formula yields

$$\binom{2\ell-2}{\ell-1} = (1+o(1))\frac{4^{\ell-1}}{\sqrt{\pi\ell}}\,, \tag{2.7}$$

where $o(1) \to 0$ for $\ell \to \infty$, while

$$2 \uparrow (2(\ell-2)+1) = \frac{4^{\ell-1}}{2}\,.$$

2.4.2. Lower bounds. Proving a lower bound like $R^{(k)}(\ell, \ell) > n$ requires to prove the existence of a partition $E_1 \dot\cup E_2 = \binom{X}{k}$ for an n-element set X such that every ℓ-set $Y \subseteq X$ has the property $\binom{Y}{k} \cap E_1 \neq \emptyset$ and $\binom{Y}{k} \cap E_2 \neq \emptyset$. It seems a natural attempt, to give an explicit example of such a partition.

For example, for $k = 2$ it is easy to show $R^{(2)}(\ell, \ell) > (\ell-1)^2$. For that split X into $\ell - 1$ sets $V_1 \dot\cup \ldots \dot\cup V_{\ell-1} = X$ of size $\ell - 1$. Let E_1 be the set of all pairs which are contained in some V_i for $i \in [\ell-1]$ and, hence, let E_2 consist of all pairs which intersect two of the classes V_i. This way we obtain a quadratic lower bound on $R^{(2)}(\ell, \ell)$. However, the upper bound given in (2.7) is exponential, which leaves a lot of room for improvement for at least one of the bounds.

It turned out to be very hard to establish good constructive lower bounds for $R^{(k)}(\ell, \ell)$ for $k \ge 2$. Abbott [1] found a construction giving $R^{(2)}(\ell, \ell) > c\ell^\alpha$ for $\alpha = \log_2(5) \approx 2.32$ and Nagy [57] gave a cubic bound. The first constructive superpolynomial bound was obtained by Frankl [32] and the currently best known constructive bound due to Frankl and Wilson [34] achieves

$$R^{(2)}(\ell, \ell) \ge \ell^{(1-o(1))\frac{\ln(\ell)}{4\ln\ln(\ell)}}\,.$$

On the other hand, Erdős [19] obtained a much better lower bound in a non-constructive way. Often the proof of this result is presented in the language of basic probability theory. In fact, [19] is considered to be one of the first applications of the *probabilistic method* in combinatorics, which grew into a very active branch in modern combinatorics itself (see, e.g., [4]). Below we present the proof in form of a simple counting argument (similarly as it appears in [19]) and avoid the introduction of necessary notions from probability theory.

THEOREM 2.4 (Erdős 1947). *If* $2\binom{n}{\ell} < 2^{\binom{\ell}{k}}$ *for* $n \geq \ell \geq k$, *then* $R^{(k)}(\ell,\ell) > n$. *In particular, for* $k=2$ *we obtain*

$$R^{(2)}(\ell,\ell) > \frac{\ell}{\mathrm{e}\sqrt{2}}2^{\ell/2} \tag{2.8}$$

and for general k *we have*

$$R^{(k)}(\ell,\ell) > 2^{c\ell^{k-1}} \tag{2.9}$$

for some constant $c = c(k) > 0$.

PROOF. For an n-element set X there are $2^{\binom{n}{k}}$ partitions of $\binom{X}{k}$ into two parts. On the other hand, there are at most $2\binom{n}{\ell}2^{\binom{n}{k}-\binom{\ell}{k}}$ partitions of $\binom{X}{k}$ into two parts such that one of the parts contains all k-element subsets of some ℓ-set $Y \subseteq X$. (We can choose the ℓ-set Y in $\binom{n}{\ell}$ ways, decide which of the two partition classes contains $\binom{Y}{k}$, and there are $2^{\binom{n}{k}-\binom{\ell}{k}}$ ways to distribute $\binom{X}{k} \setminus \binom{Y}{k}$ among the two two partition classes.) The assumption $2\binom{n}{\ell} < 2^{\binom{\ell}{k}}$ yields

$$2\binom{n}{\ell}2^{\binom{n}{k}-\binom{\ell}{k}} < 2^{\binom{n}{k}}.$$

Hence, there exists a partition of $\binom{X}{k}$ into two parts such that no ℓ-set Y has the property that $\binom{Y}{k}$ is contained in one of the partition classes, i.e., $R^{(k)}(\ell,\ell) > n$.

Solving $2\binom{n}{\ell} < 2^{\binom{\ell}{k}}$ for the largest possible n gives (2.9). For the proof of (2.8) the estimate $\binom{n}{\ell} \leq (\frac{en}{\ell})^{\ell}$ is useful. □

Both the lower and the upper bound for $R^{(2)}(\ell,\ell)$ given in (2.7) and (2.8) are exponential in ℓ, while the lower and upper bounds for $R^{(k)}(\ell,\ell)$ for $k \geq 3$ given in *(ii)* of Theorem 2.3 and (2.9) are away by $(k-2)$-times iterated exponential functions.

Despite a lot of effort those bounds for $k=2$ were only improved slightly. The best known lower bound is based on a more refined probabilistic tool (the so-called *local lemma* from [26]) given by Spencer [82] and improves (2.8) by a factor of two.

The upper bound was improved by a constant factor by Frasnay [35]. The first substantial improvement is due to Rödl (unpublished, see [42]), which was further improved by Thomason [87]. Currently the best bound is due to Conlon [7]. The best known bounds for $k=2$ are

$$\left(\frac{\sqrt{2}}{\mathrm{e}} - o(1)\right)\ell 2^{\ell/2} \leq R^{(2)}(\ell,\ell) \leq \ell^{-C\ln\ell/\ln\ln\ell}\binom{2\ell-2}{\ell-1}, \tag{2.10}$$

for some constant $C > 0$.

For hypergraphs the gap between the bounds in *(ii)* of Theorem 2.3 and (2.9) could be closed up to one exponential in the iterated exponential function. For $k \geq 3$ the following result of Erdős, Hajnal, and Rado [25, Lemma 6, p.140] (see also [24, Theorem 26.3]) "lifts" a lower bound for $R^{(k)}(\ell,\ell)$ by one additional exponential to a lower bound for $R^{(k+1)}(\ell',\ell')$ (for some appropriate ℓ'). Moreover, it is known that for $k=2$ a similar result is true, if $r \geq 4$.

THEOREM 2.5 ((negative) stepping-up lemma).

(i) *If* $k \geq 3$ *and* $R^{(k)}(\ell,\ell) > n$, *then* $R^{(k)}(\ell',\ell') > 2^n$ *where* $\ell' = 2\ell + k - 4$.
(ii) *If* $R^{(2)}(\ell;4) > n$, *then* $R^{(3)}(\ell+1;4) > 2^n$. □

For the proof we refer the reader to [24, Theorem 26.3] (see also [44, Section 4.7]).

The negative stepping-up lemma combined with (2.9) for $k=3$ yields a much better lower bound for $R^{(k)}(\ell,\ell)$ for $k\geq 4$. In fact, we obtain

$$R^{(k)}(\ell,\ell) > \underbrace{2\uparrow\ldots\uparrow 2\uparrow}_{(k-2)\text{-times}} c\ell^2\,,$$

for every $k\geq 3$ and constants $c=c(k)>0$. Together with (2.6) we obtain for every $k\geq 3$ some constants c, $C>0$ depending only on k such that

$$c\ell^2 < \log_2^{(k-2)}\big(R^{(k)}(\ell,\ell)\big) < 2^{C\ell}\,. \tag{2.11}$$

It turned out to be a very hard problem to close the exponential gap in (2.11). Erdős, Hajnal, and Rado [25, Section 16] conjectured that the lower bound can be improved to match the upper bound (up to the constant C) and (ii) of Theorem 2.5 gives some evidence for this believe. In view of (i) of Theorem 2.5 it would suffice to settle the following conjecture.

Conjecture 2.6 (Erdős, Hajnal & Rado 1965). There is some $c>0$ such that

$$R^{(3)}(\ell,\ell)\geq 2\uparrow 2\uparrow c\ell = 2^{2^{c\ell}}\,.$$

Conjecture 2.6 is one of the major open problems related to Ramsey's theorem and Erdős offered \$ 500 for its solution [20].

2.5. Infinite version of Ramsey's theorem

We briefly discuss the so-called *infinite version* of Ramsey's theorem, which was already considered in [75].

Theorem 2.7 (Ramsey 1930). *For all integers r, $k\geq 1$, and every (countably) infinite set X the following holds. For any partition $E_1\dot\cup\ldots\dot\cup E_r=\binom{X}{k}$ of the k-element subsets of X, there exists an index $j\in[r]$ and an infinite subset $Y\subseteq X$ such that $\binom{Y}{k}\subseteq E_j$.*

Clearly, Theorem 2.7 implies Theorem 1.1 and, hence, it also implies Theorem 2.1 by the compactness principle. However, while Theorem 2.1 implies Theorem 1.1, it does not immediately yield Theorem 2.7. This is because Theorem 2.7 asserts the existence of an infinite monochromatic set Y for every finite partition of $\binom{X}{k}$ for a countably infinite set X, while a straight forward application of Theorem 2.1 only yields the existence of monochromatic subsets Y of arbitrary (unbounded) size in an infinite set X.

However, the second and the third proof of the finite version of Ramsey's theorem given in Section 2.3 can be adjusted to prove Theorem 2.7. Again the proof proceeds by induction on k. Roughly speaking, we simply have to ensure the existence of infinite sequences $(x_i)_{i\in\mathbb{N}}$ and $(X_i)_{i\in\mathbb{N}}$ where X_i is always an infinite set (instead of a large finite set satisfying the lower bounds stated in (i) in those proofs). In the end, we conclude the proof by an application of the infinite version of the pigeonhole principe (Theorem 2.7 for $k=1$) in case of the second proof and by an application of the Theorem 2.7 for $k-1$ in case of the third proof. We omit the details here. Such an infinite version of the second proof from Section 2.3 can be found for example in [15, Proof of Theorem 9.1.2].

Below we give a proof of Theorem 2.7 based on ultrafilters (see Appendix A.3 for the necessary background). Ideas connecting ultrafilters with Ramsey theory can be traced back to Fred Galvin in connection with Hindman's theorem [47]. The earliest proofs of the infinite Ramsey theorem based on ultrafilter we could find are [6, Theorem 3.3.7.] and [49, Problem 7.5.1]. The proof given below follows the presentation from [24, Section 10].

Proof of Theorem 2.7. Let integers r and k be given. Theorem 2.7 is trivial if $r = 1$. We proceed by induction on k. The theorem follows from the pigeonhole principle for $k = 1$, which establishes the induction start. Hence, we may assume that $r \geq 2$ and $k \geq 2$ and that the theorem holds for $k - 1$.

Let X be a countably infinite set. Fix a non-principal ultrafilter $\mathcal{F}$ on X, which is guaranteed to exist by Theorem A.30.

Let $E_1 \dot\cup \ldots \dot\cup E_r$ be an arbitrary partition of $\binom{X}{k}$. For any $(k-1)$-set $K \in \binom{X}{k-1}$ the sets

$$A_{j,K} = \{x \in X \setminus K \colon\ (K + x) \in E_j\}$$

for $j \in [r]$ form a partition of $X \setminus K$. Since $\mathcal{F}$ is a non-principal ultrafilter and $X \setminus K$ is co-finite, the set $(X \setminus K) \in \mathcal{F}$ (see Proposition A.29). Moreover, it follows from Proposition A.26 there exists a unique index $j_K \in [r]$ such that $A_{j_K,K} \in \mathcal{F}$.

We claim that there exists an infinite set $X' = \{x_1, x_2, \ldots\}$ such that

($*$) for every $(k-1)$-set $K = \{x_{i_1}, \ldots, x_{i_{k-1}}\} \in \binom{X'}{k-1}$ with $i_1 < \cdots < i_{k-1}$ and every $x_{i_k} \in X'$ with $i_k > i_{k-1}$ we have $(K + x_{i_k}) \in E_{j_K}$, i.e., $x_{i_k} \in A_{j_K,K}$.

For that we show that every finite set X' with property ($*$) can be enlarged, which means that that there are no maximal finite sets with this property and, therefore, there must be an infinite set with this property. So let X' be a finite set satisfying ($*$). Consequently,

$$X'' = \bigcap_{K \in \binom{X'}{k-1}} A_{j_K,K}$$

is a finite intersection of sets from the filter and, hence, $X'' \in \mathcal{F}$. In particular, X'' is infinite (since $\mathcal{F}$ is a non-principal ultrafilter) and, hence, there exists some $x \in X'' \setminus X'$. It follows from the definition of x, that $X' + x$ has property ($*$).

Let X' be an infinite set with property ($*$). We consider the corresponding partition $E'_1 \dot\cup \ldots \dot\cup E'_r$ of $\binom{X'}{k-1}$, where we let E'_j consist of those $(k-1)$-sets $K \in \binom{X'}{k-1}$ with $j_K = j$. Owing to the induction assumption there exists an infinite subset $Y \subseteq X$ and an index $j \in [r]$ such that $\binom{Y}{k-1} \subseteq E'_j$. Since $Y \subseteq X'$, it posseses property ($*$) and, hence, $\binom{Y}{k} \subseteq E_j$. $\square$

CHAPTER 3

Arithmetic progressions

This chapter is devoted to van der Waerden's theorem (Theorem 1.4) and its extensions. We will give the combinatorial proof of van der Waerden in Section 3.1. In Section 3.2 we present a topological proof of Theorem 1.4.

In Sections 3.3.1 and 3.3.2 we consider the first nontrivial case of Szemerédi's theorem (Theorem 1.6 for $k = 3$), which is due to Roth [**77, 78**]. We give a combinatorial proof due to Szemerédi in Section 3.3.1 and in Section 3.3.1 we follow the analytical argument of Roth from [**78**].

3.1. Combinatorial proof of van der Waerden's theorem

For $r = 2$ Theorem 1.4 was conjectured by Schur and van der Waerden heard it from Baudet. Van der Waerden discussed the problem with Schreier and Artin during his research stay in Hamburg in 1926/27. In those discussions it turned out the following finite version of Theorem 1.4 is helpful for an inductive proof (see, e.g., [**90, 91**] for more details).

Theorem 3.1 (van der Waerden 1927). *For all integers $r \geq 1$ and $k \geq 1$ there exists some integer n_0 such that for every $n \geq n_0$ the following holds. For any partition $E_1 \dot\cup \ldots \dot\cup E_r = [n]$ there exists some $j \in [r]$ such that E_j contains an arithmetic progression of length k, i.e., there exists an $a \in \mathbb{N}$ and $\lambda > 0$ such that $a + i\lambda \in E_j$ for every $i = 0, \ldots, k-1$.*

We note that an infinite version of Theorem 3.1 in the spirit of of the infinite version of Ramsey's theorem (Theorem 2.7) is obviously not true. For example, split $\mathbb{N}$ into two classes so that every class contains intervals of unbounded size (e.g., let $E_1 = \bigcup_{i \in \mathbb{N}_0}\{2^{2i}, \ldots, 2^{2i+1}-1\}$ and $E_2 = \mathbb{N} \setminus E_1$). Suppose that one of the classes contains an *infinite* arithmetic progression of the form $\{a + i\lambda \colon i \in \mathbb{N}_0\}$ for a and $\lambda \in \mathbb{N}$. Since such an arithmetic progression cannot "jump" over an interval of length at least λ from the other class, but each class contains infinitely many intervals of length at least λ, we derive at a contradiction.

We denote by $W(k; r)$ the smallest integer n for which the conclusion of Theorem 3.1 holds. Clearly, $W(k; 1) = k$ for every integer $k \geq 1$, and for $r \geq 1$ we have $W(1; r) = 1$ and $W(2; r) = r + 1$. The first non-trivial case concerns $W(3; 2)$. The partition given by $E_1 = \{1, 2, 5, 6\}$ and $E_2 = \{3, 4, 7, 8\}$ shows that $W(3; 2) > 8$. A simple (but somewhat tedious) case analysis shows indeed $W(3; 2) \leq 9$ and, hence, $W(3; 2) = 9$. Below we show the considerably weaker bound

$$W(3; 2) \leq 325\,. \tag{3.1}$$

However, this proof can be generalises and leads to van der Waerden's original proof of Theorem 3.1. For the proof of (3.1) it is helpful to note

$$325 = 65 \cdot 5 = (2 \cdot 2^5 + 1) \cdot 5$$

where the 5 comes from $5 = 2W(2; 2) - 1$. Let $E_1 \dot\cup E_2 = [325]$ be some partition. We split $[325]$ into 65 intervals of size 5, i.e., $[365] = \dot\bigcup_{i=1}^{65} I_i$ for $I_i = \{5(i-1)+1, \ldots, 5i\}$. Such an interval of can intersect E_1 and E_2 in $2^5 = 32$ different ways. Consequently, there are at least 2 intervals among the first 33 intervals which intersect E_1 and E_2 in the same way, i.e., I_{i_1} and I_{i_2} have the same pattern according to the partition of I_{i_1} and I_{i_2} induced by E_1 and E_2. More precisely, there exist indices $1 \leq i_1 < i_2 \leq 33$ such that $5(i_1 - 1) + \ell$ and $5(i_2 - 1) + \ell$ are in the same partition class E_1 or E_2 for every $\ell \in [5]$.

Every partition of an interval of length 5 into two classes there exists an arithmetic progression of length 3 such that the first and the second member are from the same class. In fact, among the first three elements of such an interval two elements are in the same class and the third number, which completes this pair to a three term arithmetic progression, is also contained in this interval of length 5.

Let $1 \leq \ell_1 < \ell_2 < \ell_3 \leq 5$ be the indices of this three term arithmetic progression in I_{i_1} for the partition E_1 and E_2. Without loss of generality we may assume that $5(i_1-1)+\ell_1$ and $5(i_1-1)+\ell_2$ are contained in E_1 and $5(i_1-1)+\ell_3 \in E_2$. Since I_{i_1} and I_{i_2} have the same pattern, we also have $5(i_2-1)+\ell_1$, $5(i_2-1)+\ell_2 \in E_1$ and $5(i_2-1)+\ell_3 \in E_2$.

Finally, we consider the l_3-rd element from I_{i_3} for $i_3 = i_2 + (i_2 - i_1)$. Note that $i_3 \leq 65$, since $i_2 \leq 33$, and $5(i_3 - 1) + \ell_3 \leq 325$. Furthermore, we observe that $5(i_3 - 1) + \ell_3$ is the third element for the two three term arithmetic progressions

$$5(i_1 - 1) + \ell_1\,, \quad 5(i_2 - 1) + \ell_2\,, \quad 5(i_3 - 1) + \ell_3$$

and

$$5(i_1 - 1) + \ell_3\,, \quad 5(i_2 - 1) + \ell_3\,, \quad 5(i_3 - 1) + \ell_3\,.$$

Since $5(i_1-1)+\ell_1$, $5(i_2-1)+\ell_2 \in E_1$ and $5(i_1-1)+\ell_3$, $5(i_2-1)+\ell_3 \in E_2$, it follows that either E_1 or E_2 contains a three term arithmetic progression and we conclude (3.1).

Roughly speaking, in the argument given above, we ensured the existence of two arithmetic progressions of length two one in E_1 and one in E_2, which *focused* on the same number (on $5(i_3 - 1) + \ell_3$). In some sense in the proof we used the (trivial) fact that the $W(2;r) \leq r+1$ is true for $r = 32$. So we used the induction assumption (on k) for a much bigger number of partition classes (32 instead of 2). In order to generalize this proof we have to consider how to establish $W(3;r)$ for $r > 2$. Iterating the proof of (3.1) yields

$$W(3;3) \leq (2 \cdot 3^{(2 \cdot 3^7 + 1)\cdot 7} + 1)(2 \cdot 3^7 + 1) \cdot 7$$

and, more generally, for $W(3;r)$ we set $w_1 = 2 \cdot W(2;r) - 1 = 2r + 1$ and for $j = 2, \ldots, r$

$$W_{j-1} = \prod_{i=1}^{j-1} w_i \qquad \text{and} \qquad w_j = 2 \cdot r^{W_{j-1}} + 1 = 2 \cdot W(2; r^{W_{j-1}}) - 1$$

and obtain

$$W(3;r) \leq W_r\,. \tag{3.2}$$

A proof along the lines the $W(3;2)$ yields inductively that for every $j = 1\ldots, r$ and every partition $E_1 \dot\cup \ldots \dot\cup E_r$ of an interval of length W_j the existence of j arithmetic progression of length three with the same endpoint z_j where the first two elements from each of those arithmetic progressions are contained in a different class E_i. In other words, if we assume that there exists no three term arithmetic progression completely contained in one of the classes, then at least j of the classes E_i are *forbidden* for the point z_j.

In the next step we split $[n]$ into blocks of length W_j and note that there are at most r^{W_j} different intersection patters for those intervals with the partition $E_1 \dot\cup \ldots \dot\cup E_r$. Moreover, among $w_{j+1} = 2 \cdot r^{W_j} + 1 = 2 \cdot W(2; r^{W_j}) - 1$ consecutive intervals of this length we find two with the same pattern and a third which is in a three term arithmetic progression with the first two intervals, i.e., the first element of all three intervals form a three term arithmetic progression. In particular, the three copies of z_j form a three term arithmetic progression and since the first two copies belong to the same partition class, say E_i, now there exist also two elements from E_i, which this third copy of z_j completes to a three term arithmetic progression. Let this third copy be z_{j+1}. Moreover, one can still ensure that the other j classes E_i are still forbidden for z_{j+1}. In fact, if x and y were such a pair in the interval of length W_j, i.e., x, y, z_j form a three term arithmetic progression and x, y are in the same class from $E_1\dot\cup\dots\dot\cup E_r$. Then the copy of x in the first of the two equally partitioned intervals of length W_j and the copy of y from the second such interval are still both contained in the same class from the partition $E_1\dot\cup\dots\dot\cup E_r$ and those copies of x and y form an arithmetic progression with $z_{\ell+1}$.

We have shown that for every partition of an interval of length $w_{j+1}\cdot W_j = W_{j+1}$ there exists a number z_{j+1} with the property that at least $(j+1)$ three term arithmetic progressions end in z_{j+1} with the property that the first two elements for each of those progressions are contained in in the same class, but for diffrenent classes for different progressions. This establishes the induction step and for $j = r$ it implies (3.2). In other words, we proved Theorem 3.1 for $k = 3$ and every $r \geq 1$.

It is left to generalize the argument for arbitrary k. However, also this is straight forward. We proceed by induction on k and assume Theorem 3.1 holds for $k-1$ and every $r \geq 1$. For fixed $r \geq 2$ we repeat the definition from (3.2), but we can replace 2 by $\frac{k-1}{k-2}$ and have to replace $W(2;r)$ by $W(k-1;r)$. (Terms of the form $Z = 2\cdot Y - 1 = 2\cdot(Y-1)+1$ ensured that the third point of a three term arithmetic progression with the first two being contained in an interval of length Y are contained in an interval of length Z. Now we need that, the k-th element of a k-term arithmetic progression with the first $k-1$ be contained in an interval of length Y are contained in an interval of length Z, which is guaranteed by $Z = (k-1)\lfloor\frac{Y-1}{k-2}\rfloor + 1$.)

We set $w_1 = (k-1)\lfloor\frac{W(k-1;r)-1}{k-2}\rfloor + 1$ and for $j = 2,\dots,r$

$$W_{j-1} = \prod_{i=1}^{j-1} w_i \qquad \text{and} \qquad w_j = (k-1)\left\lfloor\frac{W(k-1;r^{W_{j-1}})-1}{k-2}\right\rfloor + 1$$

and claim

$$W(k;r) \leq W_r\,. \tag{3.3}$$

It is not hard to verify (3.3) along the lines the proof of (3.2) and we omit the details here. In fact, such a proof was carried out in the original work of van der Waerden [89]. Below we give essentially the same proof in a more concise form due to Graham and Rothschild [43].

Proof of Theorem 3.1. For two integers $a \leq b \in \mathbb{Z}$ we write $[a,b]$ for the interval $\{a,\dots,b\}$. For $k\in\mathbb{N}$ and $m\in\mathbb{N}$ we define an equivalence relation $\sim_k$ on $[0,k]^m$ by

$$(x_1,\dots,x_m) \sim_k (y_1,\dots,y_m)$$

if there exists some $i_0 \in [0,m]$ such that $x_i = y_i$ for $i \in [i_0]$ with $x_{i_0} = y_{i_0} = k$ and $x_{i_0} < k$ and $y_{i_0} < k$ for every $i \in [i_0+1,m]$. We note that the equivalence classes of $\sim_k$ are of the form

$$[0,k-1]^m$$

if $i_0 = 0$ and for $i_0 \geq 1$ they are of the form

$$\vec{z}_{i_0-1} \times \{k\} \times [0,k-1]^{m-i_0}$$

with $\vec{z}_{i_0-1} \in [0,k]^{i_0-1}$.

Below we verify the following statement for all integers k and $m \geq 1$

$\boldsymbol{vdW}(k,m)$**:** *For every integer $r \geq 1$ there exists some integer $N = N(k,m,r)$ such that for every partition $E_1 \dot\cup \ldots \dot\cup E_r$ of $[N]$ there exist a, $\lambda_1, \ldots, \lambda_m \in \mathbb{N}$ with $a + k\sum_{i=1}^m \lambda_i \leq N$ such that for every equivalence class X of $\sim_k$ on $[0,k]^m$ there exists some $j \in [r]$ such that $f_{a,\lambda_1,\ldots,\lambda_m}(X) \subseteq E_j$, where*

$$f_{a,\lambda_1,\ldots,\lambda_m}(x_1,\ldots,x_m) = a + \sum_{i=1}^m x_i\lambda_i\,.$$

For $k = m = 1$ there are two equivalence classes of $\sim_1$ on $\{0,1\}$ and each of them consist of exactly one element. Hence, $\boldsymbol{vdW}(1,1)$ is trivial and $N(k,m,r) = 2$ suffices for every $r \geq 1$.

Moreover, we note that $\boldsymbol{vdW}(k,1)$ implies Theorem 3.1 for k. In fact, $[0,k-1]$ and $\{k\}$ are the only two equivalence classes of $\sim_k$ on $[0,k]$. Statement $\boldsymbol{vdW}(k,1)$ guarantees for every $r \geq 1$ a constant $N = N(k,1,r)$ such that for every partition $E_1 \dot\cup \ldots \dot\cup E_r$ of $[N]$ there exist a and $\lambda \in \mathbb{N}$ and an index $j \in [r]$ such that

$$f_{a,\lambda}([0,k-1]) = \{a,\, a+\lambda,\, a+2\lambda, \ldots,\, a+(k-1)\lambda\} \subseteq E_j\,,$$

i.e., E_j contains a k-term arithmetic progression. Consequently, Theorem 3.1 follows from Lemmas 3.2 and 3.3 below. $\square$

Lemma 3.2. *If $\boldsymbol{vdW}(k,1)$ holds, then $\boldsymbol{vdW}(k,m)$ holds for every $m \geq 1$.*

Proof of Lemma 3.2. We prove Lemma 3.2 by induction on m. The assumption of Lemma 3.2 states that $\boldsymbol{vdW}(k,1)$ holds, which establishes the induction start. So we assume $\boldsymbol{vdW}(k,m)$ holds and we will deduce $\boldsymbol{vdW}(k,m+1)$.

Let $r \geq 1$ be fixed and let $N_1 = N(k,m,r)$ and $N_2 = N(k,1,r^{N_1})$ be given by induction assumption. We set

$$N(k,m+1,r) = N_1 \cdot N_2\,.$$

Let $E_1 \dot\cup \ldots \dot\cup E_r$ be a partition of $[N]$ for $N = N_1N_2$. We split $[N]$ into N_2 intervals $I_1, \ldots, I_{N_2}$ of length N_1. We say two intervals I_{i_1} and I_{i_2} have the same *pattern* under the partition $E_1 \dot\cup \ldots \dot\cup E_r$, if for every $\ell \in [N_1]$ the numbers $(i_1-1)N_1+\ell$ and $(i_2-1)N_1+\ell$ are contained in the same class of the partition $E_1 \dot\cup \ldots \dot\cup E_r$. There are $R = r^{N_1}$ different pattern. We fix some enumeration of all patterns and we consider the following auxiliary partition $E'_1 \dot\cup \ldots \dot\cup E'_R = [N_2]$, where E'_J contains those indices $i \in [N_2]$ for which I_i has the J-th pattern (of the fixed enumeration).

Owing to the definition of $N_2 = N(k,1,R)$ the statement $\boldsymbol{vdW}(k,1)$ guarantees the existence of A and $\Lambda \in \mathbb{N}$ such that $A + k\Lambda \leq N_2$ and for each of the two equivalence classes $[0,k-1]$ and $\{k\}$ of $\sim_k$ on $[0,k]$ there exist J_1, $J_2 \in [R]$ such that $f_{A,\Lambda}([0,k-1]) \subseteq E'_{J_1}$ and $f_{A,\Lambda}(k) \subseteq E'_{J_2}$. In particular, there exists some pattern (the J_1-st pattern) such that all blocks I_i with $i \in f_{A,\Lambda}([0,k-1])$ have this pattern.

Moreover, owing to the choice of $N_1 = N(k,m,r)$ we can apply $\boldsymbol{vdW}(k,m)$ to one (and hence to all) block(s) I_i with $i \in f_{A,\Lambda}([0,k-1])$ and we get a_0, $\lambda_1, \ldots, \lambda_m$ with $a_0 + k\sum_{i=1}^m \lambda_i \leq N_1$ such that for every equivalence class X of $\sim_k$ on $[0,k]^m$ there exists $j \in [r]$ such that

$$f_{a_0,\lambda_1,\ldots,\lambda_m}(X) \subseteq E_j\,.$$

We set

$$a = (A-1)N_1 + a_0 \quad \text{and} \quad \lambda_{m+1} = \Lambda \cdot N_1$$

and note that

$$\begin{aligned} a + k\sum_{i=1}^{m+1}\lambda_i = a + k\sum_{i=1}^{m}\lambda_i + k\Lambda N_1 &= (A-1)N_1 + \left(a_0 + k\sum_{i=1}^{m}\lambda_i\right) + k\Lambda N_1 \\ &\leq (A-1)N_1 + N_1 + k\Lambda N_1 = (A+k\Lambda)N_1 \leq N_2N_1 = N\,. \end{aligned}$$

We will establish $\boldsymbol{vdW}(k, m+1)$ by showing that for every equivalence class Y of $\sim_k$ on $[0,k]^{m+1}$ there exists $j \in [r]$ such that

$$f_{a,\lambda_1,\dots,\lambda_{m+1}}(Y) \subseteq E_j\,. \tag{3.4}$$

Note that elements $(x_1,\dots,x_{m+1}) \in [0,k]^{m+1}$ with $x_{m+1} = k$ form trivial equivalence classes in $\sim_k$ on $[0,k]^{m+1}$, which consist of only one elements. Hence, (3.4) holds for those equivalence classes.

So let Y be an equivalence class contained in $[0,k]^m \times [0,k-1]\}$. Owing to the application of $\boldsymbol{vdW}(k,1)$ (which ensures the repetition the same pattern on the intervals I_i of length N_1 with i of the form $A + x\Lambda$ for $x \in [1,k-1]$) we infer that for every $\ell \in [N_1]$ there exists some $j \in [r]$ such that

$$\begin{aligned}\big\{(A-1)N_1 + \ell + x\lambda_{m+1} \colon x \in [0,k-1]\big\}&\\ = \big\{(A-1)N_1 + \ell + x\Lambda N_1 \colon x \in [0,k-1]\big\} &\subseteq E_j\,.\end{aligned}$$

For given $(x_1,\dots,x_{m+1})$ from an equivalence class contained in $[1,k]^m \times [0,k-1]$ we apply the last observation with $\ell = a_0 + \sum_{i=1}^m x_i\lambda_i \le N_1$ and obtain that

$$a + \sum_{i=1}^{m+1} x_i\lambda_i = (A-1)N_1 + \left(a_0 + \sum_{i=1}^{m} x_i\lambda_i\right) + x_{m+1}\lambda_{m+1}$$

and

$$a + \sum_{i=1}^{m} x_i\lambda_i = (A-1)N_1 + \left(a_0 + \sum_{i=1}^{m} x_i\lambda_i\right)$$

are contained in the same partition class, say E_j. Moreover, it follows from the application of $\boldsymbol{vdW}(k,m)$ that also $(a + \sum_{i=1}^m y_i\lambda_i) \in E_j$ for every vector $(y_1,\dots,y_m) \in [0,k]^m$ with the property $(x_1,\dots,x_m) \sim_k (y_1,\dots,y_m)$. Since for all vectors $(x_1,\dots,x_m,x_{m+1})$, $(y_1,\dots,y_m,y_{m+1}) \in Y \subseteq [1,k]^m \times [0,k-1]$ we have $(x_1,\dots,x_m) \sim_k (y_1,\dots,y_m)$ on $[1,k]^m$ assertion (3.4) follows. □

It is left to prove the following lemma, which allows an induction on k.

Lemma 3.3. *If $\boldsymbol{vdW}(k,m)$ holds for every $m \ge 1$, then $\boldsymbol{vdW}(k+1,1)$ holds.*

Proof of Lemma 3.3. Let $r \ge 1$. Set $N = N(k+1,1,r) = 2N(k,r,r)$, where $N(k,r,r)$ is given by $\boldsymbol{vdW}(k,m)$ applied with $m = r$. Let $E_1\dot\cup\dots\dot\cup E_r = [N]$ be arbitrary. The statement $\boldsymbol{vdW}(k,m)$ yields integers a, $\lambda_1,\dots,\lambda_r \in \mathbb{N}$ such that $a + k\sum_{i=1}^r \lambda_i \le N(k,r,r)$ and for every equivalence class X of $\sim_k$ on $[0,k]^r$ we have $f_{a,\lambda_1,\dots,\lambda_r}(X) \subset E_j$ for some $j \in [r]$.

It follows from the pigeonhole principle that there exist numbers s, $t \in [0,r]$ with $s < t$ such that $a + k\sum_{i=1}^s \lambda_i$ and $a + k\sum_{i=1}^t \lambda_i$ are contained in the same partition class, say E_j. We set

$$A = a + k\sum_{i=1}^{s}\lambda_i \quad\text{and}\quad \Lambda = \sum_{i=s+1}^{t}\lambda_i$$

and claim that $f_{A,\Lambda}([0,k]) \subseteq E_j$. Since $[0,k]$ is the only non-trivial equivalence class of $\sim_{k+1}$ on $[0,k+1]$ and $A + (k+1)\Lambda \le a + (k+1)\sum_{i=1}^r \lambda_i \le 2N(k,r,r) = N$ this establishes $\boldsymbol{vdW}(k+1,1)$ for r.

However, for every $x \in [0,k-1]$ we have

$$A + x\Lambda = \left(a + k\sum_{i=1}^{s}\lambda_i\right) + x\left(\sum_{i=s+1}^{t}\lambda_i\right) \subseteq f_{a,\lambda_1,\dots,\lambda_r}(X)\,,$$

where X is the equivalence class $\{k\}^s \times [0, k-1]^{r-s}$ of $\sim_k$ on $[0, k]^r$. Hence, there exists some j such that $A + x\Lambda \in E_j$ for every $x \in [0, k-1]$. Finally, we note that also $A + k\Lambda \in E_j$, since $A = A + 0 \cdot \Lambda \in E_j$ and since

$$A + k\Lambda = a + k\sum_{i=1}^{t} \lambda_i \quad \text{and} \quad A = a + k\sum_{i=1}^{s} \lambda_i$$

are in the same partitions class due to the choice of s and t above. $\square$

3.2. Topological proof of van der Waerden's theorem

In this section we give another proof of van der Waerden's theorem. This proof uses ideas from topological dynamics (see Appendix A.2 for the necessary background). Roughly speaking, we deduce Theorem 1.4 from a *multiple recurrence theorem* (see Section 3.2.2), which can be viewed as a strengthening of Theorem A.17. Such a proof of van der Waerden's theorem first appeared in the work of Furstenberg and Weiss [38] (see also [37]) and it grew out of Furstenberg's proof of Szemerédi's theorem [36].

3.2.1. The topological van der Waerden theorem. We first state the following recurrence result, which can be viewed as a topological version of Theorem 1.4.

THEOREM 3.4 (topological van der Waerden theorem). *Let (X, T) be a dynamical system, where X is a compact metric space with metric ϱ and let $k \in \mathbb{N}$. There exists some $x \in X$ such that for every $\varepsilon > 0$ there exists an $n \in \mathbb{N}$ such that*

$$\varrho(x, T^{in}(x)) < \varepsilon$$

simultaneously for every $i \in [k]$.

In Section 3.2.2 we prove the multiple recurrence theorem (Theorem 3.6) which is more general than the topological van der Waerden theorem. Below we first deduce Theorem 1.4 from Theorem 3.4. In this proof the following metric space will play a prominent rôle.

Let A be a finite set. We define a metric ϱ on the set of all functions $\boldsymbol{x}$, $\boldsymbol{y}\colon \mathbb{N} \to A$ as follows

$$\varrho(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} 0\,, & \text{if } \boldsymbol{x}(n) = \boldsymbol{y}(n) \text{ for every } n \in \mathbb{N}, \\ \frac{1}{m}\,, & \text{where } m = \min\{n \in \mathbb{N}\colon\ \boldsymbol{x}(n) \neq \boldsymbol{y}(n)\}. \end{cases} \tag{3.5}$$

It is straightforward to check that $(A^{\mathbb{N}}, \varrho)$ is a metric space and we omit this here. In fact, more is true and below we show that the metric ϱ induces the product topology (of the discrete topologies) on $A^{\mathbb{N}}$. Indeed, for $\boldsymbol{x} \in A^{\mathbb{N}}$ and $\varepsilon > 0$ we see that the open ball

$$\mathcal{B}_\varepsilon(\boldsymbol{x}) = \{\boldsymbol{y} \in A^{\mathbb{N}}\colon\ \varrho(\boldsymbol{x}, \boldsymbol{y}) < \varepsilon\}$$

in the metric space $(A^{\mathbb{N}}, \varrho)$ consists precisely of those functions $\boldsymbol{y} \in A^{\mathbb{N}}$ for which $\boldsymbol{y}(j) = \boldsymbol{x}(j)$ for every $j = 1, \dots, \lfloor 1/\varepsilon \rfloor$. Consequently,

$$\mathcal{B}_\varepsilon(\boldsymbol{x}) = \{\boldsymbol{x}(1)\} \times \dots \times \{\boldsymbol{x}(\lfloor 1/\varepsilon \rfloor)\} \times \prod_{n > 1/\varepsilon} A\,,$$

which is in the basis of product topology on $A^{\mathbb{N}}$. Conversely, since A is finite, every basic open set in the product topology on $A^{\mathbb{N}}$ can be written as the finite union of sets of the form $\{a_1\} \times \dots \times \{a_i\} \times \prod_{n>i} A$ with $a_1, \dots, a_i \in A$. In other words, every basic open set in product topology on $A^{\mathbb{N}}$ can be written as the union of open balls of the form $\mathcal{B}_\varepsilon(\boldsymbol{x})$ for some $\boldsymbol{x} \in A^{\mathbb{N}}$ and $\varepsilon > 0$. Hence, we have shown that the metric ϱ induces the product topology on $A^{\mathbb{N}}$ and owing to Tychonoff's theorem (Theorem A.10) the metric space $(A^{\mathbb{N}}, \varrho)$ is compact.

We shall consider the *left shift* on $A^{\mathbb{N}}$ defined by

$$(T \circ \boldsymbol{x})(n) = \boldsymbol{x}(n+1) \tag{3.6}$$

for every $\boldsymbol{x} \in A^{\mathbb{N}}$ and every $n \in \mathbb{N}$. Clearly, if $\varrho(\boldsymbol{x}, \boldsymbol{y}) \leq 1/(n+1)$ for some $n \in \mathbb{N}$, then $\varrho(T(\boldsymbol{x}), T(\boldsymbol{y})) \leq 1/n$ and, therefore, T is uniformly continuous on $(A^{\mathbb{N}}, \varrho)$. We summarise the above in the following proposition.

PROPOSITION 3.5. *Let A be a finite set and let $\varrho\colon\ A^{\mathbb{N}} \to A^{\mathbb{N}}$ be defined by (3.5). Then $(A^{\mathbb{N}}, \varrho)$ is a compact metric space and the induced topology is the product topology of the discrete topology on A. Moreover, the left shift $T\colon\ A^{\mathbb{N}} \to A^{\mathbb{N}}$ defined in (3.6) is uniformly continuous.* □

After these preparations we prove van der Waerden's theorem from the topological version.

PROOF: THEOREM 3.4 ⇒ THEOREM 1.4. Let integers $r \geq 1$ and $k \geq 1$ be fixed and let $E_1 \dot\cup \ldots \dot\cup E_r$ be some partition of $\mathbb{N}$. We interpret this partition as an element $\boldsymbol{x}$ from $[r]^{\mathbb{N}}$ defined by $\boldsymbol{x}^{-1}(j) = E_j$ for every $j \in [r]$. Let ϱ be the metric on $[r]^{\mathbb{N}}$ given in (3.5) and let T be the left shift defined in (3.6). We let X be the closure in $([r]^{\mathbb{N}}, \varrho)$ of the orbit of $\boldsymbol{x}$ under T, i.e.,

$$O_{\boldsymbol{x}} = \{T^n(\boldsymbol{x})\colon\ n \in \mathbb{N}\} \quad \text{and} \quad X = \mathrm{cl}(O_{\boldsymbol{x}})\,.$$

Since $([r]^{\mathbb{N}}, \varrho)$ is compact (see Proposition 3.5), we have that (X, ϱ) is a compact metric space. Moreover, T is continuous and since $T(O_{\boldsymbol{x}}) \subseteq O_{\boldsymbol{x}}$ we have

$$\begin{aligned} X = \mathrm{cl}(O_{\boldsymbol{x}}) \supseteq \mathrm{cl}(T(O_{\boldsymbol{x}})) = \bigcap_{\substack{C' \text{ closed} \\ C' \supseteq T(O_{\boldsymbol{x}})}} C' &\overset{(*)}{\supseteq} \bigcap_{\substack{C \text{ closed} \\ C \supseteq O_{\boldsymbol{x}}}} T(C) \\ &\supseteq T\Bigg(\bigcap_{\substack{C \text{ closed} \\ C \supseteq O_{\boldsymbol{x}}}} C\Bigg) = T(\mathrm{cl}(O_{\boldsymbol{x}})) = T(X)\,, \end{aligned}$$

where the continuity of T was used for $(*)$.

Consequently, (X, T) is a dynamical system, which satisfies the assumptions of Theorem 3.4 and applying it with $k-1$ and $\varepsilon = 1/3$, yields some $\boldsymbol{y} \in X$ and $\lambda \in \mathbb{N}$ such that

$$\varrho\big(\boldsymbol{y}, T^{i\lambda}(\boldsymbol{y})\big) < \frac{1}{3} \tag{3.7}$$

for every $i \in [k-1]$. Moreover, we infer from the continuity of T that $T^{i\lambda}$ is continuous for every $i \in [k-1]$. Hence, there exists some $\delta > 0$ such that for every $\boldsymbol{z} \in X$ with $\varrho(\boldsymbol{y}, \boldsymbol{z}) < \delta$ we have

$$\varrho\big(T^{i\lambda}(\boldsymbol{y}), T^{i\lambda}(\boldsymbol{z})\big) < \frac{1}{3} \tag{3.8}$$

for every $i \in [k-1]$. In particular, there exists some $\boldsymbol{z} \in O_{\boldsymbol{x}}$ satisfying (3.8) and, therefore, there exists some $a \in \mathbb{N}$ such that (3.8) holds for $\boldsymbol{z} = T^a(\boldsymbol{x})$. With out loss of generality, we may assume $\varrho(\boldsymbol{y}, T^a(\boldsymbol{x})) < \delta < 1/3$. Since $T^a(T^b(\boldsymbol{z})) = T^{a+b}(\boldsymbol{z})$ by definition of the left shift, we infer from (3.7) and (3.8) applied to $\boldsymbol{z} = T^a(\boldsymbol{x})$ that

$$\varrho\big(T^a(\boldsymbol{x}), T^{a+i\lambda}(\boldsymbol{x})\big) \leq \varrho\big(T^a(\boldsymbol{x}), \boldsymbol{y}\big) + \varrho\big(\boldsymbol{y}, T^{i\lambda}(\boldsymbol{y})\big) + \varrho\big(T^{i\lambda}(\boldsymbol{y}), T^{a+i\lambda}(\boldsymbol{x})\big) < 1$$

for every $i \in [k-1]$. In particular, $(T^a \circ \boldsymbol{x})(1) = (T^{a+i\lambda} \circ \boldsymbol{x})(1)$ for every $i \in [k-1]$ and, therefore,

$$T^a(\boldsymbol{x}(1)) = T^{a+\lambda}(\boldsymbol{x}(1)) = T^{a+2\lambda}(\boldsymbol{x}(1)) = \cdots = T^{a+(k-1)\lambda}(\boldsymbol{x}(1))\,.$$

Recalling that T is the left shift on $[r]^{\mathbb{N}}$ we infer that there exists some $j \in [r]$ such that

$$j = \boldsymbol{x}(a+1) = \boldsymbol{x}((a+1)+\lambda) = \boldsymbol{x}((a+1)+2\lambda) = \cdots = \boldsymbol{x}((a+1)+(k-1)\lambda)\,,$$

which means that $((a+1)+i\lambda) \in E_j$ for every $i = 0, \ldots, k-1$, i.e., E_j contains an arithmetic progression of length k. $\square$

3.2.2. The multiple recurrence theorem. In this section we verify Theorem 3.4. In fact, we show the following more general *multiple recurrence theorem.*

THEOREM 3.6 (multiple recurrence theorem). *Let (X, ϱ) be a compact metric space and let $T_1, \ldots, T_k$ be a commuting family of continuous maps on X. Then there exists an $x \in X$ such that for every $\varepsilon > 0$ there exists an $n \in \mathbb{N}$ such that*

$$\varrho(x, T_i^n(x)) < \varepsilon$$

simultaneously for every $i \in [k]$.

The topological van der Waerden theorem (Theorem 3.4) follows from the multiple recurrence theorem by setting for $i \in [k]$

$$T_i = T^i = \underbrace{T \circ \cdots \circ T}_{i\text{-times}},$$

which are clearly commuting.

The proof of Theorem 3.6 splits into two parts. First we prove it under the more restrictive assumption that the continuous commuting maps $T_1, \ldots, T_k$ are indeed homeomorphisms (see Proposition 3.9). We remark that van der Waerden's theorem for partitions of $\mathbb{Z}$ can be deduced directly from Proposition 3.9, since in this case we would consider the left shift T on $[r]^{\mathbb{Z}}$, which is a homeomorphism. On the other hand, it is not hard reduce Theorem 3.6 to Proposition 3.9 and we give those details in Section 3.2.2.2

3.2.2.1. *Multiple recurrence for homeomorphisms.* In the is section we verify a version of Theorem 3.6, where we additionally assume that the continuous maps T_i are homeomorphisms on X. We begin with the following lemma concerning *homogeneous* subsets.

DEFINITION 3.7 (homogeneous set). Let (X, ϱ) be a compact metric space and $T\colon X \to X$ be continuous. A closed non-empty set $Z \subseteq X$ is *homogeneous with respect to T*, if there exists a group of homeomorphisms Γ of X such that each $S \in \Gamma$ commutes with T (i.e., $S \circ T = T \circ S$) and $S(Z) \subseteq Z$ for every $S \in \Gamma$ and such that (Z, Γ) is a minimal dynamical system.

PROPOSITION 3.8. *Let (X, ϱ) be a compact metric space, let $T\colon X \to X$ be a continuous map on X, and let $Z \subseteq X$ be homogeneous with respect to T. Suppose that for every $\varepsilon > 0$ there exist x, $y \in Z$ and $m \in \mathbb{N}$ such that $\varrho(x, T^m(y)) < \varepsilon$. Then the following holds*

- (*i*) *for every $\delta > 0$ and every $z \in Z$ there exist $z' \in Z$ and $n \in \mathbb{N}$ such that $\varrho(z, T^n(z')) < \delta$,*
- (*ii*) *for every $\delta > 0$ there exist $z \in Z$ and $n \in \mathbb{N}$ such that $\varrho(z, T^n(z)) < \delta$,*
- (*iii*) *for every $\delta > 0$ such that $Z_\delta = \{z \in Z\colon \exists n \in \mathbb{N} \text{ s.t. } \varrho(z, T^n(z)) < \delta\}$ is dense in Z, and*
- (*iv*) *there exists some $z \in Z$ such that for every $\delta > 0$ there exists $n \in \mathbb{N}$ such that $\varrho(z, T^n(z)) < \delta$.*

PROOF. We first verify part (*i*) and then we show

$$(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv).$$

The proof given below of assertions *(i)* and *(ii)* of Proposition 3.8 was attributed to Rufus Bowen (see [38, Proposition 1.2] or [37, Section 2.1]).

Proof of *(i)*: Let $\delta > 0$ and $z \in Z$ be given and let Γ be the group of homeomorphism guaranteed by the assumption that Z is homogeneous. We appeal to Proposition A.21, applied to (Z, Γ) with $\delta/2$ and obtain a finite subset $\Gamma' \subseteq \Gamma$ with the property that for every x, $y \in Z$ there exists some $S \in \Gamma'$ such that $\varrho(x, S(y)) < \delta/2$. In particular, for $x = z$ we have that for every $y \in Z$ there exists $S \in \Gamma'$ with

$$\varrho(z, S(y)) < \delta/2\,. \tag{3.9}$$

Moreover, by the assumption of Proposition 3.8 there exist y, $y' \in Z$ and $m \in \mathbb{N}$ such that

$$\varrho(y, T^m(y')) < \varepsilon\,,$$

where ε is sufficiently small so that for every $S \in \Gamma'$ the continuity of S implies that $\varrho(S(x), S(x')) < \delta/2$ for all x, $x' \in X$ with $\varrho(x, x') < \varepsilon$. Note that we used the notion of uniform continuity here. However, this is justified since between compact metric spaces every continuous function is uniformly continuous (see Theorem A.13).

In particular, this choice of ε guarantees for every $S \in \Gamma'$ that

$$\varrho\big(S(y), T^m(S(y'))\big) = \varrho\big(S(y), S(T^m(y'))\big) < \delta/2\,,$$

where we used that every $S \in \Gamma'$ commutes with T. It follows from (3.9) applied with $y = y'$ that there exists some $S \in \Gamma'$ such that we have

$$\varrho\big(z, T^m(S(y'))\big) \leq \varrho\big(z, S(y')\big) + \varrho\big(S(y'), T^m(S(y'))\big) < \delta\,.$$

Setting $z' = S(y')$ yields *(i)*.

(i) $\Rightarrow$ *(ii)*: Let $\delta > 0$ be given. We shall repeatedly apply *(i)*. Fix some $z_0 \in Z$. We will inductively apply *(i)*, and the j-th application will yield $z_j \in Z$ and $n_j \in \mathbb{N}$. Suppose for $j \geq 0$ the points $z_0, z_1, \dots, z_j \in Z$ and $n_1, \dots, n_j$ were chosen already.

Let $z_{j+1} \in Z$ and $n_{j+1} \in \mathbb{N}$ be given by assertion *(i)* applied for $\varepsilon_{j+1} \leq \delta/2^{j+2}$ and z_j, where $\varepsilon_{j+1} > 0$ was chosen in such a way that the continuity of T implies

$$\varrho\big(T^{n_s+\dots+n_t}(x), T^{n_s+\dots+n_t}(x')\big) < \frac{\delta}{2^{t+2}} \tag{3.10}$$

for all $1 \leq s \leq t \leq j$ and all x, $x' \in Z$ with $\varrho(x, x') < \varepsilon_{j+1}$.

This way we obtain sequences $(z_j)_{j \in \mathbb{N}_0}$ and $(n_j)_{j \in \mathbb{N}}$ such that

$$\varrho(z_j, T^{n_{j+1}}(z_{j+1})) < \varepsilon_{j+1} \leq \frac{\delta}{2^{j+2}} \tag{3.11}$$

for every $j \in \mathbb{N}_0$. A simple inductive argument shows that for every $0 \leq s < t$ we have

$$\varrho\big(z_s, T^{n_{s+1}+\dots+n_t}(z_t)\big) < \frac{\delta}{2} \sum_{j=s+1}^{t} 2^{-j} < \frac{\delta}{2}\,. \tag{3.12}$$

In fact, this estimate holds for $t = s + 1$ due to (3.11) and, hence, by induction we have

$$\begin{aligned}
&\varrho\big(z_s, T^{n_{s+1}+\dots+n_{t+1}}(z_{t+1})\big)\\
&\quad\leq \varrho\big(z_s, T^{n_{s+1}+\dots+n_t}(z_t)\big) + \varrho\big(T^{n_{s+1}+\dots+n_t}(z_t), T^{n_{s+1}+\dots+n_t}(T^{n_{t+1}}(z_{t+1}))\big)\\
&\qquad\qquad < \frac{\delta}{2} \sum_{j=s+1}^{t} 2^{-j} + \frac{\delta}{2^{t+2}} = \frac{\delta}{2} \sum_{j=s+1}^{t+1} 2^{-j}\,,
\end{aligned}$$

where we used that $\varrho(z_{t-1}, T^{n_t}(z_t)) < \varepsilon_t$ and (3.10) for the last estimate.

The compactness of the metric space Z (inherited from the compactness of X) implies that there exist integers s and t with $0 \leq s < t$ such that $\varrho(z_s, z_t) < \delta/2$ and, hence, (3.12) yields

$$\varrho\big(z_t, T^{m_{s+1}+\cdots+m_t}(z_t)\big) \leq \varrho(z_t, z_s) + \varrho\big(z_s, T^{m_{s+1}+\cdots+m_t}(z_t)\big) < \frac{\delta}{2} + \frac{\delta}{2} = \delta\,,$$

which concludes the proof of assertion *(ii)* of Proposition 3.8.

(ii) $\Rightarrow$ *(iii)*: Let $\delta > 0$ be given and let $U \supseteq \mathcal{B}_\varepsilon(x)$ be some open set in Z. Again we appeal to Proposition A.21, applied to (Z, Γ) with ε and obtain the finite set $\Gamma' \subseteq \Gamma$. Moreover, we apply the conclusion of part *(ii)* of Proposition 3.8 for $\delta' > 0$ sufficiently small, so that for every $S \in \Gamma'$ we have $\varrho(S(x'), S(x'')) < \delta$ for all x', $x'' \in Z$ with $\varrho(x', x'') < \delta'$. Part *(ii)* yields some $n \in \mathbb{N}$ and $z \in Z$ such that $\varrho(z, T^n(z)) < \delta'$ and the choice of δ' yields $\varrho(S(z), S(T^n(z))) < \delta$ for every $S \in \Gamma'$. Moreover, owing to the properties of Γ' there exists some $S \in \Gamma'$ such that $\varrho(x, S(z)) < \varepsilon$.

Summarising, we have shown that $S(z) \in U$ and

$$\varrho\big(S(z), T^n(S(z))\big) = \varrho\big(S(z), S(T^n(z))\big) < \delta\,,$$

which means $S(z) \in Z_\delta$. Since U was arbitrary, this means that every open set in Z contains some element from Z_δ, i.e., Z_δ is dense in Z.

(iii) $\Rightarrow$ *(iv)*: Finally, we verify conclusion *(iv)* of Proposition 3.8. For every $N \in \mathbb{N}$ we consider the sets $Z_{1/N}$. It follows from the continuity of T^n for every $n \in \mathbb{N}$ that $Z_{1/N}$ is open for every $N \in \mathbb{N}$. Moreover, $Z_{1/N}$ is dense due to *(iii)*. Hence, it follows for example from Baire's catagory theorem that $\bigcap_{N\in\mathbb{N}} Z_{1/N} \neq \emptyset$ (in fact, it is dense itself) and every $z \in \bigcap_{N\in\mathbb{N}} Z_{1/N}$ satisfies conclusion *(iv)* of Proposition 3.8.

Alternatively, we can show $\bigcap_{N\in\mathbb{N}} Z_{1/N} \neq \emptyset$ directly. Clearly, $Z_1 \supseteq Z_{1/2} \supseteq Z_{1/3} \supseteq \dots$ form a chain (under inclusion) of non-empty open sets and, hence, we can find a chain of non-empty closed subsets $(Y_{1/N})_{N\in\mathbb{N}}$ with $Y_{1/N} \subseteq Z_{1/N}$. The chain $(Y_{1/N})_{N\in\mathbb{N}}$ has the finite intersection property and, hence, the compactness of Z yields $\bigcap_{N\in\mathbb{N}} Z_{1/N} \supseteq \bigcap_{N\in\mathbb{N}} Y_{1/N} \neq \emptyset$ (see Proposition A.8). $\square$

Now we are ready to prove the Theorem 3.6 for homeomorphisms.

PROPOSITION 3.9 (multiple recurrence theorem for homeomorphisms). *Let (X, ϱ) be a compact metric space and let $T_1, \dots, T_k$ be a commuting family of homeomorphisms on X. Then there exists an $x \in X$ such that for every $\varepsilon > 0$ there exists an $n \in \mathbb{N}$ such that*

$$\varrho(x, T_i^n(x)) < \varepsilon$$

simultaneously for every $i \in [k]$.

PROOF. We proceed by induction on k. For $k = 1$ Proposition 3.9 follows from Birkhoff's recurrence theorem (Theorem A.17).

Let $k > 1$ and let Γ be the group generated by $T_1, \dots, T_k$. Owing to Proposition A.22 we may assume without loss of generality that (X, Γ) is minimal dynamical system.

In the proof we consider the k-fold product $\boldsymbol{X} = X \times \dots \times X$ and let $\Delta(\boldsymbol{X})$ be the diagonal elements in $\boldsymbol{X}$, i.e., $\Delta(\boldsymbol{X}) = \{(x, \dots, x) \colon x \in X\} \subseteq \boldsymbol{X}$. We can fix the following metric $\boldsymbol{\varrho}$ on $\boldsymbol{X}$ by setting

$$\boldsymbol{\varrho}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{k} \varrho(x_i, y_i)$$

for every $\boldsymbol{x} = (x_1, \dots, x_k)$, $\boldsymbol{y} = (y_1, \dots, y_k) \in \boldsymbol{X}$. We fix $\boldsymbol{T} = T_1 \times \dots \times T_k \colon \boldsymbol{X} \to \boldsymbol{X}$ in the obvious way

$$\boldsymbol{T}(\boldsymbol{x}) = (T_1(x_1), \dots, T_k(x_k))$$

for every $\boldsymbol{x} = (x_1, \ldots, x_k) \in \boldsymbol{X}$. It follows from those definitions that $(\boldsymbol{X}, \boldsymbol{\varrho})$ is a compact metric space and T is continuous on $\boldsymbol{X}$ (actually it is even a homeomorphism). Moreover, for every $S \in \Gamma$ let $\boldsymbol{S} = S \times \cdots \times S$ and $\boldsymbol{\Gamma} = \{\boldsymbol{S} \colon S \in \Gamma\}$. The minimality of (X, Γ) implies that $(\Delta(\boldsymbol{X}), \boldsymbol{\Gamma})$ is a minimal dynamical system. Moreover, every $\boldsymbol{S} \in \boldsymbol{\Gamma}$ is a homeomorphism on $\boldsymbol{X}$ and commutes with $\boldsymbol{T}$. Hence, $\Delta(\boldsymbol{X})$ is homogeneous with respect to $\boldsymbol{T}$. Furthermore, we claim that $\Delta(\boldsymbol{X})$ satisfies the assumption of Proposition 3.8. In fact, we apply the induction assumption of Proposition 3.9 to the maps $R_1, \ldots, R_{k-1}$ defined by $R_i = T_i \circ T_k^{-1}$. The induction assumption yields some $x \in X$ such that for every $\varepsilon > 0$ there exists some $m \in \mathbb{N}$ such that $\varrho(x, R_i^m(x)) < \varepsilon$ for every $i \in [k-1]$ and, hence, for $\boldsymbol{x} = (x, \ldots, x) \in \Delta(\boldsymbol{X})$ and $\boldsymbol{y} = (T_k^{-m}(x), \ldots, T_k^{-m}(x)) \in \Delta(\boldsymbol{X})$, we have

$$\boldsymbol{\varrho}(\boldsymbol{x}, \boldsymbol{T}^m(\boldsymbol{y})) = \sum_{i=1}^{k} \varrho(x, T_i^m(T_k^{-m}(x)) = \sum_{i=1}^{k-1} \varrho(x, R_i^m(x)) + \varrho(x, x) < (k-1)\varepsilon\,.$$

Consequently, Proposition 3.8 part *(iv)* yields the existence of some $\boldsymbol{z} = (z, \ldots, z) \in \Delta(\boldsymbol{X})$ such that for every $\delta > 0$ there exists some $n \in \mathbb{N}$ such that $\boldsymbol{\varrho}(\boldsymbol{z}, \boldsymbol{T}^n(\boldsymbol{z})) < \delta$ and, in particular, $\varrho(z, T_i^n(z)) < \delta$ for every $i \in [k]$. $\square$

3.2.2.2. *Multiple recurrence for continuous maps.* Below we verify Theorem 3.6 and reduce it to the "homeomorphism version" (Proposition 3.9).

Proof of Theorem 3.6. Let (X, ϱ) be a compact metric space and $T_1, \ldots, T_k$ be a commuting family of continuous maps on X. We consider the product space $\boldsymbol{X}$ of all functions from $\mathbb{Z}^k$ to X, i.e., $\boldsymbol{X} = X^{\mathbb{Z}^k}$. Owing to Tychonoff's theorem (Theorem A.10) the topological space $\boldsymbol{X}$ is compact. Moreover, for fix an enumeration τ of $\mathbb{Z}^k$ (i.e., $\tau\colon \mathbb{Z}^k \to \mathbb{N}$ is a bijection) and set $\boldsymbol{\varrho}\colon \boldsymbol{X} \times \boldsymbol{X} \to \mathbb{R}_{\geq 0}$

$$\boldsymbol{\varrho}_\tau(f, g) = \sum_{\vec{z} \in \mathbb{Z}^k} \frac{\varrho(f(\vec{z}), g(\vec{z}))}{2^{\tau(\vec{z})}} \tag{3.13}$$

for all functions $f,\, g\colon \mathbb{Z}^k \to X$. It is easy to check that $\boldsymbol{\varrho}_\tau$ defines a metric on $\boldsymbol{X}$ and that $\boldsymbol{\varrho}_\tau$ induces the product topology on $\boldsymbol{X}$ (the proof is similar to the proof that following (3.5) in Section 3.2.1).

For $i \in [k]$ we consider left shift in the i-th coordinate $S_i\colon \boldsymbol{X} \to \boldsymbol{X}$ defined for every $f \in \boldsymbol{X}$ by

$$(S_i \circ f)(\vec{z}) = f(z_1, \ldots, z_{i-1}, z_i + 1, z_{i+1}, \ldots, z_k)\,,$$

for every $\vec{z} = (z_1, \ldots, z_k) \in \mathbb{Z}^k$. It is straightforward to check that S_i is a homeomorphism on $\boldsymbol{X}$ for every $i \in [k]$ and, clearly, $S_1, \ldots, S_k$ commute. Summarising, $(\boldsymbol{X}, \boldsymbol{\varrho}_\tau)$ is a compact metric space and $S_1, \ldots, S_k$ are commuting homeomorphisms on $\boldsymbol{X}$.

We will apply Proposition 3.9 to the following closed subset $\boldsymbol{Y} \subseteq \boldsymbol{X}$. Let $\boldsymbol{Y}$ contain those functions $f \in \boldsymbol{X}$ satisfying

$$(S_i \circ f)(\vec{z}) = T_i(f(\vec{z})) \quad \text{for every } i \in [k] \text{ and } \vec{z} \in \mathbb{Z}^k\,.$$

For later reference we note that the definition of $\boldsymbol{Y}$ yields for every $f \in \boldsymbol{Y}$, $n \in \mathbb{N}$, $\vec{z} \in \mathbb{Z}^k$, and $i \in [k]$ that

$$(S_i^n \circ f)(\vec{z}) = T_i^n(f(\vec{z}))\,. \tag{3.14}$$

It is easy to see that $\boldsymbol{X} \setminus \boldsymbol{Y}$ is open. In fact, if $f \in \boldsymbol{X} \setminus \boldsymbol{Y}$, then there exists $\vec{z} = (z_1, \ldots, z_k) \in \mathbb{Z}^k$ and $i \in [k]$ such that for $\vec{z}_i = (z_1, \ldots, z_i + 1, \ldots, z_k)$

$$T(f(\vec{z})) \neq f(\vec{z}_i)\,.$$

Consequently, every function $g \in \boldsymbol{X}$ which takes the same values on $\vec{z}$ and $\vec{z}_i$ as f is not in $\boldsymbol{Y}$. The set of all those functions g defines a basic open set in the product topology of $\boldsymbol{X}$, which is contained in $\boldsymbol{X} \setminus \boldsymbol{Y}$ and, hence, $\boldsymbol{X} \setminus \boldsymbol{Y}$ is open.

Moreover, it follows from the definition of $\boldsymbol{Y}$ that for every $f \in \boldsymbol{Y}$ we have $S_i(f) \in \boldsymbol{Y}$ and $S_i^{-1}(f) \in \boldsymbol{Y}$ for every $i \in [k]$. In other words, $\boldsymbol{Y}$ is invariant under S_i and S_i^{-1} for every $i \in [k]$.

We are going to apply Proposition 3.9 to $S_1, \dots, S_k$ and the $(\boldsymbol{Y}, \boldsymbol{\varrho}_\tau)$. However, in order to ensure that $(\boldsymbol{Y}, \boldsymbol{\varrho}_\tau)$ is indeed a (compact) metric space, we still have to verify that $\boldsymbol{Y}$ is non-empty. For every $n \in \mathbb{N}$ we define a function $f_n \colon \mathbb{Z}^k \to X$ as follows: Fix some $x \in X$ and for $\vec{z} = (z_1, \dots, z_k) \in \mathbb{Z}^k$ with $z_i < -n$ for some $i \in [k]$ we set $f_n(\vec{z}) = x$ and for all other $\vec{z} = (z_1, \dots, z_k) \in \mathbb{Z}^k$ we set

$$f_n(\vec{z}) = T_1^{z_1+n}(T_2^{z_2+n}(\dots(T_k^{z_k+n}(x))\dots)).$$

Since $\boldsymbol{X}$ is compact there exists a converging subsequence $(f_{n_j})_{j \in \mathbb{N}}$ and we set $f = \lim_{j \to \infty} f_{n_j}$. Moreover, for every $n \in \mathbb{N}$, $\vec{z} \in \mathbb{Z}^k$ with $z_1, \dots, z_k \geq -n$, and $i \in [k]$ it follows from the assumption that $T_1, \dots, T_k$ are commuting that

$$\begin{aligned} (S_i \circ f_n)(\vec{z}) &= f_n(z_1, \dots, z_i + 1, \dots, z_k) \\ &= T_1^{z_1+n}(\dots(T^{z_i+1+n}(\dots(T_k^{z_k+n}(x))\dots))\dots) \\ &= T_i(T_1^{z_1+n}(\dots(T^{z_i+n}(\dots(T_k^{z_k+n}(x))\dots))\dots) = T_i(f_n(\vec{z})). \end{aligned}$$

Consequently, $(S_i \circ f)(\vec{z}) = T_i(f(\vec{z}))$ for every $\vec{z} \in \mathbb{Z}^k$ and every $i \in [k]$ and, therefore, $f \in \boldsymbol{Y}$

Summarising the above, we have shown that $(\boldsymbol{Y}, \boldsymbol{\varrho}_\tau)$ is a compact metric space, which is invariant under $S_1, \dots, S_k$ and $S_1^{-1}, \dots, S_k^{-1}$. Hence, it follows from the fact that $S_1, \dots, S_k$ are commuting homeomorphism on $\boldsymbol{X}$, that $S_1, \dots, S_k$ are also commuting homeomorphism on $\boldsymbol{Y}$. In other words, $(\boldsymbol{Y}, \boldsymbol{\varrho}_\tau)$ and $S_1, \dots, S_k$ satisfy the assumptions of Proposition 3.9, which yields a function $h \in \boldsymbol{Y}$ such that for every $\varepsilon > 0$ there exists $n \in \mathbb{N}$ with

$$\boldsymbol{\varrho}_\tau(h, S_i^n(h)) < \varepsilon \tag{3.15}$$

for every $i \in [k]$. Recalling, that τ is some fixed enumeration of $\mathbb{Z}^k$. We set $\vec{z}^* = \tau^{-1}(1) \in \mathbb{Z}^k$. It follows from (3.15) and the definition of $\boldsymbol{\varrho}_\tau$ in (3.13), that

$$\varrho\big(h(\vec{z}^*), (S_i^n \circ h)(\vec{z}^*)\big) < 2\varepsilon.$$

Hence, for $x^* = h(\vec{z}^*) \in X$ we obtain

$$\varrho\big(x^*, T_i^n(x^*)\big) = \varrho\big(h(\vec{z}^*), T_i^n(h(\vec{z}))\big) \overset{(3.14)}{=} \varrho\big(h(\vec{z}^*), (S_i^n \circ h)(\vec{z}^*)\big) < 2\varepsilon$$

for every $i \in [k]$, which concludes the proof of Theorem 3.6. □

3.3. Two proofs of Roth's theorem

In this section we focus on Roth's theorem, the first non-trivial case of Szemerédi's theorem, i.e., Theorem 1.6 for $k = 3$. This result was first obtained by Roth [77] and in Section 3.3.2 we give the improved proof of Roth from [78]. Before in Section 3.3.1 we present a combinatorial proof of Roth's theorem due to Szemerédi (see [84, Remark on page 94]).

We begin with a few observations. Throughout this section we mean by an AP_k for an integer $k \in \mathbb{N}$ an (non-trivial) arithmetic progression of length k. Note that for $k = 3$ an AP_3 consists of three integers $x < y < z$ satisfying $x + z = 2y$. We say a set $A \subseteq \mathbb{N}$ is *AP_k-free*, if it contains no AP_k. For $k \in \mathbb{N}$ and a subset $X \subseteq \mathbb{N}$ we set

$$r_k(X) = \max\{|A| \colon A \subseteq X \text{ and } A \text{ is } AP_k\text{-free}\}$$

and for $X = [n] \subseteq \mathbb{N}$ we omit the square brackets and simply write $r_k(n)$. Szemerédi's theorem is equivalent to the assertion that for every $k \in \mathbb{N}$ the ratio $r_k(n)/n$ tends to zero as $n \to \infty$. It follows from the definition that $r_k(X) = r_k(n)$ whenever X is an arithmetic progression of length n and, hence, in particular, when X is an interval of length n. Consequently, the pigeonhole principle yields

$$r_k(n+m) = r_k([n+m]) \le r_k([n]) + r_k([n+1, n+m]) = r_k(n) + r_k(m)$$

for all integers n, $m \in \mathbb{N}$. In other words, $r_k(n)$ is a subadditive function and owing to Fekete's lemma (Proposition B.2) the limit $\tau_k = \lim_{n\to\infty} r_k(n)/n$ exists and for every $n \in \mathbb{N}$ we have

$$\frac{r_k(n)}{n} \ge \tau_k \,. \tag{3.16}$$

In Sections 3.3.1 and 3.3.2 we will show Roth's theorem from [77] and establish

$$\tau_3 = 0 \,. \tag{3.17}$$

in two very different ways. First we give a combinatorial argument due to Szemerédi (see [84, page 94]) in Section 3.3.1 and then we give the improved version of Roth's original argument from [78] in Section 3.3.2.

3.3.1. Szemerédi's combinatorial proof of Roth's theorem. The key ingredient in Szemerédi's argument is the following *density version* of Hilbert's cube lemma (Theorem 1.2). We recall that for integers a, $\lambda_1, \dots, \lambda_k$ the *k-cube* $\mathcal{C}(a; \lambda_1, \dots \lambda_k)$ *spanned by* $a, \lambda_1, \dots, \lambda_k$ is defined by

$$\mathcal{C}(a; \lambda_1, \dots \lambda_k) = \left\{ a + \sum_{i=1}^{k} \delta_i \lambda_i \colon\ (\delta_1, \dots, \delta_k) \in \{0,1\}^k \right\}.$$

The following result asserts that sets $X \subseteq [n]$ of non-vanishing density contain cubes of dimension $\log\log(n)$.

Proposition 3.10 (Cube lemma (density version)). *Let $\varepsilon > 0$, $n \in \mathbb{N}$, and $X \subseteq [n]$ satisfy $|X| \ge \varepsilon n$. Suppose for some integer $k \in \mathbb{N}$ we have*

$$n \ge \frac{4^{2^k - 1}}{\varepsilon^{2^k}}$$

holds, then there exist $a, \lambda_1, \dots, \lambda_k \in \mathbb{N}$ such that $\mathcal{C}(a; \lambda_1, \dots \lambda_k) \subseteq X$.

In particular, every set $X \subseteq J$ with $|X| \ge \varepsilon |J|$ contains a k-cube for

$$k \ge \lfloor \log_2 \log_2(|J|) - \log_2 \log_2(4/\varepsilon) \rfloor$$

for every interval $J \subseteq \mathbb{N}$.

We refer to Proposition 3.10 as a density version of Theorem 1.2 since it guarantees a k-cube for subsets of $[n]$ of positive density, while Theorem 1.2 (via the compactness principle) is an assertion for partitions of $[n]$. Similarly, we may view Szemerédi's theorem (Theorem 1.6) as a density version of van der Waerden's theorem (Theorem 1.4). Also note that k in Proposition 3.10 is not necessarily fixed and can slowly grow with n, which will become important later. We remark that Proposition 3.10 is best possible in the sense that a a simple probabilistic argument (or a counting argument as in the proof of Theorem 2.4) shows that for sufficiently large n there are subsets $X \subseteq [n]$ with $|X| \ge n/2$, which contain no k-cube with $k \ge (1+\xi)\log_2\log_2(n)$.

Proof. Let ε, n, X, and k satisfy the assumptions of Proposition 3.10 Set $X_0 = X$. Inductively we define for $i = 1, \dots k$ integers $\lambda_i \in \mathbb{N}$ and sets $X_i \subseteq X_{i-1}$ such that

$$\{\lambda_i + x \colon\ x \in X_i\} \subseteq X_{i-1} \quad \text{and} \quad |X_i| \ge \frac{|X_{i-1}|^2}{4n} \ge \frac{\varepsilon^{2^i} n}{4^{2^i - 1}} \,.$$

Note that the lower bound on the size of X_i yields $|X_k| \geq 1$ and $|X_{i-1}| \geq 2$ for every $i = 0, \dots, k-1$. Suppose $\lambda_1, \dots, \lambda_{i-1} \in \mathbb{N}$ and $X_0 \supseteq X_1 \supseteq \dots \supseteq X_{i-1}$ were defined already. Every pair $\{x, y\} \in \binom{X_{i-1}}{2}$ with $x < y$ has a difference $y - x \in [n-1]$ and, hence, one difference appears for at least

$$\frac{\binom{|X_{i-1}|}{2}}{n-1} \geq \frac{|X_{i-1}|^2}{4n} \tag{3.18}$$

pairs, where we used $|X_{i-1}| \geq 2$. Let $\lambda_i \in \mathbb{N}$ such a difference and set

$$X_i = \{x \in X_{i-1} \colon x + \lambda_i \in X_{i-1}\}.$$

Clearly, the choice of X_i and λ_i yields $X_i \subseteq X_{i-1}$ and $\{\lambda_i + x \colon x \in X_i\} \subseteq X_{i-1}$. Moreover, (3.18) and the inductive bound on $|X_{i-1}|$ gives

$$|X_i| \geq \frac{|X_{i-1}|^2}{4n} \geq \frac{(\varepsilon^{2^{i-1}} n / 4^{2^{i-1}-1})^2}{4n} = \frac{\varepsilon^{2^i} n}{4^{2^i - 1}}.$$

Finally, we fix some $a \in X_k$. The definition of $\lambda_1, \dots, \lambda_k$ and of the sets $X_1, \dots, X_k$ yields by induction that the $(k-i+1)$-cube $\mathcal{C}(a; \lambda_k, \dots, \lambda_i)$ is contained in X_{i-1}. Consequently, the k-cube $\mathcal{C}(a; \lambda_k, \dots, \lambda_1) = \mathcal{C}(a; \lambda_1, \dots, \lambda_k) \subseteq X_0 = X$. □

After these preparations we verify (3.17).

Szemerédi's proof of (3.17). We assume for a contradiction that $\tau_3 = \tau$ for some $\tau > 0$. We set

$$\xi = \min\left\{\frac{\tau}{200}, \frac{\tau^2}{10}\right\}. \tag{3.19}$$

In view of (3.16) we may choose $m_0 \in \mathbb{N}$ to be sufficiently large such that

$$\tau \leq \frac{r_3(m)}{m} \leq \tau + \xi \tag{3.20}$$

for every $m \geq m_0$. Finally, we let n be sufficiently large so that

$$5\Big\lfloor \log_2 \log_2\left(\sqrt{n/100}\right) - \log_2 \log_2(8/\tau)\Big\rfloor \geq \log_2 \log_2(n) \tag{3.21}$$

and

$$\frac{\tau^2}{100} \log_2 \log_2(n) \geq m_0 \tag{3.22}$$

and let $X \subseteq [n]$ be an AP_3-free set with $|X| \geq \tau n$. Without loss of generality, we may assume that n is a square number and that it is divisible by 100. Owing to (3.20) we have

$$|X \cap Y| \leq (\tau + \xi)|Y|$$

for every set Y, which forms an arithmetic progression of length at least m_0, since otherwise $|X \cap Y|$ would contain an AP_3. In particular,

$$|X \cap [0.49n]| + |X \cap [0.5n + 1, n]| \leq (\tau + \xi) \cdot 0.99n$$

and, consequently, for the interval $I = [0.49n + 1, 0.5n]$ we infer from $\xi \leq \tau/200$ that

$$|X \cap I| \geq \tau n - (\tau + \xi) \cdot 0.99n \geq \frac{\tau}{2}|I|. \tag{3.23}$$

Splitting I into $\sqrt{n/100}$ intervals of length $\sqrt{n/100}$ we infer that there exists such an interval $J \subset I$ with $|X \cap J| \geq \tau |J|/2$. Hence, Proposition 3.10 yields integers a and $\lambda_1, \dots, \lambda_k \in \mathbb{N}$ such that the k-cube $\mathcal{C}(a; \lambda_1, \dots, \lambda_k)$ is contained in $X \cap J$ with

$$k = \Big\lfloor \log_2 \log_2\left(\sqrt{n/100}\right) - \log_2 \log_2(8/\tau)\Big\rfloor \tag{3.24}$$

and we have

$$\sum_{i=1}^{k} \lambda_i \leq |J| = \frac{\sqrt{n}}{10}. \tag{3.25}$$

We set $\mathcal{C}_0 = \{a\}$ and for $i \in k$ set $\mathcal{C}_i = \mathcal{C}(a; \lambda_1, \dots, \lambda_i)$, i.e., $\mathcal{C}_i$ is the i-cube spanned by a and $\lambda_1, \dots, \lambda_i$. Moreover, for $i = 0, \dots, k$ we set

$$Z_i = \{2y - x\colon\ x \in X \cap [0.49n] \text{ and } y \in \mathcal{C}_i\}\,.$$

Note that for every $z \in Z_i$ there exist $x \in X$ and $y \in \mathcal{C}_i \subseteq X$ with $x < y < z$ such that $x + z = 2y$, i.e., x, y, z forms an AP_3 in $[n]$. In particular, $Z_i \cap X = \emptyset$, since X is AP_3-free.

Similar as in (3.23) one can show that $|X \cap [0.49n]| \geq \frac{\tau}{2} \cdot 0.49n$ and, therefore,

$$|Z_0| \geq |X \cap [0.49n]| \geq 0.245\tau n\,. \tag{3.26}$$

Moreover, we have $\mathcal{C}_i = \mathcal{C}_{i-1} \cup \{y + \lambda_i\colon\ y \in \mathcal{C}_{i-1}\}$,

$$Z_i = Z_{i-1} \cup \{z + 2\lambda_i\colon\ z \in Z_{i-1}\}\,, \tag{3.27}$$

and

$$Z_0 \subseteq Z_1 \subseteq \cdots \subseteq Z_k \subseteq [n]\,. \tag{3.28}$$

In particular, there exists some $i \in [k]$ such that

$$|Z_i \setminus Z_{i-1}| \leq \frac{n}{k}\,. \tag{3.29}$$

Recalling that $k = \Omega(\log_2 \log_2(n))$, we have $|Z_i \setminus Z_{i-1}| \leq n/k = o(n)$, which is crucial for this proof of Roth's theorem. Below we shall show that Z_{i-1} can be decomposed into "few" (in fact, into $n/k = o(n)$) arithmetic progressions with difference $2\lambda_i$. As a consequence we will infer that $[n] \setminus Z_{i-1}$ can be decomposed into $2\lambda_i + o(n) = o(n)$ arithmetic progressions with difference $2\lambda_i$ Since $X \cap Z_{i-1} = \emptyset$ and $|Z_{i-1}| \geq |Z_0| = \Omega(n)$ it follows that the density of X on $[n] \setminus Z_{i-1}$ is "substantially" larger compared to the density of X on $[n]$. Consequently, the density of X on one of the arithmetic progressions A with difference $2\lambda_i$ in $[n] \setminus Z_i$ must be "large". In fact, $|X \cap A| > (\tau + \xi)|A|$ for some sufficiently large arithmetic progression A, so that (3.20) yields that $X \cap A$ contains an AP_3. However, this contradicts the assumption that X is AP_3-free. Below we give the details of this outline.

Let $\mathcal{A}$ be the family of maximal arithmetic progressions with difference $2\lambda_i$ completely contained in Z_{i-1}. Clearly, the arithmetic progressions in $\mathcal{A}$ are pairwise disjoint and $\{b, b + 2\lambda_i, \dots, b + (\ell - 1) \cdot 2\lambda_i\} \subseteq Z_{i-1}$ is a set in $\mathcal{A}$ (for some $b \in Z_{i-1}$ and $\ell \in \mathbb{N}$) if and only if $b + \ell \cdot 2\lambda_i \in Z_i \setminus Z_{i-1}$. Conversely, we infer from (3.27) that for every element of $z \in Z_i \setminus Z_{i-1}$ we have $z - 2\lambda_i \in Z_{i-1}$. Hence, there is a one-to-one correspondence between the arithmetic progression in $\mathcal{A}$ and the elements in $Z_i \setminus Z_{i-1}$ and owing to (3.29) we have

$$|\mathcal{A}| \leq \frac{n}{k}\,.$$

Next we consider the family $\mathcal{B}$ of maximal arithmetic progressions with difference $2\lambda_i$ completely contained in $[n] \setminus Z_{i-1}$. Firstly, note that $[n]$ can be decomposed into $2\lambda_i$ arithmetic progressions with difference $2\lambda_i$ (the equivalence classes of $[n]$ modulo $2\lambda_i$). Every member $A \in \mathcal{A}$ is contained in one of those equivalence classes and removing A from $[n]$ splits this equivalence class into two arithmetic progressions with difference $2\lambda_i$. Repeating this argument for every $A \in \mathcal{A}$ shows that the set $[n] \setminus Z_{i-1}$ can be decomposed into at most $2\lambda_i + |\mathcal{A}|$ arithmetic progressions with difference $2\lambda_i$, i.e.,

$$|\mathcal{B}| \leq 2\lambda_i + \frac{n}{k}\,. \tag{3.30}$$

Finally, we say an arithmetic progression $B \in \mathcal{B}$ is *short* if its length is is less than m_0 and otherwise it is called *long*. Note that for every long B we must have

$|X \cap B| \leq (\tau + \xi)|B|$, due to (3.20). Moreover, since trivially $|X \cap B| \leq m_0$ for every short $B \in \mathcal{B}$ and since $X \cap Z_{i-1} = \emptyset$ we obtain

$$|X| = |X \cap ([n] \setminus Z_{i-1})| \leq \sum_{\substack{B \in \mathcal{B} \\ B \text{ is long}}} (\tau+\xi)|B| + \sum_{\substack{B \in \mathcal{B} \\ B \text{ is short}}} m_0 \leq (\tau+\xi)(n - |Z_{i-1}|) + m_0|\mathcal{B}|\,.$$

Owing to (3.22), (3.25), (3.26), (3.28), and (3.30) we have

$$|X| \leq \tau n + \xi n - 0.245\tau^2 n + \frac{\tau^2}{100}\log_2\log_2(n)\left(\frac{\sqrt{n}}{5} + \frac{n}{k}\right)$$

Straightforward calculations involving the choice of $\xi \leq \tau^2/10$ from (3.19) and (3.24) combined with (3.21) show that

$$\xi n + \frac{\tau^2}{100}\log_2\log_2(n)\left(\frac{\sqrt{n}}{5} + \frac{n}{k}\right) < 0.245\tau^2 n\,,$$

which yields the contradiction $|X| < \tau n$. □

3.3.2. Roth's analytical proof. In this section we follow the argument of Roth and establish (3.17) by means of Fourier analysis (see Appendix B.2 for relevant basic facts and notation). We will show the following quantitative version of (3.17).

THEOREM 3.11. *For every $\tau > 0$ the following holds. If $A \subseteq [n]$ with $|A| \geq \tau n$ for some $n \geq \exp\exp(1200/\tau)$, then A contains an AP_3.*

For the proof of Theorem 3.11 it will be convenient to move away from subsets of $[n]$ and to consider subsets of of cyclic group $\mathbb{Z}/n\mathbb{Z}$ instead. The main step in the proof is based on a density-increment argument (similar as in the proof given in Section 3.3.1), which is provided by Proposition 3.13. Roughly speaking, this lemma asserts that every subset A of $\mathbb{Z}/n\mathbb{Z}$ satisfies one the following two conditions: either A is distributed "uniformly" in the sense that all nontrivial Fourier coefficients of the indicator function of A are small compared to the trivial Fourier coefficient $\hat{\mathbf{1}}_A(0) = |A|$ (see alternative *(i)* in Proposition 3.13) or A is denser on some arithmetic progression within $\mathbb{Z}/n\mathbb{Z}$ of size $\Omega(\sqrt{n})$ (see alternative *(ii)*). We will show that the uniformity given in alternative *(i)* yields "many" AP_3's in A (see Proposition 3.14 and as a consequence Theorem 3.11 follows from iterated applications of Proposition 3.13. However, in order to be prepare for iterated applications of Proposition 3.13 we have to make sure that the arithmetic progressions in $\mathbb{Z}/n\mathbb{Z}$ which we consider are arithmetic progressions in $\mathbb{Z}$ as well.

DEFINITION 3.12 ($\mathbb{Z}$-progression). We say an arithmetic progression P in $\mathbb{Z}/n\mathbb{Z}$ is a *$\mathbb{Z}$-progression*, if P viewed as a subset of $[0, n-1] \subset \mathbb{Z}$ also forms an arithmetic progression.

For example, $\{5, 1, 4\} \subset \mathbb{Z}/7\mathbb{Z}$ is an AP_3 in $\mathbb{Z}/7\mathbb{Z}$, but it is not a $\mathbb{Z}$-progression. On the other hand, $\{2, 4, 6\} \subset \mathbb{Z}/7\mathbb{Z}$ is also a $\mathbb{Z}$-progression.

PROPOSITION 3.13. *If $n \geq 50$ and $A \subseteq \mathbb{Z}/n\mathbb{Z}$ with $|A| \geq \tau n$, then one of the following holds*

(i) either $|\hat{\mathbf{1}}_A(r)| \leq \tau^2 n/100$ for every $r \in \mathbb{Z}/n\mathbb{Z} \setminus \{0\}$,
(ii) or there is a $\mathbb{Z}$-progression $P \subseteq \mathbb{Z}/n\mathbb{Z}$ of length at least $|P| \geq \tau^2\sqrt{n}/5000$ such that $|A \cap P| \geq (\tau + \tau^2/800)|P|$.

Before we prove Proposition 3.13 we establish the connection between the condition in *(i)* and the existence of AP_3's in $|A|$.

PROPOSITION 3.14. *Suppose $n > 50/\tau^3$ for some $\tau \in (0,1]$ and $A \subseteq \mathbb{Z}/n\mathbb{Z}$ satisfies $|A| \geq \tau n$. If $|\hat{\mathbf{1}}_A(r)| \leq \tau^2 n/100$ for every $r \in \mathbb{Z}/n\mathbb{Z} \setminus \{0\}$, then one of the following holds*

(*i*) *either A contains an AP_3, which is a $\mathbb{Z}$-progression,*
(*ii*) *or there exists a $\mathbb{Z}$-progression $P \subseteq \mathbb{Z}/n\mathbb{Z}$ of length at least $|P| \geq \lfloor n/3 \rfloor$ such that $|A \cap P| \geq (\tau + \tau/6)|P|$.*

Theorem 3.11 follows from iterated applications of Propositions 3.13 and 3.14 and we first give the details of this reduction.

PROOF OF THEOREM 3.11. Let τ and n satisfy

$$n \geq \exp\exp(1200/\tau)\,. \tag{3.31}$$

Let $A \subseteq [n]$ with $|A| \geq \tau n$ and suppose for a contradiction that A contains no AP_3. We set $A_0 = \{a - 1 \colon a \in A\}$ and view A_0 as a subset of $\mathbb{Z}/n\mathbb{Z}$. Next we appeal to Proposition 3.13. If alternative (*ii*) occurs, then we obtain a $\mathbb{Z}$-progression $P_1 \subseteq \mathbb{Z}/n\mathbb{Z}$ on which the relative density of A_0 is increased. We then pass to the set $A_1 = A_0 \cap P$ which we may view as a subset of $\mathbb{Z}/n_1\mathbb{Z}$, where $n_1 = |P|$. More formally, we set $A_1 = \{a - \min_{z \in P} z \colon a \in A_0 \cap P\}$ and view A_1 as a subset of $\mathbb{Z}/n_1\mathbb{Z}$.

Note that every $\mathbb{Z}$-progression $Q \subseteq A_1 \subseteq \mathbb{Z}/n_1\mathbb{Z}$ corresponds to a sub-progression of P. Moreover, since P is a $\mathbb{Z}$-progression in $\mathbb{Z}/n\mathbb{Z}$ the progression Q also corresponds to a $\mathbb{Z}$-progression in $\mathbb{Z}/n\mathbb{Z}$. In other words, $\mathbb{Z}$-progressions Q in A_1 correspond to progressions in $A \subseteq [n]$. This would not be necessarily true if either Q or P would not be a $\mathbb{Z}$-progression and, in fact, this is the reason why we restrict our attention in Propositions 3.13 and 3.14 to $\mathbb{Z}$-progressions only.

If, on the other hand, after the first application of Proposition 3.13 alternative (*i*) occurs, then we see that A_0 satisfies the assumption of of Proposition 3.14. Owing to the assumption that A is AP_3-free we see that alternative (*i*) is impossible . Consequently, alternative (*ii*) of Proposition 3.14 occurs, which guarantees an even bigger density increment on an longer $\mathbb{Z}$-progression (compared with alternative (*ii*) of Proposition 3.13).

We iterate the same argument k times (k determined later). Let $A_1, \dots, A_k$ be the resulting "subsets" of A_0, $P_1, \dots, P_k$ be the involved $\mathbb{Z}$-progressions of lengths $n_1, \dots, n_k$ and let $\tau_j = |A_j \cap P_j|/|P_j|$ for every $j = 1, \dots, k$. In each step we can guarantee that

$$n_j \geq \frac{\tau^2}{5000}\sqrt{n_{j-1}} \quad \text{and} \quad \tau_j \geq \tau_{j-1} + \frac{\tau_{j-1}^2}{800}\,,$$

where $\tau_0 = \tau$ and $n_0 = n$. A simple calculation shows that for every $j \in [k]$ we have

$$n_j \geq \left(\frac{\tau^2}{5000}\right)^{\sum_{\ell=0}^{j-1} 2^{-\ell}} n^{1/2^j} \geq \frac{\tau^4}{5000^2} n^{1/2^j}\,. \tag{3.32}$$

Clearly, at most $k^* = 800/\tau^2$ such iterations are possible, since then the density would be bigger than 1. However, this would only give a contradiction if we would require n to be doubly exponential in C/τ^2 for some appropriate C (instead of C/τ as chosen in (3.31)). So for a better estimate on the maximal number of iterations we take also into account, that the τ_j increase and, hence, the density increments increase.

For that we observe that after at most $j_1 = \lceil 800/\tau \rceil$ steps we have $\tau_{j_1} \geq 2\tau$ and from then on in each step we obtain a density increment of $4\tau^2/800 = \tau^2/200$. After additional $400/\tau$ steps the density doubled again, i.e., for $j_2 = \lceil 800/\tau \rceil + \lceil 400/\tau \rceil$ we have $\tau_{j_2} \geq 4\tau$ and from then on we obtain density increments of at least $16\tau^2/800$. Continuing this way we see that after at most

$$k = \sum_{\ell=0}^{\lceil \log_2(1/\tau) \rceil} \left\lceil \frac{800}{2^\ell \tau} \right\rceil < \lceil \log_2(1/\tau) \rceil + \sum_{\ell=0}^{\infty} \frac{800}{2^\ell \tau} = \lceil \log_2(1/\tau) \rceil + \frac{1600}{\tau} \leq \frac{1601}{\tau}$$

iterations we would arrive at the contradiction $\tau_k > 1$ if $n_{k-1} \geq 50/\tau^3$ (otherwise we would not have been allowed to apply Proposition 3.14 in the kth step). Hence, in view of (3.32) we obtain the desired contradiction if

$$\frac{\tau^4}{5000^2} n^{1/2^{k-1}} > \frac{50}{\tau^3},$$

which is ensured by

$$\ln\ln(n) \overset{(3.31)}{>} \frac{1200}{\tau} > \frac{1601\ln(2)}{\tau} + \ln\ln(5^3 10^7 \tau^{-7}).$$

This concludes the proof of Theorem 3.11 based on Propositions 3.13 and 3.14. □

It is left to verify Propositions 3.13 and 3.14. We begin with Proposition 3.14, which establishes the crucial connection between small non-trivial Fourier coefficients of the function $\mathbf{1}_A$ and the existence of AP_3's in A.

Proof of Proposition 3.14. We split $\mathbb{Z}/n\mathbb{Z}$ evenly into the three intervals $I_1 = \{0, \dots, \lfloor n/3 \rfloor - 1\}$, $I_2 = \{\lfloor n/3 \rfloor, \dots, \lfloor 2n/3 \rfloor - 1\}$, and $I_3 = \{\lfloor 2n/3 \rfloor, \dots, n-1\}$ and set $B = A \cap I_2$. Note that I_1, I_2, and I_3 form $\mathbb{Z}$-progressions of length either $\lfloor n/3 \rfloor$ of $\lceil n/3 \rceil$.

If $|B| < \tau n/5$, then it follows from $n > 42$, that either $|A \cap I_1| \geq 7\tau|I_1|/6$ or $|A \cap I_3| \geq 7\tau|I_3|/6$. In other words, if we assume that alternative *(ii)* does not occur, then

$$|B| \geq \frac{\tau}{5} n. \tag{3.33}$$

We shall estimate the number $N(A)$ of triples $(x, y, z) \in A \times B \times B$ with the property $x + z \equiv 2y \pmod{n}$. Roughly speaking, $N(A)$ is a lower bound on the number of AP_3's in A. Moreover, owing to the choice of B we have that $N(A)$ only counts $\mathbb{Z}$-progressions. On the other hand, we do not require the elements of (x, y, z) to be different and, as a consequence, $N(A)$ includes degenerate progressions of the form $x = y = z$. However, degenerate AP_3's are fixed by choosing one element from B, i.e., there are at most $|B| \leq n$ degenerate AP_3's and the number of (non-degenerate) AP_3's contained in A which are $\mathbb{Z}$-progressions is at least $N(A) - n$. Hence, it suffices to show

$$N(A) > n. \tag{3.34}$$

Letting $e_n(\cdot)$ denote the function from $\mathbb{Z}/n\mathbb{Z}$ to $\mathbb{C}$ defined by $x \mapsto \exp(2\pi \boldsymbol{i} x/n)$, we obtain the following crucial identity

$$\begin{aligned} N(A) &\overset{(B.2)}{=} \frac{1}{n} \sum_{r \in \mathbb{Z}/n\mathbb{Z}} \sum_{x \in A} \sum_{y \in B} \sum_{z \in B} e_n((2y - x - z)r) \\ &= \frac{1}{n} \sum_{r \in \mathbb{Z}/n\mathbb{Z}} \left(\sum_{x \in \mathbb{Z}/n\mathbb{Z}} \mathbf{1}_A(x) e_n(-xr) \sum_{y \in \mathbb{Z}/n\mathbb{Z}} \mathbf{1}_B(y) e_n(2yr) \sum_{z \in \mathbb{Z}/n\mathbb{Z}} \mathbf{1}_B(z) e_n(-zr) \right) \\ &\overset{(B.3)}{=} \frac{1}{n} \sum_{r \in \mathbb{Z}/n\mathbb{Z}} \hat{\mathbf{1}}_A(r) \cdot \hat{\mathbf{1}}_B(-2r) \cdot \hat{\mathbf{1}}_B(r). \end{aligned}$$

Consequently, it follows from the Cauchy-Schwarz inequality that

$$N(A) \geq \frac{1}{n}|A||B|^2 - \frac{1}{n} \max_{r \neq 0} |\hat{\mathbf{1}}_A(r)| \cdot \left(\sum_{r \neq 0} |\hat{\mathbf{1}}_B(-2r)|^2 \right)^{1/2} \cdot \left(\sum_{r \neq 0} |\hat{\mathbf{1}}_B(r)|^2 \right)^{1/2}.$$

Since there are at most two solutions for the equation $s \equiv -2r \pmod{n}$ for every fixed $s \in \mathbb{Z}/n\mathbb{Z}$ we have $\sum_{r\neq 0} |\hat{\mathbf{1}}_B(-2r)|^2 \leq 2\sum_{r\neq 0} |\hat{\mathbf{1}}_B(r)|^2$ and appealing to the assumption of Proposition 3.14 we obtain

$$N(A) \geq \frac{|A||B|^2}{n} - \frac{\tau^2}{50\sqrt{2}} \sum_{r\neq 0} |\hat{\mathbf{1}}_B(r)|^2 \,.$$

Parseval's identity yields

$$\begin{aligned} N(A) &\geq \frac{|A||B|^2}{n} - \frac{\tau^2}{50} \sum_{r\in\mathbb{Z}/n\mathbb{Z}} |\hat{\mathbf{1}}_B(r)|^2 \\ &\overset{(B.6)}{=} \frac{|A||B|^2}{n} - \frac{\tau^2}{50} n \sum_{x\in\mathbb{Z}/n\mathbb{Z}} |\mathbf{1}_B(r)|^2 \\ &\geq \frac{|A||B|^2}{n} - \frac{\tau^2}{50} n|B| \,. \end{aligned}$$

Finally, owing to $|A| \geq \tau n$, $B \subseteq A$, and (3.33) we infer

$$N(A) \geq |A| \left(\frac{|B|^2}{n} - \frac{\tau^2}{50} n \right) \geq \frac{\tau^3}{50} n^2$$

and (3.34) follows from $n > 50/\tau^3$. $\square$

Finally we verify Proposition 3.13.

Proof of Proposition 3.13. Let $A \subseteq \mathbb{Z}/n\mathbb{Z}$ $|A| \geq \tau n$ for some $\tau > 0$ and assume that alternative (i) of Proposition 3.13 fails, i.e., we assume

$$|\hat{\mathbf{1}}_A(r)| \geq \frac{\tau^2}{100} n \quad \text{for some } r \in \mathbb{Z}/n\mathbb{Z} \setminus \{0\}. \tag{3.35}$$

We define the weighted indicator function f_A of A by

$$f_A(x) = \mathbf{1}_A(x) - \alpha \,, \quad \text{where} \quad \alpha = \frac{|A|}{n} \geq \tau \,,$$

for which we have

$$\sum_{x\in\mathbb{Z}/n\mathbb{Z}} f_A(x) = 0 \,. \tag{3.36}$$

For the moment let $Q \subseteq \mathbb{Z}/n\mathbb{Z}$ be some arithmetic progression (later we will make a more careful choice) and for $x \in \mathbb{Z}/n\mathbb{Z}$ let Q_x be the arithmetic progression given by

$$Q_x = \{x - z \colon\; z \in Q\} \,.$$

We consider the convolution of f_A with $\mathbf{1}_Q$, i.e., we set $h_{A,Q} = f_A * \mathbf{1}_Q$, and observe that

$$\begin{aligned} h_{A,Q}(x) &= \sum_{y\in\mathbb{Z}/n\mathbb{Z}} f_A(y)\mathbf{1}_Q(x-y) \\ &= \sum_{y\in\mathbb{Z}/n\mathbb{Z}} \mathbf{1}_A(y)\mathbf{1}_Q(x-y) - \alpha \sum_{y\in\mathbb{Z}/n\mathbb{Z}} \mathbf{1}_Q(x-y) \\ &= \sum_{y\in\mathbb{Z}/n\mathbb{Z}} \mathbf{1}_A(y)\mathbf{1}_{Q_x}(y) - \alpha \sum_{y\in\mathbb{Z}/n\mathbb{Z}} \mathbf{1}_{Q_x}(y) \\ &= |A \cap Q_x| - \alpha|Q_x| \,. \end{aligned} \tag{3.37}$$

Hence, ignoring for a moment that we are interested in $\mathbb{Z}$-progressions only, we want to find an arithmetic progression Q and an $x \in \mathbb{Z}/n\mathbb{Z}$ such that

$$h_{A,Q}(x) = c|Q_x| = c|Q| \quad \text{and} \quad |Q| = |Q_x| \geq c'\sqrt{n} \tag{3.38}$$

for appropriate constants c, $c' > 0$. For that we appeal to the inequality

$$|\hat{f}_A(r)||\hat{\mathbf{1}}_Q(r)| \overset{(B.7)}{=} |\hat{h}_{A,Q}(r)| \overset{(B.4)}{\leq} \sum_{x\in\mathbb{Z}/n\mathbb{Z}} |h_{A,Q}(x)|\,. \tag{3.39}$$

Owing to $r \neq 0$ we have

$$\hat{f}_A(r) = \hat{\mathbf{1}}_A(r) - \alpha \sum_{x\in\mathbb{Z}/n\mathbb{Z}} e_n(-xr) \overset{(B.2)}{=} \hat{\mathbf{1}}_A(r)\,. \tag{3.40}$$

Moreover, it follows from the definition of $h_{A,Q}$ that

$$\begin{aligned}\sum_{x\in\mathbb{Z}/n\mathbb{Z}} h_{A,Q}(x) &= \sum_{x\in\mathbb{Z}/n\mathbb{Z}}\sum_{y\in\mathbb{Z}/n\mathbb{Z}} f_A(y)\mathbf{1}_Q(x-y)\\ &= \sum_{y\in\mathbb{Z}/n\mathbb{Z}} f_A(y) \sum_{x\in\mathbb{Z}/n\mathbb{Z}} \mathbf{1}_Q(x-y) = \sum_{y\in\mathbb{Z}/n\mathbb{Z}} f_A(y)|Q| \overset{(3.36)}{=} 0\,.\end{aligned} \tag{3.41}$$

Below we shall show that there exists an arithmetic progression $Q \subseteq \mathbb{Z}/n\mathbb{Z}$ of length $2m+1$, where m is the largest integer such that

$$m \leq \frac{n}{6\lceil\sqrt{n}\,\rceil} \tag{3.42}$$

such that

$$|\hat{\mathbf{1}}_Q(r)| \geq \frac{|Q|}{2} \tag{3.43}$$

and we deduce (3.38) from this. In fact, summarising the above we infer

$$\begin{aligned}\sum_{x\in\mathbb{Z}/n\mathbb{Z}} \big(h_{A,Q}(x) + |h_{A,Q}(x)|\big) &\overset{(3.41)}{=} \sum_{x\in\mathbb{Z}/n\mathbb{Z}} |h_{A,Q}(x)|\\ &\overset{(3.39)}{\geq} |\hat{f}_A(r)||\hat{\mathbf{1}}_Q(r)| \overset{(3.35),(3.43)}{\geq} \frac{\tau^2}{200} n|Q|\,.\end{aligned}$$

Consequently, there exists an $x \in \mathbb{Z}/n\mathbb{Z}$ such that $h_{A,Q}(x) + |h_{A,Q}(x)| \geq \tau^2|Q|/200$ and, therefore,

$$h_{A,Q}(x) \geq \frac{\tau^2}{400}|Q|\,,$$

which yields (3.38).

Next we verify the existence of a progression Q of length $2m+1$, which satisfies (3.43). We consider the $\lceil\sqrt{n}\,\rceil + 1$ elements of the form $jr \pmod n$ for $j = 0, \dots, \lceil\sqrt{n}\,\rceil$. Owing to the pigeonhole principle, there exist numbers j_1 and j_2 with $0 \leq j_1 < j_2 \leq \lceil\sqrt{n}\,\rceil$ such that $j_2 r \pmod n$ and $j_1 r \pmod n$ are at most $\lceil\sqrt{n}\,\rceil$ apart. In other words, for $\lambda = j_2 - j_1$ we have

$$\lambda \leq \lceil\sqrt{n}\,\rceil \quad \text{and} \quad \lambda r \leq \lceil\sqrt{n}\,\rceil\,. \tag{3.44}$$

We then set

$$Q = \{-m\lambda, (m-1)\lambda, \dots, -\lambda, 0, \lambda, \dots, m\lambda\} \subseteq \mathbb{Z}/n\mathbb{Z}\,. \tag{3.45}$$

Owing to (3.42) and (3.44) we have $m\lambda \leq m\lceil\sqrt{n}\,\rceil \leq n/6$ and, hence, the arithmetic progression Q is the union of two $\mathbb{Z}$-progressions. Moreover, for every $x \in \mathbb{Z}/n\mathbb{Z}$ the shifted progression Q_x is a union of two $\mathbb{Z}$-progressions.

Using the identity $e_n(-z) + e_n(z) = 2\cos(2\pi z/n)$ we obtain

$$\hat{\mathbf{1}}_Q(r) = \sum_{x\in\mathbb{Z}/n\mathbb{Z}} \mathbf{1}_Q(x)e_n(-xr) = \sum_{k=-m}^{m} e_n(-k\lambda r) = 1 + 2\sum_{k=1}^{m} \cos\Big(\frac{2\pi k\lambda r}{n}\Big)\,.$$

Since

$$\frac{2\pi m\lambda r}{n} \overset{(3.44)}{\leq} \frac{2\pi m\lceil\sqrt{n}\,\rceil}{n} \overset{(3.42)}{\leq} \frac{\pi}{3}$$

and since $\cos(\xi) \geq 1/2$ for every ξ with $0 \leq \xi \leq \pi/3$ we arrive at

$$\hat{\mathbf{1}}_Q(r) \geq 1 + m > \frac{|Q|}{2},$$

which gives (3.43).

Summarising, we have shown that there exists an arithmetic progression Q in $\mathbb{Z}/n\mathbb{Z}$ of length $2m+1$, which is the union of two $\mathbb{Z}$-progressions and some $x \in ZZ$ such that $h_{A,Q}(x) \geq \tau^2|Q|/400$ and, hence, we infer from (3.37) that

$$|A \cap Q_x| \geq \alpha|Q_x| + \frac{\tau^2}{400}|Q_x| \geq \left(\tau + \frac{\tau^2}{400}\right)|Q_x|.$$

Finally, we shall pass to a $\mathbb{Z}$-progression. Recall that Q_x is the union of two $\mathbb{Z}$-progressions, say P and P'. If $|P'| < \tau^2|Q_x|/800$, then

$$|A \cap P| \geq |A \cap Q_x| - \frac{\tau^2}{800}|Q_x| \geq \left(\tau + \frac{\tau^2}{800}\right)|Q_x| \geq \left(\tau + \frac{\tau^2}{800}\right)|P|$$

and

$$|P| \geq \left(1 - \frac{\tau^2}{800}\right)(2m+1) > \frac{\tau^2}{5000}\sqrt{n}.$$

In other words, in this case we arrived at conclusion *(ii)*. If on the other hand, both P and P' are $\mathbb{Z}$-progressions of length at least $\tau^2|Q_x|/800$, then for at least one of them, say P, we have

$$\frac{|A \cap P|}{|P|} \geq \frac{|A \cap Q_x|}{|Q_x|} \geq \tau + \frac{\tau^2}{400}$$

and standard calculation using $n \geq 50$ and the definition of m in (3.42) give

$$\frac{\tau^2}{800}|Q_x| = \frac{\tau^2}{800}(2m+1) \geq \frac{\tau^2}{5000}\sqrt{n}.$$

In other words, we again arrive at conclusion *(ii)* of Proposition 3.13. $\square$

CHAPTER 4

The Hales–Jewett theorem

This chapter is devoted to the Hales–Jewett theorem (Theorem 1.11). The Hales–Jewett theorem can be used to deduce many other results in Ramsey theory and in Section 4.1 we give a few example. In Section 4.2 we present Shelah's proof [80] of Theorem 1.11. Here we follow the presentation of Alon from [68] (see also [50]).

4.1. Applications of the Hales–Jewett theorem

We begin with some applications of Theorem 1.11. First we derive a multidimensional version of the Hales–Jewett theorem in Section 4.1.1. In Section 4.1.2 we deduce Theorem 1.7 and in Section 4.1.3 we reduce Theorems 1.8 and 1.9 from the Hales–Jewett theorem.

4.1.1. Multidimensional version of the Hales–Jewett theorem. We recall the definition of a combinatorial line, Definition 1.10 and, more generally, we define combinatorial spaces.

DEFINITION 4.1. Let A be a finite set of cardinality k and let n, $d \geq 1$ be integers. For an integer $m \geq 1$ let A^m be the set of all functions from $[m]$ to A.

We say a k^d-element subset $\mathcal{S} = \{f_{\mathfrak{a}} \colon \mathfrak{a} \in A^d\} \subseteq A^n$ is a *combinatorial d-space in A^n* if there exist d pairwise disjoint non-empty subsets $X_1 \dot\cup \dots \dot\cup X_d \subseteq [n]$ and a function $g \colon [n] \setminus \bigcup_{j\in[d]} X_j \to A$ such that for every $\mathfrak{a} \in A^d$ we have

(i) $f_{\mathfrak{a}}(x) = g(x)$ for every $x \in [n] \setminus \bigcup_{j\in[d]} X_j$ and
(ii) $f_{\mathfrak{a}}(x) = \mathfrak{a}(i)$ for every $i \in [d]$ and $x \in X_i$.

Clearly, combinatorial 1-spaces are combinatorial lines (see Definition 1.10) and the following can be viewed as a multidimensional version of Theorem 1.11.

THEOREM 4.2. *For all integers r, $d \geq 1$, and every finite alphabet A there exists some integer n_0 such that for every $n \geq n_0$ the following holds. For any partition $E_1 \dot\cup \dots \dot\cup E_r$ of A^n there exists some $j \in [r]$ such that E_j contains a combinatorial d-space.*

We derive Theorem 4.2 as a simple corollary of Theorem 1.11.

PROOF OF: THEOREM 1.11 $\Rightarrow$ THEOREM 4.2. Let r, d, and $A = \{a_1, \dots, a_k\}$ be given. Let $N = n_0(r, B)$ be the constant ensured by Theorem 1.11 applied with r and the alphabet $B = A^d$ and set $n_0 = dN$.

Let $n \geq n_0$. Without loss of generality we may assume that n is divisible by d. Indeed if $n = dm + s$ for some $m \geq n_0/d$ and $1 \leq s < d$, then we consider an arbitrary projection of A^n to A^{dm} by fixing s coordinates. For example we consider only those functions $f \in A^n$ with $f(dm+1) = \dots = f(dm+s) = a_1$. Clearly, any combinatorial d-space within this restriction forms a combinatorial d-space in A^n and we are done.

Therefore, let $n = dm$ for some $m \geq n/d \geq N$. We partition (in an arbitrary way) $[n]$ into m ordered sets of size d. For simplicity let $I_1 \dot\cup \dots \dot\cup I_m = [n]$, where

I_j is the interval $\{(j-1)d+1,\dots,jd\}$. Now we may view A^n as a copy of B^m by letting $f\in A^n$ correspond to $F_f\in B^m$ where

$$F_f(j)=f(I_j):=\big(f((j-1)d+1),\dots,f(jd)\big)\in A^d=B\,.$$

It is easy to check that this correspondence defines a bijection between A^n and B^m and for every partition $E_1\dot\cup\dots\dot\cup E_r$ of A^n we obtain a partition $E_1'\dot\cup\dots\dot\cup E_r$ of B^m, where $F_f\in E_j'$ if and only if $f\in E_j$. Since $m\geq N$, there exists some $j\in[r]$ such that E_j' contains a combinatorial line in B^m. This line consists of $|B|=|A|^d$ functions from B^m and it follows from the bijection between A^n and B^m that the corresponding $|A|^d$ functions in A^n form a combinatorial d-space. $\square$

4.1.2. Proof of the Gallai–Witt theorem. Next we deduce the multidimensional version of van der Waerden's theorem (so-called Gallai–Witt theorem, Theorem 1.7) from the Hales–Jewett theorem.

Proof of: Theorem 1.11 $\Rightarrow$ Theorem 1.7. Let r, $d\geq 1$ be integers and let $F=\{\vec u_1,\dots,\vec u_k\}\subset\mathbb{N}^d$. Let N be given by Theorem 1.11 applied with r and alphabet F. Set n_0 be sufficiently large such that $m_1\vec u_1+\dots+m_k\vec u_k\in[n_0]^d$ for every choice of integers $m_1,\dots,m_k\geq 0$ with $m_1+\dots+m_k=N$.

Let $n\geq n_0$ and $E_1\dot\cup\dots\dot\cup E_r$ be a partition of $[n]^d$. We have to show that one of the partition classes $E_1,\dots,E_r$ contains a homothetic copy of F. We consider the following partition $E_1',\dot\cup\dots\dot\cup E_r'$ of F^N defined for every $f\colon[N]\to F$ by

$$f\in E_j'\qquad\Longleftrightarrow\qquad\sum_{x=1}^{N}f(x)\in E_j\,.$$

Note that $\sum_{x=1}^N f(x)$ is an element in $[n_0]^d\subseteq[n]^d$ due to the choice of n_0. (For $N\geq 2$ the map $f\mapsto\sum_{x=1}^N f(x)$ from A^N to $[n]^d$ is not injective, but we do not have to worry about that here.)

Theorem 1.11 guarantees a $j\in[r]$ and a combinatorial line $\{f_1,\dots,f_k\}\subseteq E_j'$, i.e., there exist a non-empty set $X\subseteq[N]$ and $g\colon[N]\setminus X\to F$ such that for every $i\in[k]$ we have

$$f_i(x)=\begin{cases}\vec u_i & \text{if } x\in X,\\ g(x) & \text{otherwise,}\end{cases}\qquad\text{and}\qquad f_i\in E_j'\,.$$

Owing to the definition of the partition $E_1'\dot\cup\dots\dot\cup E_r'$ we have

$$\sum_{x=1}^{N}f_i(x)\in E_j$$

and setting $\vec v_0=\sum_{x\in[N]\setminus X}g(x)$ and $\lambda=|X|>0$ we see that

$$\vec v_0+\lambda\vec u_i=\sum_{x=1}^{N}f_i(x)\in E_j$$

for every $i\in[k]$ In other words, E_j contains a homothetic copy of F. $\square$

4.1.3. Ramsey's theorem for vector spaces. In this section we present Spencer's proof from [83] of Ramsey's theorem for (affine) vector spaces over finite fields, Theorems 1.8 and 1.9. Those results were conjectured by Gian-Carlo Rota and first proved by Graham, Leeb, and Rothschild [41]. We begin with the proof of the affine version.

4.1.3.1. *Proof of Theorem 1.9.* In this section let $\mathbb{F} = \mathrm{GF}(q)$ be the finite field consisting of q elements. For an affine subspace $\mathcal{U} \subseteq \mathbb{F}^n$ we denote by $\dim(\mathcal{U})$ the dimension of $\mathcal{U}$ and for an integer $k \geq 0$ we denote by $\left[\begin{smallmatrix}\mathcal{U}\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$ the set of all k-dimensional affine subspaces of $\mathcal{U}$. In particular, $\left[\begin{smallmatrix}\mathbb{F}^n\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$ denotes the set of k-dimensional affine subspaces of $\mathbb{F}^n$ and clearly we have $\left[\begin{smallmatrix}\mathcal{U}\\ k\end{smallmatrix}\right]_{\mathrm{aff}} \subseteq \left[\begin{smallmatrix}\mathbb{F}^n\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$.

Suppose $\mathcal{U} \in \left[\begin{smallmatrix}\mathbb{F}^n\\ d+1\end{smallmatrix}\right]_{\mathrm{aff}}$ has the property that the projection $\pi_d\colon \mathcal{U} \to \mathbb{F}^d$ onto the first d coordinates is surjective, i.e., $\pi_d(\mathcal{U}) = \mathbb{F}^d$. Note that for every $\mathcal{W} \in \left[\begin{smallmatrix}\mathcal{U}\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$ we have $\dim(\pi_d(\mathcal{W})) \in \{k-1, k\}$. We say $\mathcal{W}$ is *transversal* if $\dim(\pi_d(\mathcal{W})) = k$.

Let $E_1 \dot\cup \ldots \dot\cup E_r$ be a partition of $\left[\begin{smallmatrix}\mathbb{F}^n\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$. We say $\mathcal{U} \in \left[\begin{smallmatrix}\mathbb{F}^n\\ d+1\end{smallmatrix}\right]_{\mathrm{aff}}$ is *special* (with respect to $E_1 \dot\cup \ldots \dot\cup E_r = \left[\begin{smallmatrix}\mathbb{F}^n\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$) if $\pi_d(\mathcal{U}) = \mathbb{F}^d$ and for all transversal $\mathcal{W}_1$, $\mathcal{W}_2 \in \left[\begin{smallmatrix}\mathcal{U}\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$ with $\pi_d(\mathcal{W}_1) = \pi_d(\mathcal{W}_2)$ we have that $\mathcal{W}_1$ and $\mathcal{W}_2$ are contained in the same partition class from $E_1 \dot\cup \ldots \dot\cup E_r = \left[\begin{smallmatrix}\mathbb{F}^n\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$. In other words, $\mathcal{U}$ is special if the partition class of every transversal $\mathcal{W} \in \left[\begin{smallmatrix}\mathcal{U}\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$ is determined by its image under the projection π_d. The following proposition is the key lemma in the proof of Theorem 1.9. It asserts the existence of special spaces $\mathcal{U} \in \left[\begin{smallmatrix}\mathbb{F}^n\\ d+1\end{smallmatrix}\right]_{\mathrm{aff}}$ for any $d \geq k$ in any partition in a sufficiently high dimensional vector space over $\mathbb{F}$.

PROPOSITION 4.3. *Let $\mathbb{F}$ be a finite field. For all integers $d \geq k \geq 1$ and $r \geq 1$ there exists an m_0 such that for every integer $m \geq m_0$, every partition $E_1 \dot\cup \ldots \dot\cup E_r$ of $\left[\begin{smallmatrix}\mathbb{F}^m\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$, there exists a $(d+1)$-dimensional affine space $\mathcal{U} \in \left[\begin{smallmatrix}\mathbb{F}^m\\ d+1\end{smallmatrix}\right]_{\mathrm{aff}}$ which is special with respect to the partition $E_1 \dot\cup \ldots \dot\cup E_r$.*

PROOF. Let $\mathbb{F}$ be a finite field and $d \geq k \geq 1$ and $r \geq 1$ be integers. Let μ be the number of k-dimensional affine subspaces in a d-dimensional affine space over $\mathbb{F}$, i.e.,

$$\mu = \left| \begin{bmatrix}\mathbb{F}^d\\ k\end{bmatrix}_{\mathrm{aff}} \right| .$$

The proof relies on an application of the Hales–Jewett theorem and we let n_0 be given by Theorem 1.11 applied with the alphabet $\mathbb{F}^{d+1}$ and number of partition classes r^μ. We set

$$m_0 = n_0 + d$$

and for $m \geq m_0$ let $E_1 \dot\cup \ldots \dot\cup E_r$ be a partition of $\left[\begin{smallmatrix}\mathbb{F}^m\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$. As above we denote by π_d the projection of $\mathbb{F}^m$ onto the first d coordinates. For convenience we set

$$m' = m - d \geq m_0 \,. \tag{4.1}$$

For $\vec{c} = (c_0, \ldots, c_d) \in \mathbb{F}^{d+1}$ we consider the affine transformation $f_{\vec{c}}\colon \mathbb{F}^d \to \mathbb{F}$ defined for every $\vec{x} = (x_1, \ldots, x_d) \in \mathbb{F}^d$ by

$$f(\vec{x}) = c_0 + \sum_{i=1}^{d} c_i x_i \,.$$

Furthermore, for any $\boldsymbol{C} = (\vec{c}_1, \ldots, \vec{c}_{m'}) \in (\mathbb{F}^{d+1})^{m'}$ let $F_{\boldsymbol{C}}\colon \mathbb{F}^d \to \mathbb{F}^m$ be the lift of $\mathbb{F}^d$ to $\mathbb{F}^m$ defined for every $\vec{x} = (x_1, \ldots, x_d) \in \mathbb{F}^d$ by

$$F_{\boldsymbol{C}} = \big(x_1, \ldots, x_d, f_{c_1}(\vec{x}), \ldots, f_{c_{m'}}(\vec{x})\big) \,.$$

It follows from those definitions that for every $\boldsymbol{C} \in (\mathbb{F}^{d+1})^{m'}$ the map $F_{\boldsymbol{C}}$ is an injective, affine transformation, which is inverse to the projection π_d. In particular, for every $\mathcal{W}' \in \left[\begin{smallmatrix}\mathbb{F}^d\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$ the image $F_{\boldsymbol{C}}(\mathcal{W}')$ is again a k-dimensional affine space in $\mathbb{F}^m$, i.e., $F_{\boldsymbol{C}}(\mathcal{W}') \in \left[\begin{smallmatrix}\mathbb{F}^m\\ k\end{smallmatrix}\right]_{\mathrm{aff}}$ and, therefore, for every such $\mathcal{W}'$ the space $F_{\boldsymbol{C}}(\mathcal{W}')$ is contained in one of the partition classes $E_1, \ldots, E_r$.

Next we consider a partition $\dot\bigcup_{\vec r\in[r]^\mu} E'_{\vec r}$ of $(\mathbb{F}^{d+1})^{m'}$ into r^μ parts. For that fix some enumeration $\mathcal{W}'_1,\dots,\mathcal{W}'_\mu$ of $\begin{bmatrix}\mathbb{F}^d\\ k\end{bmatrix}_{\mathrm{aff}}$. For $\vec r=(r_1,\dots,r_\mu)\in[r]^\mu$ we include $\boldsymbol{C}\in(\mathbb{F}^{d+1})^{m'}$ in $E'_{\vec r}$ if

$$F_{\boldsymbol{C}}(\mathcal{W}'_i)\in E_{r_i}\quad\text{for every } i\in[\mu]\,. \tag{4.2}$$

Owing to (4.1) and the choice of m_0 we infer from the Hales–Jewett theorem, Theorem 1.11, that there exists some $\vec s=(s_1,\dots,s_\mu)\in[r]^\mu$ such that $E'_{\vec s}$ contains a combinatorial line $\mathcal{L}$ in $(\mathbb{F}^{d+1})^{m'}$. Below we will show that the combinatorial line $\mathcal{L}$ can be used to define a special space $\mathcal{U}\in\begin{bmatrix}\mathbb{F}^m\\ d+1\end{bmatrix}_{\mathrm{aff}}$.

Since $\mathcal{L}$ is a combinatorial line in $(\mathbb{F}^{d+1})^{m'}$ there exist a non-empty set $M\subseteq[m']$ and a function $g\colon\ [m']\setminus M\to\mathbb{F}^{d+1}$ such that

$$\mathcal{L}=\big\{\boldsymbol{C}_{\vec c}=(\vec c_1,\dots,\vec c_{m'})\colon\ \vec c\in\mathbb{F}^{d+1}\big\},\quad\text{where } \vec c_i=\begin{cases}\vec c & \text{if } i\in M,\\ g(i) & \text{otherwise.}\end{cases} \tag{4.3}$$

Without loss of generality we may assume that M consists of the last $|M|$ elements of $[m']$, i.e., $[m']\setminus M=[m'']$ for $m''=m'-|M|$. We set

$$\begin{aligned}\mathcal{U}&=\bigcup_{\boldsymbol{C}\in\mathcal{L}}\ \bigcup_{\vec x\in\mathbb{F}^d}F_{\boldsymbol{C}}(\vec x)=\bigcup_{\vec c\in\mathbb{F}^{d+1}}\ \bigcup_{\vec x\in\mathbb{F}^d}F_{\boldsymbol{C}_{\vec c}}(\vec x)\\ &=\bigcup_{\substack{\vec x\in\mathbb{F}^d\\ \vec x=(x_1,\dots,x_d)}}\ \bigcup_{\vec c\in\mathbb{F}^{d+1}}\Big\{\big(x_1,\dots,x_d,f_{g(1)}(\vec x),\dots,f_{g(m'')}(\vec x),\underbrace{f_{\vec c}(\vec x),\dots,f_{\vec c}(\vec x)}_{|M|\text{-times}}\big)\Big\}\end{aligned} \tag{4.4}$$

and claim that $\mathcal{U}$ has the desired properties.

Recall that for every $\vec c\in\mathbb{F}^{d+1}$ the map $f_{\vec c}$ maps $\mathbb{F}^d$ to $\mathbb{F}$ and $\bigcup_{\vec c\in\mathbb{F}^d}f_{\vec c}(\vec x)=\mathbb{F}$ for every $\vec x\in\mathbb{F}^d$. Consequently,

$$\mathcal{U}=\bigcup_{\vec x=(x_1,\dots,x_d)\in\mathbb{F}^d}\ \bigcup_{y\in\mathbb{F}}\Big\{\big(x_1,\dots,x_d,f_{g(1)}(\vec x),\dots,f_{g(m'')}(\vec x),\underbrace{y,\dots,y}_{|M|\text{-times}}\big)\Big\}\,. \tag{4.5}$$

Moreover, $f_{\vec c}$ is an affine transformation for every $\vec c\in\mathbb{F}^{d+1}$ and, hence,

$$(x_1,\dots,x_d,y)\mapsto\big(x_1,\dots,x_d,f_{g(1)}(\vec x),\dots,f_{g(m'')}(\vec x),\underbrace{y,\dots,y}_{|M|\text{-times}}\big)$$

is an affine (and clearly injective) transformation. In other words, $\mathcal{U}$ is isomorphic to $\mathbb{F}^{d+1}$, i.e., $\mathcal{U}\in\begin{bmatrix}\mathbb{F}^n\\ d+1\end{bmatrix}_{\mathrm{aff}}$. It also follows from (4.5) that $\pi_d(\mathcal{U})=\mathbb{F}^d$.

Let $\mathcal{W}_1,\mathcal{W}_2\in\begin{bmatrix}\mathcal{U}\\ k\end{bmatrix}_{\mathrm{aff}}$ be transversal with

$$\pi_d(\mathcal{W}_1)=\pi_d(\mathcal{W}_2)\in\begin{bmatrix}\mathbb{F}^d\\ k\end{bmatrix}_{\mathrm{aff}}.$$

Recall, that we enumerated the elements of $\begin{bmatrix}\mathbb{F}^d\\ k\end{bmatrix}_{\mathrm{aff}}$ and let $\mathcal{W}'_i=\pi_d(\mathcal{W}_1)=\pi_d(\mathcal{W}_2)$ for some $i\in[\mu]$.

In order to show that $\mathcal{U}$ is special, it is left to show that $\mathcal{W}_1$ and $\mathcal{W}_2$ are elements from the same partition class of the given partition $E_1\dot\cup\dots\dot\cup E_r=\begin{bmatrix}\mathbb{F}^m\\ k\end{bmatrix}_{\mathrm{aff}}$. Observe that there exist affine transformation $F_1,F_2\colon\ \mathbb{F}^d\to\mathcal{U}\subseteq\mathbb{F}^m$ such that

$$F_1(\mathcal{W}'_i)=\mathcal{W}_1\quad\text{and}\quad F_2(\mathcal{W}'_i)=\mathcal{W}_2\,.$$

In view of (4.3) and (4.4) there exist $\vec c_1,\vec c_2\in\mathbb{F}^{d+1}$ such that the functions

$$F_1=F_{\boldsymbol{C}_{\vec c_1}}\quad\text{and}\quad F_2=F_{\boldsymbol{C}_{\vec c_2}}$$

have this property. Owing to $\boldsymbol{C}_{\vec c_1},\boldsymbol{C}_{\vec c_2}\in\mathcal{L}$ we have $\boldsymbol{C}_{\vec c_1},\boldsymbol{C}_{\vec c_2}\in E'_{\vec s}$ and it follows from (4.2) that

$$W_1=F_1(W'_i)=F_{\boldsymbol{C}_{\vec c_1}}(\mathcal{W}'_i)\in E_{s_i}\quad\text{and}\quad W_2=F_2(W'_i)=F_{\boldsymbol{C}_{\vec c_2}}(\mathcal{W}'_i)\in E_{s_i}\,,$$

which concludes the proof of Proposition 4.3. □

Next we deduce Theorem 1.9 from Proposition 4.3. For the inductive proof it will be convenient to consider the following version of Theorem 1.9.

PROPOSITION 4.4. *For all integers $r \geq 1$ and $\ell_1, \dots, \ell_r \geq k \geq 0$ and every finite field $\mathbb{F}$ there exists some integer $N^{(k)}(\ell_1, \dots, \ell_r)$ such that for every integer $n \geq N^{(k)}(\ell_1, \dots, \ell_r)$ the following holds. For any partition $E_1 \dot\cup \dots \dot\cup E_r$ of $\begin{bmatrix}\mathbb{F}^n\\ k\end{bmatrix}_{\mathrm{aff}}$ there exist $j \in [r]$ and $\mathcal{S} \in \begin{bmatrix}\mathbb{F}^n\\ \ell_j\end{bmatrix}_{\mathrm{aff}}$ such that $\begin{bmatrix}\mathcal{S}\\ k\end{bmatrix}_{\mathrm{aff}} \subseteq E_j$.*

PROOF. Let $\mathbb{F}$ be a finite field. We proceed by double induction on k and $\sum_{j=1}^{r} \ell_j$.

For $k = 0$ Proposition 4.4 follows from the multidimensional Hales–Jewett theorem. In fact, applying Theorem 4.2 with alphabet $\mathbb{F}$ and $d = \max_{j\in[r]} \ell_j$ yields for sufficiently large n and any partition $E_1 \dot\cup \dots \dot\cup E_r$ of $\mathbb{F}^n = \begin{bmatrix}\mathbb{F}^n\\ 0\end{bmatrix}_{\mathrm{aff}}$ a d-dimensional subspace $\mathcal{S}'$ completely contained in some E_j. Since every combinatorial d-space in $\mathbb{F}^n$ is also a d-dimensional affine space in $\mathbb{F}^n$, any choice of $\mathcal{S} \in \begin{bmatrix}\mathcal{S}'\\ \ell_j\end{bmatrix}_{\mathrm{aff}}$ yields the conclusion in this case. Furthermore, Proposition 4.4 is trivial if $\ell_j = k$ for some $j \in [r]$ as every k-dimensional affine space only contains one k-dimensional affine subspace. This establishes the induction start.

Let integers $\ell_1, \dots, \ell_r > k > 0$ be given. Applying the induction assumptions we obtain integers $\ell'_1, \dots, \ell'_r$ defined for every $j \in [r]$ by

$$\ell'_j = N^{(k)}(\ell_1, \dots, \ell_{j-1}, \ell_j - 1, \ell_{j+1}, \dots, \ell_r)$$

and we set

$$d = N^{(k-1)}(\ell'_1, \dots, \ell'_r)\,.$$

Applying Proposition 4.3 for $\mathbb{F}$ with d, k, and r yields a constant m_0 and finally we set

$$N^{(k)}(\ell_1, \dots, \ell_r) = m_0\,. \tag{4.6}$$

Let $n \geq N^{(k)}(\ell_1, \dots, \ell_r)$ and let $E_1 \dot\cup \dots \dot\cup E_r$ be an arbitrary partition of $\begin{bmatrix}\mathbb{F}^n\\ k\end{bmatrix}_{\mathrm{aff}}$. Since $n \geq m_0$ there exists some $(d+1)$-dimensional affine subspace $\mathcal{U}$, which is special with respect to the partition $E_1 \dot\cup \dots \dot\cup E_r$. In particular, for the projection $\pi_d \colon \mathbb{F}^n \to \mathbb{F}^d$ onto the first d coordinates we have $\pi_d(\mathcal{U}) = \mathbb{F}^d$. We observe that for every integer $\ell \leq d$ and any $\mathcal{V}' \in \begin{bmatrix}\mathbb{F}^d\\ \ell\end{bmatrix}_{\mathrm{aff}}$ we have $(\pi_d^{-1}(\mathcal{V}') \cap \mathcal{U}) \in \begin{bmatrix}\mathcal{U}\\ \ell+1\end{bmatrix}_{\mathrm{aff}}$.

Consequently, we can define a partition $E'_1 \dot\cup \dots \dot\cup E'_r$ of $\begin{bmatrix}\mathbb{F}^d\\ k-1\end{bmatrix}_{\mathrm{aff}}$. We include $\mathcal{V}' \in \begin{bmatrix}\mathbb{F}^d\\ k-1\end{bmatrix}_{\mathrm{aff}}$ in the class E'_j if $(\pi_d^{-1}(\mathcal{V}') \cap \mathcal{U}) \in E_j$. The choice of d allows us to apply the induction assumption for $k-1$ to the partition $E'_1 \dot\cup \dots \dot\cup E'_r = \begin{bmatrix}\mathbb{F}^d\\ k-1\end{bmatrix}_{\mathrm{aff}}$ and as a result we obtain some $j' \in [r]$ and an affine subspace

$$\mathcal{S}' \in \begin{bmatrix}\mathbb{F}^d\\ \ell'_{j'}\end{bmatrix}_{\mathrm{aff}} \quad \text{such that} \quad \begin{bmatrix}\mathcal{S}'\\ k-1\end{bmatrix}_{\mathrm{aff}} \subseteq E'_{j'}$$

We set $\mathcal{S} = \pi_d^{-1}(\mathcal{S}') \cap \mathcal{U}$. Owing to $\mathcal{S} \in \begin{bmatrix}\mathcal{U}\\ \ell'_j+1\end{bmatrix}_{\mathrm{aff}}$, we have $\mathcal{S}$ is special with respect to the partition $E_1 \dot\cup \dots \dot\cup E_r$ of $\begin{bmatrix}\mathbb{F}^n\\ k\end{bmatrix}_{\mathrm{aff}}$. Moreover, it follows from the definition of the partition $E'_1 \dot\cup \dots \dot\cup E'_r = \begin{bmatrix}\mathbb{F}^d\\ k-1\end{bmatrix}_{\mathrm{aff}}$ and the properties of $\mathcal{S}'$ that every non-transversal subspace $\mathcal{W} \in \begin{bmatrix}\mathcal{S}\\ k\end{bmatrix}_{\mathrm{aff}}$ is an element of $E_{j'}$.

Next we consider a partition $E''_1 \dot\cup \dots \dot\cup E''_r$ of $\begin{bmatrix}\mathcal{S}'\\ k\end{bmatrix}_{\mathrm{aff}}$. For any transversal space $\mathcal{W} \in \begin{bmatrix}\mathcal{S}\\ k\end{bmatrix}_{\mathrm{aff}}$ we let $\pi_d(\mathcal{W}) \in E''_j$ if $\mathcal{W} \in E_j$. Note that this is well defined and the partition class of $\pi_d(W)$ is indeed independent of $\mathcal{W}$. This is because $\mathcal{S}$ is special with respect to $E_1\dot\cup\dots\dot\cup E_r$ and, therefore, all transversal spaces in $\genfrac{[}{]}{0pt}{}{\mathcal{S}}{k}_{\mathrm{aff}}$ with the same image under π_d are contained in the same partition class of $E_1\dot\cup\dots\dot\cup E_r$.

Owing to the choice of $\ell'_{j'}$ and since $\dim(\mathcal{S}') = \ell'_{j'}$ we can appeal to the induction assumption for $\ell_1,\dots,\ell_{j'}-1,\dots,\ell_r \geq k$. Consequently, one of the following two assertions must hold

(i) either there exists some $j'' \in [r]$ with $j'' \neq j'$ and some $\mathcal{S}'' \in \genfrac{[}{]}{0pt}{}{\mathcal{S}'}{\ell_{j''}}_{\mathrm{aff}}$ such that $\genfrac{[}{]}{0pt}{}{\mathcal{S}''}{k}_{\mathrm{aff}} \subseteq E''_{j''}$,

(ii) or there exists some $\mathcal{S}'' \in \genfrac{[}{]}{0pt}{}{\mathcal{S}'}{\ell_{j'}-1}_{\mathrm{aff}}$ such that $\genfrac{[}{]}{0pt}{}{\mathcal{S}''}{k}_{\mathrm{aff}} \subseteq E''_{j'}$.

We begin with case (i). Recall that $\dim(\mathcal{S}') = \ell'_{j'}$ and $\dim(\mathcal{S}) = \ell'_{j'}+1$. Hence, $\dim(\pi_d^{-1}(\mathcal{S}'')\cap\mathcal{S}) = \ell_{j''}+1$. Consequently, we can fix some $\ell_{j''}$-dimensional affine subspace $\mathcal{S}^*$ in $\pi_d^{-1}(\mathcal{S}'')\cap\mathcal{S}$ with the property $\pi_d(\mathcal{S}^*) = \mathcal{S}''$. Owing to this choice, every $\mathcal{W} \in \genfrac{[}{]}{0pt}{}{\mathcal{S}^*}{k}_{\mathrm{aff}}$ is transversal. Hence it follows from the definition of the partition $E''_1\dot\cup\dots\dot\cup E''_r$ and the property of $\mathcal{S}''$ that $\genfrac{[}{]}{0pt}{}{\mathcal{S}^*}{k}_{\mathrm{aff}} \subseteq E_{j''}$, which concludes the proof of Proposition 4.4 in this case.

Finally, we consider case (ii). In this case we set $\mathcal{S}^* = \pi_d^{-1}(\mathcal{S}'')\cap\mathcal{S}$ and it follows that $\dim(\mathcal{S}^*) = \dim(\mathcal{S}'')+1 = \ell_{j'}$. Similarly, as in case (i) it follows that every transversal $\mathcal{W} \in \genfrac{[}{]}{0pt}{}{\mathcal{S}^*}{k}_{\mathrm{aff}}$ is contained in $E_{j'}$. Since $\dim(\mathcal{S}^*) = \dim(\mathcal{S}'')+1$, there are non-transversal spaces $\mathcal{W} \in \genfrac{[}{]}{0pt}{}{\mathcal{S}^*}{k}_{\mathrm{aff}}$. However, every non-transversal space $\mathcal{W} \in \genfrac{[}{]}{0pt}{}{\mathcal{S}^*}{k}_{\mathrm{aff}} \subseteq \genfrac{[}{]}{0pt}{}{\mathcal{S}}{k}_{\mathrm{aff}}$ is also non-transversal in $\mathcal{S}$. Owing to the properties of $\mathcal{S}$ each such non-transversal $\mathcal{W}$ is contained in $E_{j'}$ and this concludes the proof in this case. □

We close this section be reducing Theorem 1.8 to the affine version.

Proof of Theorem 1.8. For $\mathbb{F}$, r, ℓ and k, we let n_0 be given by Theorem 1.9 applied with the same parameters. For $n \geq n_0$ let $E_1\dot\cup\dots\dot\cup E_r$ be a partition of the k-dimensional subspaces of $\mathbb{F}^n$. Note that for every affine subspace $\mathcal{V}$ in $\mathbb{F}^n$ there exists a unique vector $\vec{v}_{\mathcal{V}} \in \mathbb{F}^n$ such that the translate $\mathcal{V}-\vec{v}_{\mathcal{V}}$ is subspace of $\mathbb{F}^n$. We will use this to extend the given partition to a partition $E'_1\dot\cup\dots\dot\cup E'_r$ of $\genfrac{[}{]}{0pt}{}{\mathbb{F}^n}{k}_{\mathrm{aff}}$.

In fact, we include $\mathcal{W}$ in E'_j if $\mathcal{W}-\vec{v}_{\mathcal{W}} \in E_j$. Then Theorem 1.9 yields $\mathcal{S} \in \genfrac{[}{]}{0pt}{}{\mathbb{F}^n}{\ell}_{\mathrm{aff}}$ and $j \in [r]$ with $\genfrac{[}{]}{0pt}{}{\mathcal{S}}{k}_{\mathrm{aff}} \in E'_j$. Consequently, $\mathcal{W}-\vec{v}_{\mathcal{W}} \in E_j$ for every $\mathcal{W} \in \genfrac{[}{]}{0pt}{}{\mathcal{S}}{k}_{\mathrm{aff}} \in E'_j$ and, hence, $\mathcal{S}-\vec{v}_{\mathcal{S}}$ has the property that all its k-dimensional subspaces belong to E_j. □

4.2. Shelah's proof of the Hales–Jewett theorem

In this section we prove Theorem 1.11. We present the proof of Shelah from [80]. Currently this proof gives the best bound on n_0 in Theorem 1.11. In fact, Shelah proved the multidimensional version (Theorem 4.2) directly, which gives better bounds for Theorem 4.2. However, we only present Shelah's proof for the case of combinatorial lines. This case is a bit simpler and we closely follow the presentation from [50, Chapter 29] (see also [68]).

The following notation will be useful for the proof.

Definition 4.5 (line template). Let $A = \{a_1,\dots,a_k\}$ be a finite set and let $* \notin A$. For $n \geq 1$ we say $\Lambda \in (A\cup\{*\})^n$ is a *line template* in A^n if at least one coordinate of Λ equals $*$.

Moreover, for $a \in A$ we denote by $\Lambda(a)$ the element of A^n which we obtain from Λ be replacing every $*$ by a. We set

$$\mathcal{L}_A(\Lambda) = \{\Lambda(a)\colon\ a \in A\}\,,$$

i.e., $\mathcal{L}_A(\Lambda)$ is the combinatorial line in A corresponding to the template Λ.

For a sequence $\mathbf{\Lambda} = (\Lambda_1, \dots, \Lambda_m)$ of line templates with $\Lambda_\ell \in (A \cup \{*\})^{n_\ell}$ for $\ell \in [m]$ and some $\boldsymbol{a} = (a_1, \dots, a_m) \in A^m$ we denote by $\mathbf{\Lambda}(\boldsymbol{a})$ the element in $A^{\sum_{\ell=1}^m n_\ell}$ given by

$$\mathbf{\Lambda}(\boldsymbol{a}) = \Lambda_1(a_1) \circ \dots \circ \Lambda_m(a_m)\,,$$

where $\circ$ denotes the concatenation of vectors.

For fixed $r \geq 1$ the proof of Theorem 1.11 is by induction on the cardinality of the alphabet A. For $|A| = 1$ Theorem 1.11 is trivial. With out loss of generality we may assume $A = [k]$.

For the inductive step from an alphabet of size k to $A = [k+1]$ the following proposition will be crucial. We say two elements $\boldsymbol{a} = (a_1, \dots, a_m)$ and $\boldsymbol{b} = (b_1, \dots, b_m) \in [k+1]^m$ and are $(k-1)$-*indistinguishable* if for every $i \in [k-1]$ and $\ell \in [m]$ we have

$$a_\ell = i \qquad \Longleftrightarrow \qquad b_\ell = i\,.$$

In other words, $\boldsymbol{a}$ and $\boldsymbol{b}$ are $(k-1)$-indistinguishable if they only differ in those coordinates which take values k or $k+1$.

Proposition 4.6. *For all integers r, k, and $m \geq 1$ there exist positive integers $n_1, \dots, n_m \geq 1$ such that for $n = \sum_{i=\ell}^m n_\ell$ and every partition $E_1 \dot\cup \dots \dot\cup E_r = [k+1]^n$ there exists a sequence $\mathbf{\Lambda} = (\Lambda_1, \dots, \Lambda_m)$ of line templates with $\Lambda_\ell \in ([k+1] \cup \{*\})^{n_\ell}$ for $\ell \in [m]$ such that for all $(k-1)$-indistinguishable pairs $\boldsymbol{a}$, $\boldsymbol{b} \in [k+1]^m$ there exists some $j \in [r]$ such that $\mathbf{\Lambda}(\boldsymbol{a})$, $\mathbf{\Lambda}(\boldsymbol{b}) \in E_j$.*

Proof. Let r, k, and m be given. We inductively define the following sequence of integers $n_1, \dots, n_m$. Set

$$N_0 = 0 \quad \text{and} \quad n_1 = r^{(k+1)^{m-1}}$$

and for $\ell = 2, \dots, m$ set

$$N_{\ell-1} = \sum_{t=1}^{\ell-1} n_t \quad \text{and} \quad n_\ell = r^{(k+1)^{N_{\ell-1}+m-\ell}} \tag{4.7}$$

and set

$$n = N_m = \sum_{\ell=1}^{m} n_m\,.$$

Let $E_1 \dot\cup \dots \dot\cup E_m$ be an arbitrary partition of $[k+1]^n$. We obtain the required sequence of line templates inductively and for $s = 0, \dots, m$ we verify the following statement.

($\mathcal{S}_s$) For every integer $t = m-s+1, \dots, m$ there exists a line template $\Lambda_t \in ([k+1] \cup \{*\})^{n_t}$ such that $\mathbf{\Lambda}_s = (\Lambda_{m-s+1}, \dots, \Lambda_m)$ has the property that for every $(k-1)$-indistinguishable pair $\boldsymbol{a}_s = (a_{m-s+1}, \dots, a_m)$, $\boldsymbol{b}_s = (b_{m-s+1}, \dots, b_m) \in [k+1]^s$ and every $\boldsymbol{x} \in [k+1]^{N_{m-s}}$ there exists a $j \in [r]$ such that $\boldsymbol{x} \circ \mathbf{\Lambda}_s(\boldsymbol{a}_s)$ and $\boldsymbol{x} \circ \mathbf{\Lambda}_s(\boldsymbol{b}_s)$ are contained in E_j.

Clearly ($\mathcal{S}_m$) is the conclusion of Proposition 4.6 and ($\mathcal{S}_0$) is a trivial statement.

For the induction step we assume the validity of ($\mathcal{S}_{s-1}$) and let line templates $\mathbf{\Lambda}_{s-1} = (\Lambda_{m-s+2}, \dots, \Lambda_m)$ be given. For $t = 0, \dots, n_{m-s+1}$ we consider the following vectors of length n_{m-s+1} defined by

$$\boldsymbol{y}_t = (\underbrace{k-1, \dots, k-1}_{(n_{m-s+1}-t)\text{-times}}, \underbrace{k, \dots, k}_{t\text{-times}})\,.$$

For every $t = 0, \dots, n_{m-s+1}$ we consider a partition $E_1^t \dot\cup \dots \dot\cup E_r^t$ of $[k+1]^{N_{m-s}+s-1}$ defined for every $\boldsymbol{x} \circ \boldsymbol{z}$ with $\boldsymbol{x} \in [k+1]^{N_{m-s}}$ and $\boldsymbol{z} \in [k+1]^{s-1}$ by

$$\boldsymbol{x} \circ \boldsymbol{z} \in E_j^t \quad \Longleftrightarrow \quad \boldsymbol{x} \circ \boldsymbol{y}_t \circ \mathbf{\Lambda}_{s-1}(\boldsymbol{z}) \in E_j\,. \tag{4.8}$$

This way we defined $n_{m-s+1}+1$ partitions. On the other hand, there are at most

$$r^{|[k+1]^{N_{m-s+s-1}}|} \overset{(4.7)}{=} n_{m-s+1}\,.$$

Consequently, by the pigeonhole principle two of the defined partitions must be identical, i.e., there exist integers $0 \leq t < t' \leq n_{m-s+1}$ such that $E_j^t = E_j^{t'}$ for every $j \in [r]$. We define the line template $\Lambda_{m-s+1} \in ([k+1] \cup \{*\})^{n_{m-s+1}}$ as

$$\Lambda_{m-s+1} = (\underbrace{k-1,\dots,k-1}_{(n_{m-s+1}-t')\text{-times}},\ \underbrace{*,\dots,*}_{(t'-t)\text{-times}},\underbrace{k,\dots,k}_{t\text{-times}})$$

and show that $\mathbf{\Lambda}_s = (\Lambda_{m-s+1},\dots,\Lambda_m)$ has the desired property.

Let $\boldsymbol{a}_s = (a_{m-s+1},\dots,a_m)$ and $\boldsymbol{b}_s = (b_{m-s+1},\dots,b_m) \in [k+1]^s$ be a $(k-1)$-indistinguishable pair and $\boldsymbol{x} \in [k-1]^{N_{m-s}}$ be fixed.

We may assume $a_{m-s+1} \neq b_{m-s+1}$. In fact, if $a_{m-s+1} = b_{m-s+1}$, then $(\mathcal{S}_{s-1})$ applied to

$$\boldsymbol{x}' = \boldsymbol{x} \circ \Lambda_{m-s+1}(a_{m-s+1}) = \boldsymbol{x} \circ \Lambda_{m-s+1}(b_{m-s+1})$$

and the $(k-1)$-indistinguishable pair

$$\boldsymbol{a}'_{s-1} = (a_{m-s+2},\dots,a_m)\,,\ \boldsymbol{b}'_{s-1} = (b_{m-s+2},\dots,b_m) \in [k+1]^{s-1}$$

yields the claim.

So let $a_{m-s+1} \neq b_{m-s+1}$ and without loss of generality we assume $a_{m-s+1} = k$ and $b_{m-s+1} = k-1$. Observe that

$$\Lambda_{m-s+1}(k) = \boldsymbol{y}_{t'} \quad \text{and} \quad \Lambda_{m-s+1}(k-1) = \boldsymbol{y}_t\,.$$

Setting $\boldsymbol{c}_s = (k, b_{m-s+2},\dots,b_m)$ it follows from the induction assumption that the vectors

$$\boldsymbol{x} \circ \mathbf{\Lambda}_s(\boldsymbol{a}_s) = \boldsymbol{x} \circ \boldsymbol{y}_{t'} \circ \mathbf{\Lambda}_{s-1}(a_{m-s+2},\dots,a_m)$$

and

$$\boldsymbol{x} \circ \mathbf{\Lambda}_s(\boldsymbol{c}_s) = \boldsymbol{x} \circ \boldsymbol{y}_{t'} \circ \mathbf{\Lambda}_{s-1}(b_{m-s+2},\dots,b_m)$$

are contained in the same partition class, say E_j. In particular, it follows from (4.8) that $\boldsymbol{x} \circ (b_{m-s+2},\dots,b_m) \in E_j^{t'}$. Owing to the choice of t and t', we have $E_j^{t'} = E_j^t$ and, hence, (4.8) yields

$$\boldsymbol{x} \circ \mathbf{\Lambda}_s(\boldsymbol{b}_s) = \boldsymbol{x} \circ \boldsymbol{y}_t \circ \mathbf{\Lambda}_{s-1}(b_{m-s+2},\dots,b_m) \in E_j\,.$$

This establishes (S_s) and concludes the proof of Proposition 4.6. □

Below we prove Theorem 1.11 by induction on $|A|$ and the induction step will make use of Proposition 4.6.

Proof of Theorem 1.11. Let $r \geq 1$ be fixed. Since Theorem 1.11 is trivial for $|A| = 1$ we may assume Theorem 1.11 holds for alphabets of size k. Moreover, without loss of generality we may assume that $A = [k+1]$.

Let m_0 be given by the induction assumption, i.e., for every $m \geq m_0$ every partition of $[k]^m$ into r classes yields a combinatorial line completely contained in one of the classes. We appeal to Proposition 4.6 applied with r, k and m_0, which yields some integer n. We will show that this n is sufficiently large for Theorem 1.11 for r and $A = [k+1]$. It is easy to check that if the conclusion of Theorem 1.11 holds for A^n, then it also holds for $A^{n'}$ for every $n' \geq n$.

Let $E_1 \dot\cup \dots \dot\cup E_r$ be a partition of $[k+1]^n$. Applying Proposition 4.6 yields a sequence of line templates $\mathbf{\Lambda} = (\Lambda_1,\dots,\Lambda_{m_0})$ such that for all $(k-1)$-indistinguishable pairs $\boldsymbol{a}$, $\boldsymbol{b} \in [k+1]^{m_0}$ there exists some $j \in [r]$ such that $\mathbf{\Lambda}(\boldsymbol{a})$ and $\mathbf{\Lambda}(\boldsymbol{b})$ are both elements from E_j. We consider the following partition $E'_1 \dot\cup \dots \dot\cup E'_r$ of $[k]^{m_0}$ given by letting $\boldsymbol{a} \in [k]^{m_0}$ be contained in E'_j if $\mathbf{\Lambda}(\boldsymbol{a}) \in E_j$.

Owing to the choice of m_0 it follows from the induction assumption that there exists some $j \in [r]$ such that E'_j contains a combinatorial line $\mathcal{L}'$ in $[k]^{m_0}$ and let

$$\Lambda' = (\lambda'_1 \dots \lambda'_{m_0}) \in ([k] \cup \{*\})^{m_0}$$

be the template of this line, i.e.,

$$\mathcal{L}' = \mathcal{L}_{[k]}(\Lambda') = \{\Lambda'(i)\colon\ i \in [k]\} \subseteq E'_j\,.$$

We consider $\mathbf{\Lambda}(\Lambda') = \big(\Lambda_1(\lambda'_1), \dots, \Lambda_{m_0}(\lambda'_{m_0})\big)$, since Λ' is a line template, it contains at least one $*$ and, consequently, $\mathbf{\Lambda}(\Lambda')$ contains at least one $*$ and is a line template in $[k+1]^n$. We claim that $\mathcal{L}_{[k+1]}(\mathbf{\Lambda}(\Lambda'))$ is combinatorial line, which is contained in E_j.

In fact, it follows from the definition of the partition class $E'_j \subseteq [k]^{m_0}$ and from $\mathcal{L}_{[k]}(\Lambda') \subseteq E'_j$ that for every $i \in [k]$ we have

$$(\mathbf{\Lambda}(\Lambda'))(i) = \mathbf{\Lambda}(\Lambda'(i)) \in E_j\,.$$

Moreover, $\Lambda'(k)$ and $\Lambda'(k+1)$ are $(k-1)$-indistinguishable in $[k+1]^{m_0}$ and, therefore, $\mathbf{\Lambda}(\Lambda'(k))$ and $\mathbf{\Lambda}(\Lambda'(k+1))$ are contained in the same partition class of the partition $E_1 \dot\cup \dots \dot\cup E_r = [k+1]^n$. In particular, $\mathbf{\Lambda}(\Lambda'(k+1)) = (\mathbf{\Lambda}(\Lambda'))(k+1)$ is also contained in E_j, which shows that the combinatorial line

$$\mathcal{L}_{[k+1]}(\mathbf{\Lambda}(\Lambda')) = \{\mathbf{\Lambda}(\Lambda'(i))\colon\ i \in [k+1]\}$$

is contained in E_j. $\square$

4.3. Upper bounds for the Hales-Jewett numbers

We briefly introduce the following version of the Ackermann function. For every integer $x \in \mathbb{N}$ we set

$$\mathfrak{A}_1(x) = 2x$$

and for $i \geq 2$ we set

$$\mathfrak{A}_i(x) = \mathfrak{A}_{i-1}^x(1) = (\underbrace{\mathfrak{A}_{i-1} \circ \dots \circ \mathfrak{A}_{i-1}}_{x\text{-times}})(1)\,.$$

In other words, $\mathfrak{A}_i(x)$ is the x-times iterated function $\mathfrak{A}_{i-1}$ applied to 1. The diagonal function

$$\mathfrak{A}(x) = \mathfrak{A}_x(x)$$

is a variant of the *Ackermann function*. It follows from those definitions that

$$\mathfrak{A}_2(x) = 2^x \quad \text{and} \quad \mathfrak{A}_3(x) = \left.2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}\right\}\text{height } x$$

and $\mathfrak{A}_4$ is an iterated tower function and so on. The function $\mathfrak{A}$ grows extremely fast. For example, $\mathfrak{A}(4) = \mathfrak{A}_4(4)$ is given by a tower of 2's of height 2^{16}. In fact, $\mathfrak{A}$ is not primitive recursive, while $\mathfrak{A}_i$ is primitive recursive for every fixed i (see, e.g., [**3, 69, 76**] for details).

A careful analysis of the proofs of Proposition 4.6 and Theorem 1.11 shows that for $r = 2$ and an alphabet of size $k \geq 3$ the constant n_0 in Theorem 1.11 can be bounded by

$$\mathfrak{A}_4(k) < n_0 < \mathfrak{A}_4(k+1)$$

(see [**44**, Section 2.7] for details). All known proofs before the work of Shelah [**80**] gave much weaker bounds and were not primitive recursive as a function of k.

APPENDIX A

Topology

We briefly review a few definitions and results from general topology.

A.1. Topological and metric spaces

A.1.1. Topological spaces. We start with the basic definition of a topological space.

Definition A.1 (toplogical space). A *topological space* is a pair $(X, \mathscr{T})$ consisting of a set X and *topology* $\mathscr{T} \subseteq 2^X$, which satisfies the following conditions

(*i*) $\emptyset \in \mathscr{T}$ and $X \in \mathscr{T}$,
(*ii*) $\mathscr{T}$ is closed under arbitrary unions, and
(*iii*) $\mathscr{T}$ is closed under finite intersections.

The sets in $\mathscr{T}$ are called *open* sets and their complements are *closed* sets.

For a subset $Y \subseteq X$ the induced topology $\mathscr{T}_Y = \{Z \cap Y \colon Z \in \mathscr{T}\}$ is a topology and $(Y, \mathscr{T}_Y)$ is a *topological subspace*.

Whenever the topology is clear from the context, then we suppress $\mathscr{T}$ and simply identify X with the topological space. For an arbitrary set $Y \subseteq X$ we sometimes consider the "smallest closed" set $\mathrm{cl}(Y)$ which contains Y.

Definition A.2 (closure, dense). Let $(X, \mathscr{T})$ be a topological space and $Y \subseteq X$. The *closure of* Y in the topological space $(X, \mathscr{T})$ (denoted by $\mathrm{cl}(Y)$) is the minimal (under inclusion) closed set, which contains Y, i.e., $(X \setminus \mathrm{cl}(Y)) \in \mathscr{T}$ and if some set Z satisfies $Y \subseteq Z \subseteq X$ and $(X \setminus Z) \in \mathscr{T}$, then we have $\mathrm{cl}(Y) \subseteq Z$.

Moreover, we say a set $Y \subseteq X$ is *dense in* X, if $\mathrm{cl}(Y) = X$.

It is easy to check that $\mathrm{cl}(Y) = \bigcap Z$, where the intersection runs over all closed sets Z which contain Y. Note, that it follows from de Morgan's rule and property (*ii*) of a topology that $\bigcap Z$ is indeed closed.

Definition A.3 (basis). A family of open sets $\mathscr{B}$ in a topological space $(X, \mathscr{T})$ is a *basis* of the topological space if every open set in $O \in \mathscr{T}$ there exists a subfamily $\mathscr{B}_O \subseteq \mathscr{B}$ such that $O = \bigcup_{B \in \mathscr{B}_O} B$.

We say the basis $\mathscr{B}$ *generates* the topology and open sets $B \in \mathscr{B}$ are called *basic* open sets.

Example A.4. The following simple topologies can be defined for any set X

trivial/indiscrete topology: $\mathscr{T} = \{\emptyset, X\}$ has basis $\mathscr{B} = \{\emptyset, X\}$,
discrete topology: $\mathscr{T} = 2^X := \{Y \colon Y \subseteq X\}$ has basis $\mathscr{B} = \{\{x\} \colon x \in X\}$.

Definition A.5 (continuity, homeomorphism). Let $(X, \mathscr{T})$ and $(X', \mathscr{T}')$ be topological spaces. A map $T \colon X \to X'$ is *continuous* if for every open set $O' \in \mathscr{T}'$ we have

$$T^{-1}(O') = \{x \in X \colon f(x) \in O'\} \in \mathscr{T} \,.$$

A continuous map $T \colon X \to X'$ is a *homeomorphism* if it is a bijection and the inverse map T^{-1} is continuous as well, i.e., $f(O) \in \mathscr{T}'$ for every open set $O \in \mathscr{T}$.

The set of continuous maps $T \colon X \to X$ on a topological space forms a semigroup with composition and the set of homeomorphisms is a group.

A.1.2. Compactness and Tychonoff's theorem. A very important concept in mathematics is compactness.

Definition A.6 (compactness). A toplogical space $(X, \mathscr{T})$ is *compact* if for any collection of open sets $\mathscr{O} = \{O_\iota \in \mathscr{T} \colon \iota \in I\}$ for some index set I with $X = \bigcup_{\iota \in I} O_\iota$ there exists a finite set $J \subseteq I$ such that $X = \bigcup_{\iota \in J} O_\iota$.

The following well known observation yields an alternative definition of compactness through families of closed sets. Proposition A.8 follows directly from de Morgan's rule.

Definition A.7 (finite intersection property). We say a family of sets $\mathscr{C} = \{C_\iota \colon \iota \in I\}$ has the *finite intersection property* if $\bigcap_{\iota \in J} C_\iota \neq \emptyset$ for every finite subset $J \subseteq I$.

Proposition A.8. *A topological space is compact if and only if any family of closed sets $\mathscr{C} = \{C_\iota \colon \iota \in I\}$ which has the finite intersection property satisfies $\bigcap_{\iota \in I} C_\iota \neq \emptyset$.*

Proof. Let $(X, \mathscr{T})$ be a compact topological space and let $\mathscr{C} = \{C_\iota \colon \iota \in I\}$ be a collection of closed subsets which satisfies the finite intersection property. Suppose $\bigcap_{\iota \in I} C_\iota = \emptyset$. Owing to de Morgan's law we have $X = \bigcup_{\iota \in I} \overline{C}_\iota$. In other words, $\{\overline{C}_\iota \colon \iota \in I\}$ is a collection of open sets which covers X. The compactness of $(X, \mathscr{T})$ yields a finite subset $J \subseteq I$ such that $X = \bigcup_{\iota \in J} \overline{C}_\iota$. Consequently, another application of de Morgan's law implies $\emptyset = \bigcap_{\iota \in J} C_\iota$, which contradicts the assumption that $\mathscr{C}$ satisfies the finite intersection property.

The proof of the opposite implication is very similar and we omit the details here. $\square$

Definition A.9 (product space). Let I be an arbitrary (not necessarily countable) index set and for every $\iota \in I$ let $(X_\iota, \mathscr{T}_\iota)$ be a topological space with basis $\mathscr{B}_\iota$. The product space $X = \prod_{\iota \in I} X_\iota$ is endowed with the product topology $\mathscr{T}$, which is generated by the basis $\mathscr{B}$ consisting of sets of the form $\prod_{\iota \in I} O_\iota$ where for some finite set J the following holds

$$O_\iota \begin{cases} = X_\iota & \text{if } \iota \in I \setminus J \\ \in \mathscr{B}_\iota & \text{if } \iota \in J. \end{cases}$$

The following well known result of Tychonoff [88] asserts that products of compact topological spaces are compact (for a proof see, e.g., [17, 48]).

Theorem A.10 (Tychonoff 1930). *Let $I \neq \emptyset$ be an arbitrary (not necessarily countable) index set and for every $\iota \in I$ let $(X_\iota, \mathscr{T}_\iota)$ be a non-empty toplogical space.*

The product space $X = \prod_{\iota \in I} X_\iota$ (endowed with the product topology) is a compact topological space if and only if $(X_\iota, \mathscr{T}_\iota)$ is compact for every $\iota \in I$. $\square$

A.1.3. Metric spaces. Often we will impose additional assumption on a topological space X. For example, that is endowed with a metric.

Definition A.11 (metric space). Let X be a non-empty set. A function $\varrho \colon X \times X \to \mathbb{R}_{\geq 0}$ is a *metric on* X if for all x, y, $z \in X$ we have

(*i*) $\varrho(x, y) = 0$ if and only if $y = x$,
(*ii*) $\varrho(x, y) = \varrho(y, x)$, and
(*iii*) $\varrho(x, z) \leq \varrho(x, y) + \varrho(y, z)$.

A *metric space* is a pair (X, ϱ) consisting of a set and a metric on it.

Whenever the metric is clear from the context, then we suppress ϱ and simply identify X with the metric space. We recall that a metric ϱ on X induces a topology $\mathscr{T}_\varrho$ on X, where the basis of $\mathscr{T}_\varrho$ consists of all open balls

$$\mathcal{B}_\varepsilon(x) = \{y \in X \colon\ \varrho(x,y) < \varepsilon\}$$

with $x \in X$ and $\varepsilon \in \mathbb{R}_{>0}$. Moreover, we transfer definitions made for topological spaces to metric spaces, by imposing the conditions for the topology induced by the metric. For example, we say a metric space (X,ϱ) is compact if the topological space $(X,\mathscr{T}_\varrho)$ is compact. It is easy to check that for continuous maps between metric spaces, such a definition coincides with the well known ε-δ-criteria, i.e., a map $T\colon X \to X'$ between two metric spaces (X,ϱ) and (X',ϱ') is continuous if and only if for every $x \in X$ and every $\varepsilon > 0$ there exists a $\delta > 0$ such that for every $y \in \mathcal{B}_\delta(x)$ we have $T(y) \in \mathcal{B}_\varepsilon(T(x))$. In addition we define *uniform continuity* for maps between metric spaces.

Definition A.12 (uniformly continuous). Let (X,ϱ) and (X',ϱ') be metric spaces. A map $T\colon X \to X'$ is *uniformly continuous* if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $\varrho'(T(x),T(y)) < \varepsilon$ for all x, $y \in X$ with $\varrho(x,y) < \delta$.

We recall the Heine-Cantor theorem, which asserts that continuity and uniform continuity coincide on compact metric spaces.

Theorem A.13 (Heine-Cantor theorem). *Suppose (X,ϱ) and (X',ϱ') are compact metric spaces. If $f\colon X \to X'$ is continuous, then f is uniformly continuous.*

Proof. Suppose for a contradiction that f is not uniformly continuous. Hence, there exists some $\varepsilon > 0$ such that for every $n \in \mathbb{N}$ there exist some points x_n, $y_n \in X$ such that $\varrho(x_n,y_n) < 1/n$, but $\varrho'(f(x_n),f(y_n)) \geq \varepsilon$. Owing to the compactness of X there exist convergent subsequences $(x_{i_j})_{j\in\mathbb{N}}$ and $(y_{i_j})_{j\in\mathbb{N}}$, i.e., there exist x and $y \in X$ such that $\varrho(x_{i_j},x) \to 0$ and $\varrho(y_{i_j},y) \to 0$ as $j \to \infty$. Consequently,

$$\varrho(x,y) \leq \varrho(x_{i_j},x) + \varrho(x_{i_j},y_{i_j}) + \varrho(y_{i_j},y) \to 0$$

as $j \to \infty$, i.e., $x = y$. Moreover, the continuity of f yields that $\varrho'(f(x_{i_j}),f(x)) \to 0$ and $\varrho'(f(y_{i_j}),f(y)) \to 0$ for $j \to \infty$. But since $x = y$ this yields

$$\varrho'(f(x_{i_j}),f(y_{i_j})) \leq \varrho'(f(x_{i_j}),f(x)) + \varrho'(f(y_{i_j}),f(y)) \to 0$$

for $j \to \infty$. In particular, for sufficiently large j we have $\varrho'(f(x_{i_j}),f(y_{i_j})) < \varepsilon$, which contradicts the assumption $\varrho'(f(x_n),f(y_n)) \geq \varepsilon$ for every $n \in \mathbb{N}$. □

Example A.14. The following metric can be defined for any non-empty set X

discrete metric: Setting $\varrho(x,y) = 0$ if $x = y$ and $\varrho(x,y) = 1$ if $x \neq y$, defines the *discrete metric* on X. The topology induced by ϱ is in fact the discrete topology. Moreover, every map on such a space is uniformly continuous.

A.2. Topological dynamics

Topological dynamics concerns continuous maps $T\colon X \to X$ on a set X. Requiring that X is a topological space and that the continuous maps form a semigroup leads to the notion of a *dynamical system*.

A.2.1. Dynamical systems and recurrence. Recall that the set of continuous maps on a topological space X forms a semigroup (with identity) under composition. Any sub-semigroup of these continuous maps together with X is a dynamical system.

Definition A.15 (dynamical system). A *dynamical system* is a pair (X,Γ), where X is a non-empty topological space and Γ is a semigroup of continuous maps on X. If $\Gamma=\{T^n\colon n\in\mathbb{N}\}$ for a continuous map T, then we simply denote the dynamical system by (X,T).

Often we will impose additional assumptions on X and Γ. For example, that X is compact or that it is a metric space (or both) or that Γ is a subgroup of the set of homeomorphisms on X.

Dynamical systems of the form (X,T) might be viewed as a dynamical process in the following way: Suppose X is the initial state of some system of particles or elements and T describes the evolution of X over time, i.e., at time n the element $x\in X$ moved to $T^n(x)$. Basic questions concern the behaviour of $T^n(x)$ for $n\to\infty$. For example, is it true that $T^n(x)$ is periodic, i.e., does there exist an integer $m>0$ such that $T^m(x)=x$. In such a case $T^n(x)$ returns to x for infinitely many $n\in\mathbb{N}$. Weakening this notion by requiring that $T^n(x)$ only returns "arbitrarily close" to x leads to the concept of *recurrence*.

Definition A.16 (recurrent point). Let (X,Γ) be a dynamical system. A point $x\in X$ is *recurrent* for Γ is for every open set U (in the topology on X) with $x\in U$, there exists some $T\in\Gamma$ such that $T(x)\in U$.

The first recurrence result is attributed to Poincaré [70] and concerns recurrence in finite measure spaces under measure preserving maps. The following recurrence theorem for compact topological spaces goes back to Birkhoff [5].

Theorem A.17 (Birkhoff's recurrence theorem – 1927). *For every compact dynamical system (X,Γ) there exists a recurrent point $x\in X$.*

In the proof of Theorem A.17 we will make use of *minimal* subsystems.

Definition A.18 (minimal system). Let (X,Γ) be a dynamical system. We say (X,Γ) is *minimal*, if the only non-empty, closed set $Y\subseteq X$ which is invariant under Γ (i.e., $T(Y)\subseteq Y$ for every $T\in\Gamma$) is $Y=X$.

The following characterisation of minimal systems based on *orbits* is very useful.

Definition A.19 (orbit). Let (X,Γ) be a dynamical system and $x\in X$. The *orbit of x* is defined by $O_x=\{T(x)\colon T\in\Gamma\}$.

Proposition A.20. *A dynamical system (X,Γ) is minimal if, and only if, every orbit is dense in X, i.e., $\mathrm{cl}(O_x)=X$ for every $x\in X$.*

Proof. "$\Rightarrow$": By definition the set $\mathrm{cl}(O_x)$ is non-empty and closed. Moreover, since for every $y\in O_x$ there exists some $S\in\Gamma$ with $S(x)=y$ we have for every $T\in\Gamma$ that $T(y)=T(S(x))=(T\circ S)(x)\in O_x$. In other words, $T(O_x)\subseteq O_x$ for every $T\in\Gamma$. Hence, the continuity of every $T\in\Gamma$ (applied in $(*)$ below) yields

$$\mathrm{cl}(O_x)\supseteq\mathrm{cl}(T(O_x))=\bigcap_{\substack{C'\text{ closed}\\ C'\supseteq T(O_x)}}C'\overset{(*)}{\supseteq}\bigcap_{\substack{C\text{ closed}\\ C\supseteq O_x}}T(C)\supseteq T\Bigg(\bigcap_{\substack{C\text{ closed}\\ C\supseteq O_x}}C\Bigg)=T(\mathrm{cl}(O_x))\,.$$

Consequently, it follows from the minimality of (X,Γ) that $\mathrm{cl}(O_x)=X$.

"$\Leftarrow$": Let Y be non-empty, closed, and invariant and $y\in Y$. In particular, $Y\supseteq\mathrm{cl}(O_y)$ for every $y\in Y$ and, hence, $Y=X$. □

For minimal systems (X,Γ), when X is a compact metric space and Γ is a group of homeomorphisms, we have the following characterisation.

Proposition A.21. *Let (X, ϱ) be a compact metric space and let Γ be a group of homeomorphisms on X. The dynamical system (X, Γ) is minimal if, and only if, for every $\varepsilon > 0$ there exists a finite set $\Gamma' \subseteq \Gamma$ such that for every $x, y \in X$ there exists $T \in \Gamma'$ such that $\varrho(x, T(y)) < \varepsilon$.*

Proof. " $\Rightarrow$ ": Let $U \neq \emptyset$ be some open set in X and set $V_U = \bigcup_{T\in\Gamma} T^{-1}(U)$. The set V_U is open and invariant under Γ, i.e., $T(V_U) \subseteq V_U$ for every $T \in \Gamma$. Moreover, since Γ is a group of homeomorphisms, it follows that $X \setminus V_U$ is closed and invariant under Γ and, hence, the minimality of (X, Γ) implies $X \setminus V_U = \emptyset$ and $V_U = X$ for every non-empty, open set U. Owing to the compactness of X, we obtain that for every non-empty, open set $U \subseteq X$ there exists a finite subset $\Gamma' \subseteq \Gamma$ such that $X = V_U = \bigcup_{T\in\Gamma'} T^{-1}(U)$.

Let $\varepsilon > 0$ be given. It also follows from the compactness of X that there exists some finite subset $X' \subseteq X$ such that $\bigcup_{x'\in X'} \mathcal{B}_{\varepsilon/2}(x') = X$. Applying the observation above to $U_{x'} = \mathcal{B}_{\varepsilon/2}(x')$ for every $x' \in X'$ yields finite sets $\Gamma'_{x'} \subseteq \Gamma$ such that $\bigcup_{T\in\Gamma'_{x'}} T^{-1}(\mathcal{B}_{\varepsilon/2}(x')) = X$.

We claim that the finite set $\Gamma'' = \bigcup_{x'\in X'} \Gamma'_{x'}$ satisfies the conclusion of this implication of Proposition A.21. In fact, for every $x \in X$ there exists some $x' \in X'$ such that $x \in \mathcal{B}_{\varepsilon/2}(x')$ and for every $y \in X$ there exists some $T \in \Gamma'_{x'} \subseteq \Gamma''$ such that $y \in T^{-1}(\mathcal{B}_{\varepsilon/2}(x'))$ and, therefore, $\varrho(x, T(y)) \leq \varrho(x, x') + \varrho(x', T(y)) < \varepsilon$.

" $\Leftarrow$ ": This implication follows from Proposition A.20, since the assumption asserts that for every $y \in X$ the orbit O_y is dense in X. $\square$

The following result asserts that compact dynamical systems contain a minimal subsystem.

Proposition A.22. *Let (X, Γ) be a compact dynamical system. There exists a closed, non-empty subset $Z \subseteq X$ such that $T(Z) \subseteq Z$ for every $T \in \Gamma$ and (Z, Γ) is a minimal dynamical system, where Z is endowed with the subspace topology.*

The proof of Proposition A.22 presented here is based on a standard application of Zorn's lemma [93].

Theorem A.23 (Zorn's lemma). *Let $(P, \leq)$ be a partially ordered set. Suppose $(P, \leq)$ has the property, that for every totally ordered subset (a so-called* chain*) $C \subseteq P$, there exists an element $p \in P$ such that $c \leq p$ for every $c \in C$. Then there exists a maximal element in P, i.e., there exists a $p^* \in P$ such that there is no $p \in P$ with $p^* < p$.*

Moreover, for every element $p \in P$ there exists a maximal element $p^* \in P$ such that $p \leq p^*$.

Furthermore, reversing the partial order yields, that if for chain $C \subseteq P$, there exists an element $p \in P$ such that $p \leq c$ for every $c \in C$, then there exists a minimal element in P. $\square$

Remark A.24. We remark that Zorn considered the "maximum principle" involved in Theorem A.23 at the time, when he was working at Universität Hamburg (shortly before he immigrated to the United States). However, similar "maximum and minimum principles" were already considered before and, in particular, Zorn's lemma appeared already more than ten years earlier in the work of Kuratowski [54].

Proof of Proposition A.22. Let $\mathscr{Y} \subseteq 2^X$ be the family of subsets $Y \subseteq X$, which are non-empty, closed, and invariant under Γ, i.e., $T(Y) \subseteq Y$ for every $T \in \Gamma$. Clearly, $(\mathscr{Y}, \subseteq)$ is partially ordered set. Since, X is a topological space, it follows from de Morgan's rule and property *(ii)* in Definition A.1 that for every family $\mathscr{C} \subseteq \mathscr{Y}$ the set

$$Z_{\mathscr{C}} = \bigcap_{Y\in\mathscr{C}} Y$$

is closed. Moreover, since $T(\bigcap_{Y\in\mathscr{C}} Y) \subseteq \bigcap_{Y\in\mathscr{C}} T(Y)$ for any map T and any family of sets $\mathscr{C}$, we obtain from $\mathscr{C} \subseteq \mathscr{Y}$ that $T(Z_{\mathscr{C}}) \subseteq Z_{\mathscr{C}}$ for every $T \in \Gamma$. Furthermore, since $\mathscr{C}$ is a chain, it enjoys the finite intersection property and since X is compact the set $Z_{\mathscr{C}}$ is non-empty (see Proposition A.8).

Summarising the above, we verified that for every chain $\mathscr{C} \subseteq \mathscr{Y}$ the set $Z_{\mathscr{C}}$ is closed, non-empty, and invariant under Γ, i.e., $Z_{\mathscr{C}} \in \mathscr{Y}$. Therefore, $(\mathscr{Y}, \subseteq)$ satisfies the assumptions of Zorn's lemma. Theorem A.23 yields the existence of a minimal set $Z \in \mathscr{Y}$ and it follows that (Z, Γ) is a minimal dynamical system. □

Birkhoff's recurrence theorem follows directly from Propositions A.20 and A.22.

Proof of Theorem A.17. Since (X, Γ) is compact, Proposition A.22 yields a minimal subsytem (Z, Γ). Hence, by Proposition A.20 we have $Z = \mathrm{cl}(O_z)$ for every $z \in Z$. In particular, $z \in \mathrm{cl}(O_z)$. If $z \in O_z$, then there exists some $T \in \Gamma$ such that $T(z) = z$ and we are done. If, on the other hand, $z \in \mathrm{cl}(O_z) \setminus O_z$, then every open set U containing z intersects O_z, which concludes the proof of Theorem A.17. □

A.3. Ultrafilter

Definition A.25 (filter and ultrafilter). A family $\mathcal{F} \subseteq 2^X$ of subsets of a set X is a *filter on* X if

(i) $\emptyset \notin \mathcal{F}$ and $X \in \mathcal{F}$,
(ii) $\mathcal{F}$ is upwards closed, i.e., if $A \in \mathcal{F}$ and $A \subseteq B \subseteq X$, then $B \in \mathcal{F}$, and
(iii) $\mathcal{F}$ is closed under intersection, i.e., for all A, $B \in \mathcal{F}$ we have $(A \cap B) \in \mathcal{F}$.

A filter $\mathcal{F}$ on X is an *ultrafilter* if in addition

(iv) for every set $A \subseteq X$ either $A \in \mathcal{F}$ or $(X \setminus A) \in \mathcal{F}$.

It follows from properties (i) and (iii) of filters, that the "either-or" in property (iv) is in fact exclusive. Moreover, the next observation follows directly from the definition.

Proposition A.26. *Let $\mathcal{F}$ be an ultrafilter on X, let A be a member of $\mathcal{F}$, and let $A_1 \dot\cup \dots \dot\cup A_r = A$ be a partition of A. Then there exists a unique index $j \in [r]$ such that $A_j \in \mathcal{F}$.*

Proof. It follows from properties (i) and (iii) of filters, that there exists at most one such index $j \in [r]$.

Suppose $A_j \notin \mathcal{F}$ for every $j \in [r]$. Owing to property (iv) we have $(X \setminus A_j) \in \mathcal{F}$ for every $j \in [r]$. Consequently, property (iii) yields $\bigcap_{j=1}^r (X \setminus A_j) \in \mathcal{F}$. On the other hand, $\bigcap_{j=1}^r (X \setminus A_j) = X \setminus A$. Since $A \in \mathcal{F}$ by assumption, this contradicts that the "either-or" in property (iv) is exclusive. □

Example A.27.

trivial filter: For any set X is $\mathcal{F} = \{X\}$ the *trivial filter*.

principal filter: For any $\emptyset \neq Y \subseteq X$ is $\mathcal{F}_Y = \{A \in 2^X : Y \subseteq A\}$ the principal filter generated by Y.

co-finite/Fréchet filter: Let X be an infinite set. The *co-finite filter* (sometimes called *Fréchet filter*) consists of the complements of finite sets, i.e., $\mathcal{F}_{\mathrm{co}} = \{A \subseteq X : |X \setminus A| < \infty\}$.

principal/trivial/fixed ultrafilter: Let X be a set. It is easy to check that the only principal filters which are ultrafilters are generated by a one-element set, i.e., they are of the form $\mathcal{F}_x = \{A \in 2^X : x \in A\}$ for some $x \in X$. These ultrafilters are *principal* (also called *trivial* or *fixed*) ultrafilters. Ultrafilters which are not of this form are called *non-principal* or *non-trivial* or *free*.

maximal filter: A filter $\mathcal{F}$ is *maximal* if it is not properly contained in any other filter, i.e., there exists no filter $\mathcal{F}'$ with $\mathcal{F} \subsetneq \mathcal{F}'$. Clearly, every principal ultrafilter is a maximal filter and, in fact, we will see that a filter is maximal if and only if it is an ultrafilter (see Theorem A.28).

Below we give a few equivalent formulations for ultrafilters.

THEOREM A.28. *Let X be a set and let $\mathcal{F}$ be a filter on X. The following statements are equivalent:*

(*i*) *$\mathcal{F}$ is an ultrafilter;*
(*ii*) *if $A \cup B = X$, then either $A \in \mathcal{F}$ or $B \in \mathcal{F}$;*
(*iii*) *$\mathcal{F}$ is maximal.*

PROOF. (*i*) $\Rightarrow$ (*ii*): Let $A \cup B = X$. If $A \notin \mathcal{F}$, then $(X \setminus A) \in \mathcal{F}$ because of (*iv*) of Definition A.25. Since $(X \setminus A) \subseteq B$, it follows from property (*ii*) of filters that $B \in \mathcal{F}$.

(*ii*) $\Rightarrow$ (*iii*): Let $\mathcal{F}$ be a filter, which satisfies property (*ii*). Suppose $\mathcal{F} \subsetneq \mathcal{F}'$ for some filter $\mathcal{F}'$. Hence, there exists a set $A \subseteq X$ in $\mathcal{F}' \setminus \mathcal{F}$. Owing to (*ii*), we have $(X \setminus A) \in \mathcal{F}$ and, therefore, $(X \setminus A) \in \mathcal{F}'$. Consequently, A and $X \setminus A$ are both members of $\mathcal{F}'$ and property (*iii*) of filters yields $A \cap (X \setminus A) = \emptyset \in \mathcal{F}'$, which contradicts property (*i*) of filters.

(*iii*) $\Rightarrow$ (*i*): Let $\mathcal{F}$ be a maximal filter and suppose $\mathcal{F}$ is not an ultrafilter. Hence, there exists a set $Z \subseteq X$ such that neither Z nor $X \setminus Z$ is in $\mathcal{F}$. Note that

$$A \cap Z \neq \emptyset \quad \text{for every } A \in \mathcal{F}\,, \tag{A.1}$$

since otherwise we would have $A \subseteq (X \setminus Z)$, which would yield $(X \setminus Z) \in \mathcal{F}$ by property (*ii*). We set

$$\mathcal{F}' = \{A' \subseteq X \colon (A \cap Z) \subseteq A' \text{ for some } A \in \mathcal{F}\}\,.$$

It is easy to check that $\mathcal{F}'$ is a filter. In fact, owing to (A.1) we have $\emptyset \notin \mathcal{F}'$ and obviously, $X \in \mathcal{F}'$ and $\mathcal{F}'$ is upwards closed. Property (*iii*) of Definition A.25 follows since for all sets $A_1', A_2' \in \mathcal{F}'$ and for all sets A_1' and A_2' satisfying $(A_1 \cap Z) \subseteq A_1' \in \mathcal{F}'$ and $(A_2 \cap Z) \subseteq A_2' \in \mathcal{F}'$ we have $(A_1 \cap A_2) \in \mathcal{F}$ (owing to property (*iii*) of the filter $\mathcal{F}$) and $(A_1 \cap A_2 \cap Z) \subseteq (A_1' \cap A_2')$. Consequently, $(A_1' \cap A_2') \in \mathcal{F}'$ as well.

Moreover, $\mathcal{F} \subseteq \mathcal{F}'$ and $Z \in \mathcal{F}' \setminus \mathcal{F}$. This contradicts the assumption that $\mathcal{F}$ is a maximal filter. □

Principal ultrafilters can be characterised in several ways and it follows from this charecterisation, that non-principal ultrafilters only exist for infinite sets X.

PROPOSITION A.29. *Let X be a set and let $\mathcal{F}$ be an ultrafilter on X. The following statements are equivalent:*

(*i*) *$\mathcal{F}$ is a principal ultrafilter;*
(*ii*) $\bigcap_{A \in \mathcal{F}} A \neq \emptyset$*;*
(*iii*) *$\mathcal{F}$ contains a finite set.*

In particular it follows from (iii), that for every finite set X all ultrafilters are principal and for every infinite set X every non-principal ultrafilter contains the co-finite filter.

PROOF. (*i*) $\Rightarrow$ (*ii*): This implication is obvious.

(*ii*) $\Rightarrow$ (*iii*): Set $A_0 = \bigcap_{A \in \mathcal{F}} A$. Since $\mathcal{F}$ is an ultrafilter and since $(X \setminus A_0) \notin \mathcal{F}$, the set A_0 is a member of $\mathcal{F}$.

Suppose A_0 contains more than one element. Consequently, there exists a proper subset B, $\emptyset \neq B \subsetneq A_0$. Since $\mathcal{F}$ is an ultrafilter either B or $X \setminus B$ is a member of $\mathcal{F}$. Since neither B nor $X \setminus B$ contains A_0, this yields a contradiction to the definition of A_0. Therefore the set $A_0 \in \mathcal{F}$ is finite (in fact it only contains one element).

(iii) $\Rightarrow$ *(i)*: Let $A_0 \in \mathcal{F}$ be a set of smallest size. If $A_0 = \{x\}$, then every member $B \in \mathcal{F}$ must contain x and, therefore, $\mathcal{F} \subseteq \mathcal{F}_x = \{A \in 2^X \colon x \in A\}$. On the other hand, we recall that $\mathcal{F}$ is an ultrafilter and, hence, for every A with $x \in A \subseteq X$ we must have $A \in \mathcal{F}$. If $(X \setminus A) \in \mathcal{F}$, then $(X \setminus A) \cap A_0 = \emptyset \in \mathcal{F}$ due to property *(iii)* of filters, which contradicts property *(i)*. Therefore, $\mathcal{F} = \mathcal{F}_x$ and we are done.

Hence, we assume A_0 is finite, but contains more than just one element. Then A_0 contains a proper subset B, $\emptyset \neq B \subsetneq A_0$ and since $\mathcal{F}$ is an ultrafilter either B or $B \setminus X$ is in $\mathcal{F}$. Consequently, either $A_0 \cap B$ or $A_0 \cap (X \setminus B)$ is a member of $\mathcal{F}$. In either case, $\mathcal{F}$ contains a set of size smaller than A_0, which contradicts the choice of A_0. □

After we have shown that non-principal ultrafilters can only exist for infinite sets, we show that indeed for every infinite set there exists a non-principal ultrafilter. However, this result requires the axiom of choice in form of Zorn's lemma, Theorem A.23.

THEOREM A.30. *For every infinite set X there exists a non-principal ultrafilter.*

PROOF. Let $\mathscr{F} \subseteq 2^{2^X}$ be the set of all filters on X. Clearly, $(\mathscr{F}, \subseteq)$ is a partially ordered set under inclusion. It is easy to check, that if $\mathscr{C} \subseteq \mathscr{F}$ is a totally ordered subset of filters, then $\bigcup_{\mathcal{F} \in \mathscr{C}} \mathcal{F}$ is a filter as well. Consequently, for every $\mathcal{F}' \in \mathscr{C}$ we have $\mathcal{F}' \subseteq \bigcup_{\mathcal{F} \in \mathscr{C}} \mathcal{F} \in \mathscr{F}$. Hence, it follows from Zorn's lemma that every filter is contained in a maximal filter and Theorem A.28 yields that such a maximal filter is an ultrafilter.

In particular, since X is infinite there exists the co-finite filter

$$\mathcal{F}_{\mathrm{co}} = \{A \subseteq X \colon |X \setminus A| < \infty\}$$

on X and there exists a maximal filter $\mathcal{F}^*$ which contains $\mathcal{F}_{\mathrm{co}}$. Suppose A is a co-finite set (i.e., $|X \setminus A|$ is finite), then $A \in \mathcal{F}_{\mathrm{co}} \subseteq \mathcal{F}^*$. Since, $\mathcal{F}^*$ is a filter the finite set $X \setminus A$ is not contained in $\mathcal{F}^*$. Repeating this argument for every co-finite set A yields that $\mathcal{F}^*$ contains no finite set. In particular, $\mathcal{F}^*$ is not a principal ultrafilter (see Proposition A.29). □

APPENDIX B

Analysis

We review a few basic facts from analysis.

B.1. Subadditive functions

DEFINITION B.1 (subadditive). A function $f\colon \mathbb{N} \to \mathbb{R}$ is *subadditive* if for all integers n, $m \in \mathbb{N}$ we have $f(n+m) \le f(n) + f(m)$.

It follows from the definition of subadditivity that for all a and $m \in \mathbb{N}$ we have

$$f(am) \le af(m)\,.$$

The following result was used by Fekete [31].

PROPOSITION B.2 (Fekete's lemma – 1923). *Let $f\colon \mathbb{N} \to \mathbb{R}_{\ge 0}$ be a subadditive function. The limit $\alpha = \lim_{n\to\infty} f(n)/n$ exists and $\alpha \le f(n)/n$ for every $n \in \mathbb{N}$.*

PROOF. Note that $f(n)/n \ge 0$ for every $n \in \mathbb{N}$ and set $\alpha = \liminf_{n\to\infty} f(n)/n$. Let $\varepsilon > 0$ and fix some $m \in \mathbb{N}$ such that $f(m)/m < \alpha + \varepsilon/2$. Set $B = \max_{b\in[m-1]} f(b)$ and let $n \ge 2B/\varepsilon$.

If $n = am$ for some $a \in \mathbb{N}$ then

$$\frac{f(n)}{n} = \frac{f(am)}{am} \le \frac{af(m)}{am} < \alpha + \frac{\varepsilon}{2}$$

and, similarly, if $n = am + b$ for some $a \in \mathbb{N}$ and $b \in [m-1]$, then

$$\frac{f(n)}{n} = \frac{f(am+b)}{n} \le \frac{f(am)}{n} + \frac{f(b)}{n} \le \frac{af(m)}{am+b} + \frac{B}{n} \le \frac{af(m)}{am} + \frac{\varepsilon}{2} < \alpha + \frac{\varepsilon}{2} + \frac{\varepsilon}{2}\,.$$

Hence, $f(n)/n < \alpha + \varepsilon$ for every sufficiently large n, i.e., $\limsup_{n\to\infty} f(n)/n = \alpha$. Therefore,

$$\limsup_{n\to\infty} \frac{f(n)}{n} = \alpha = \liminf_{n\to\infty} \frac{f(n)}{n}$$

and $\lim_{n\to\infty} f(n)/n = \alpha$.

The second assertion easily follows by contradiction. Suppose $f(n)/n = \beta < \alpha$ for some $n \in \mathbb{N}$. Then for every $a \in \mathbb{N}$ we have

$$\frac{f(an)}{an} \le \frac{af(n)}{an} = \beta < \alpha\,,$$

which yields the contradiction $\liminf_{n\to\infty} f(n)/n \le \beta < \alpha$. $\square$

B.2. Discrete Fourier analysis

For a more concise notation we introduce the function $e_n(\cdot)$ from the cyclic group $\mathbb{Z}/n\mathbb{Z}$ to the complex numbers defined by

$$e_n(x) = \exp\Big(\frac{2\pi\boldsymbol{i}}{n}x\Big) = \cos\Big(\frac{x}{n}\Big) + \boldsymbol{i}\sin\Big(\frac{x}{n}\Big)\,.$$

Clearly, $e_n(xr) = e_n(x)^r$ for all x, $r \in \mathbb{Z}/n\mathbb{Z}$ and the complex conjugate of $e_n(x)$ is given by

$$\overline{e_n(x)} = e_n(-x)\,. \tag{B.1}$$

Moreover, since $e_n(x)$ is an n-th root of unity for any $x \in \mathbb{Z}/n\mathbb{Z}$ we obtain for $n \geq 2$ and $x \in \mathbb{Z}/n\mathbb{Z} \setminus \{0\}$

$$\sum_{r \in \mathbb{Z}/n\mathbb{Z}} e_n(xr) = \sum_{r \in \mathbb{Z}/n\mathbb{Z}} e_n(x)^r = \frac{1 - e_n(x)^n}{1 - e_n(x)} = 0\,.$$

Therefore, for every $x \in \mathbb{Z}/n\mathbb{Z}$ and $n \in \mathbb{N}$ we have

$$\sum_{r \in \mathbb{Z}/n\mathbb{Z}} e_n(xr) = \begin{cases} 0 & \text{if } x \neq 0\,, \\ n & \text{if } x = 0\,. \end{cases} \tag{B.2}$$

For a function $f\colon \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$, we define the *Fourier transform* $\hat{f}\colon \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$ to be

$$\hat{f}(r) = \sum_{x \in \mathbb{Z}/n\mathbb{Z}} f(x) e_n(-xr)\,. \tag{B.3}$$

The following estimate is a direct consequence of the definition of the Fourier transform. For every $r \in \mathbb{Z}/n\mathbb{Z}$

$$|\hat{f}(r)| = \left| \sum_{x \in \mathbb{Z}/n\mathbb{Z}} f(x) e_n(-xr) \right| \leq \sum_{x \in \mathbb{Z}/n\mathbb{Z}} |f(x)||e_n(-xr)| = \sum_{x \in \mathbb{Z}/n\mathbb{Z}} |f(x)|\,. \tag{B.4}$$

For any $x \in \mathbb{Z}/n\mathbb{Z}$ we have

$$\sum_{r \in \mathbb{Z}/n\mathbb{Z}} \hat{f}(r) e_n(xr) = \sum_{r \in \mathbb{Z}/n\mathbb{Z}} \sum_{y \in \mathbb{Z}/n\mathbb{Z}} f(y) e_n((x-y)r) \overset{\text{(B.2)}}{=} n f(x)\,,$$

which yields the inversion formula

$$f(x) = \frac{1}{n} \sum_{r \in \mathbb{Z}/n\mathbb{Z}} \hat{f}(r) e_n(xr)\,. \tag{B.5}$$

Applying the inversion formula to both functions f and $g\colon \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$ gives

$$\begin{aligned}
\sum_{x \in \mathbb{Z}/n\mathbb{Z}} f(x)\overline{g(x)} &\overset{\text{(B.5)}}{=} \sum_{x \in \mathbb{Z}/n\mathbb{Z}} \Bigg(\frac{1}{n} \sum_{r \in \mathbb{Z}/n\mathbb{Z}} \hat{f}(r) e_n(rx) \Bigg) \Bigg(\frac{1}{n} \sum_{s \in \mathbb{Z}/n\mathbb{Z}} \overline{\hat{g}(s)}\, \overline{e_n(sx)} \Bigg) \\
&\overset{\text{(B.1)}}{=} \frac{1}{n^2} \sum_{r \in \mathbb{Z}/n\mathbb{Z}} \sum_{s \in \mathbb{Z}/n\mathbb{Z}} \hat{f}(r)\overline{\hat{g}(s)} \sum_{x \in \mathbb{Z}/n\mathbb{Z}} e_n((r-s)x) \\
&\overset{\text{(B.2)}}{=} \frac{1}{n} \sum_{r \in \mathbb{Z}/n\mathbb{Z}} \hat{f}(r)\overline{\hat{g}(r)}\,.
\end{aligned}$$

For the special case $g = f$ this yields Parseval's identity

$$\sum_{r \in \mathbb{Z}/n\mathbb{Z}} |\hat{f}(r)|^2 = n \sum_{x \in \mathbb{Z}/n\mathbb{Z}} |f(x)|^2\,. \tag{B.6}$$

For two functions $f, g\colon \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$ we define the *convolution* $h = f * g$ by

$$h(x) = \sum_{y \in \mathbb{Z}/n\mathbb{Z}} f(y) g(x-y)$$

and we observe that

$$
\begin{aligned}
\hat{h}(r) &= \sum_{x\in\mathbb{Z}/n\mathbb{Z}} (f*g)(x)e_n(-xr) \\
&= \sum_{x\in\mathbb{Z}/n\mathbb{Z}} \sum_{y\in\mathbb{Z}/n\mathbb{Z}} f(y)g(x-y)e_n(-xr) \\
&= \sum_{y\in\mathbb{Z}/n\mathbb{Z}} f(y)e_n(-yr) \sum_{x\in\mathbb{Z}/n\mathbb{Z}} g(x-y)e_n(-(x-y)r) \qquad \text{(B.7)} \\
&= \sum_{y\in\mathbb{Z}/n\mathbb{Z}} f(y)e_n(-yr) \sum_{z\in\mathbb{Z}/n\mathbb{Z}} g(z)e_n(-zr) \\
&= \hat{f}(r)\hat{g}(r)\,.
\end{aligned}
$$

Notation

$\mathbb{N}$	$=$	$\{1, 2, 3, \dots\}$	–	the set of natural numbers
$\mathbb{N}_0$	$=$	$\{0, 1, 2, 3, \dots\}$	–	the set of natural numbers including zero
$\mathbb{R}$			–	the set of real numbers
$\mathbb{R}_{\geq 0}$	$=$	$\{x \in \mathbb{R}\colon\ x \geq 0\}$	–	the set of non-negative real numbers
$\mathbb{R}_{>0}$	$=$	$\{x \in \mathbb{R}\colon\ x > 0\}$	–	the set of positive real numbers
$[n]$	$=$	$\{1, 2, \dots, n\}$	–	first n positive integers ($n \in \mathbb{N}$)
$[a, b]$	$=$	$\{a, a+1, \dots, b-1, b\}$	–	integers between integers a and b (a, $b \in \mathbb{Z}$)
2^X	$=$	$\{A\colon\ A \subseteq X\}$	–	powerset of X
$\binom{X}{k}$	$=$	$\{A \in 2^X\colon\ \lvert A\rvert = k\}$	–	k-element subsets of X
$X \dot{\cup} Y$	$=$	$X \cup Y$	–	union of disjoint sets X and Y
$X + x$	$=$	$X \cup \{x\}$	–	joining one element x to a set X
$X - x$	$=$	$X \setminus \{x\}$	–	removing one element x from a set X
f^n	$=$	$f \circ f \circ \cdots \circ f$	–	n-times iterated function $f\colon\ X \to X$, where f^0 is the identity on X and if f is invertible, then f^{-n} is the n-times iterated inverse function f^{-1}

Bibliography

1. H. L. Abbott, *Lower bounds for some Ramsey numbers*, Discrete Math. **2** (1972), no. 4, 289–293. 2.4.2
2. F. G. Abramson and L. A. Harrington, *Models without indiscernibles*, J. Symbolic Logic **43** (1978), no. 3, 572–600. 1.4
3. W. Ackermann, *Zum Hilbertschen Aufbau der reellen Zahlen*, Math. Ann. **99** (1928), no. 1, 118–133. 4.3
4. N. Alon and J. H. Spencer, *The probabilistic method*, third ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., Hoboken, NJ, 2008, With an appendix on the life and work of Paul Erdős. 2.4.2
5. G. D. Birkhoff, *Dynamical systems*, Colloquium Publications, vol. 9, American Mathematical Society, 1927. A.2.1
6. C. C. Chang and H. J. Keisler, *Model theory*, North-Holland Publishing Co., Amsterdam, 1973, Studies in Logic and the Foundations of Mathematics, Vol. 73. 2.5
7. D. Conlon, *A new upper bound for diagonal Ramsey numbers*, Ann. of Math. (2) **170** (2009), no. 2, 941–960. 1.4, 2.4.2
8. D. Conlon, J. Fox, and B. Sudakov, *Hypergraph Ramsey numbers*, J. Amer. Math. Soc. **23** (2010), no. 1, 247–266. 1.4
9. N. G. de Bruijn and P. Erdős, *A colour problem for infinite graphs and a problem in the theory of relations*, Nederl. Akad. Wetensch. Proc. Ser. A. **54** = Indagationes Math. **13** (1951), 369–373. 1.3
10. W. Deuber, *Partitionstheoreme für Graphen*, Comment. Math. Helv. **50** (1975), no. 3, 311–320. 1.4
11. W. Deuber, K. Leeb, and B. Voigt, *Partitionseigenschaft von endlichen Folgen*, Abh. Math. Sem. Univ. Hamburg **53** (1983), 103–110. 1.4
12. W. A. Deuber, *Canonization*, Combinatorics, Paul Erdős is eighty, Vol. 1, Bolyai Soc. Math. Stud., János Bolyai Math. Soc., Budapest, 1993, pp. 107–123. 1.4
13. L. E. Dickson, *Lower limit for the number of sets of solutions of $x^e + y^e + z^e \equiv 0 \ (mod\, p)$*, J. Reine Angew. Math. **135** (1909), 181–188. 1.1.2
14. ———, *On the congruence $x^n + y^n + z^n \equiv 0 \ (mod\, p)$*, J. Reine Angew. Math. **135** (1909), 134–141. 1.1.2
15. R. Diestel, *Graph theory*, fourth ed., Graduate Texts in Mathematics, vol. 173, Springer, Heidelberg, 2010. 1.3, 2.5
16. B. Dushnik and E. W. Miller, *Partially ordered sets*, Amer. J. Math. **63** (1941), 600–610. 1.4
17. R. Engelking, *General topology*, second ed., Sigma Series in Pure Mathematics, vol. 6, Heldermann Verlag, Berlin, 1989, Translated from the Polish by the author. A.1.2
18. P. Erdős and P. Turán, *On some sequences of integers*, J. London. Math. Soc. **11** (1936), 261–264. 1.1.2
19. P. Erdős, *Some remarks on the theory of graphs*, Bull. Amer. Math. Soc. **53** (1947), 292–294. 1.4, 2.4.2
20. ———, *On the combinatorial problems which I would most like to see solved*, Combinatorica **1** (1981), no. 1, 25–42. 1.4, 2.4.2
21. P. Erdős, R. L. Graham, P. Montgomery, B. L. Rothschild, J. Spencer, and E. G. Straus, *Euclidean Ramsey theorems. I*, J. Combinatorial Theory Ser. A **14** (1973), 341–363. 1.4
22. ———, *Euclidean Ramsey theorems. II*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. I, North-Holland, Amsterdam, 1975, pp. 529–557. Colloq. Math. Soc. János Bolyai, Vol. 10. 1.4
23. ———, *Euclidean Ramsey theorems. III*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. I, North-Holland, Amsterdam, 1975, pp. 559–583. Colloq. Math. Soc. János Bolyai, Vol. 10. 1.4
24. P. Erdős, A. Hajnal, A. Máté, and R. Rado, *Combinatorial set theory: partition relations for cardinals*, Studies in Logic and the Foundations of Mathematics, vol. 106, North-Holland Publishing Co., Amsterdam, 1984. 1.4, 2.4.2, 2.4.2, 2.5

25. P. Erdős, A. Hajnal, and R. Rado, *Partition relations for cardinal numbers*, Acta Math. Acad. Sci. Hungar. **16** (1965), 93–196. 1.4, 2.4.2, 2.4.2
26. P. Erdős and L. Lovász, *Problems and results on 3-chromatic hypergraphs and some related questions*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II, North-Holland, Amsterdam, 1975, pp. 609–627. Colloq. Math. Soc. János Bolyai, Vol. 10. 2.4.2
27. P. Erdős and R. Rado, *A combinatorial theorem*, J. London Math. Soc. **25** (1950), 249–255. 1.4
28. ———, *Combinatorial theorems on classifications of subsets of a given set*, Proc. London Math. Soc. (3) **2** (1952), 417–439. 1.4, 2.2.2.2, 2.3
29. ———, *A partition calculus in set theory*, Bull. Amer. Math. Soc. **62** (1956), 427–489. 1.4
30. P. Erdős and G. Szekeres, *A combinatorial problem in geometry*, Compositio Math. **2** (1935), 463–470. 1.1.1, 1.4, 2.2.2.1, 2.3
31. M. Fekete, *Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten*, Math. Z. **17** (1923), no. 1, 228–249. B.1
32. P. Frankl, *A constructive lower bound for some Ramsey numbers*, Ars Combinatoria **3** (1977), 297–302. 2.4.2
33. P. Frankl and V. Rödl, *A partition property of simplices in Euclidean space*, J. Amer. Math. Soc. **3** (1990), no. 1, 1–7. 1.4
34. P. Frankl and R. M. Wilson, *Intersection theorems with geometric consequences*, Combinatorica **1** (1981), no. 4, 357–368. 2.4.2
35. C. Frasnay, *Sur des fonctions d'entiers se rapportant au théorème de Ramsey.*, C. R. Acad. Sci. Paris **256** (1963), 2507–2510. 2.4.2
36. H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256. 3.2
37. ———, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, Princeton, N.J., 1981, M. B. Porter Lectures. 3.2, 3.2.2.1
38. H. Furstenberg and B. Weiss, *Topological dynamics and combinatorial number theory*, J. Analyse Math. **34** (1978), 61–85 (1979). 3.2, 3.2.2.1
39. W. H. Gottschalk, *Choice functions and Tychonoff's theorem*, Proc. Amer. Math. Soc. **2** (1951), 172. 1.3, 1.3
40. R. L. Graham, *Recent trends in Euclidean Ramsey theory*, Discrete Math. **136** (1994), no. 1-3, 119–127, Trends in discrete mathematics. 1.4
41. R. L. Graham, K. Leeb, and B. L. Rothschild, *Ramsey's theorem for a class of categories*, Advances in Math. **8** (1972), 417–433. 1.1.3, 4.1.3
42. R. L. Graham and V. Rödl, *Numbers in Ramsey theory*, Surveys in combinatorics 1987 (New Cross, 1987), London Math. Soc. Lecture Note Ser., vol. 123, Cambridge Univ. Press, Cambridge, 1987, pp. 111–153. 1.4, 2.4.2
43. R. L. Graham and B. L. Rothschild, *A short proof of van der Waerden's theorem on arithmetic progressions*, Proc. Amer. Math. Soc. **42** (1974), 385–386. 3.1
44. R. L. Graham, B. L. Rothschild, and J. H. Spencer, *Ramsey theory*, second ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., New York, 1990, A Wiley-Interscience Publication. (document), 1.1.1, 1.4, 2.4.2, 4.3
45. A. W. Hales and R. I. Jewett, *Regularity and positional games*, Trans. Amer. Math. Soc. **106** (1963), 222–229. 1.1.3
46. D. Hilbert, *Ueber die Irreducibilität ganzer rationaler Functionen mit ganzzahligen Coefficienten*, J. Reine Angew. Math. **110** (1892), 104–129. 1.1.2
47. N. Hindman, *Finite sums from sequences within cells of a partition of N*, J. Combinatorial Theory Ser. A **17** (1974), 1–11. 2.5
48. K. Jänich, *Topology*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1984, With a chapter by Theodor Bröcker, Translated from the German by Silvio Levy. A.1.2
49. T. J. Jech, *The axiom of choice*, North-Holland Publishing Co., Amsterdam, 1973, Studies in Logic and the Foundations of Mathematics, Vol. 75. 2.5
50. St. Jukna, *Extremal combinatorics*, Texts in Theoretical Computer Science. An EATCS Series, Springer-Verlag, Berlin, 2001, With applications in computer science. 4, 4.2
51. A. S. Kechris, V. G. Pestov, and S. Todorcevic, *Fraïssé limits, Ramsey theory, and topological dynamics of automorphism groups*, Geom. Funct. Anal. **15** (2005), no. 1, 106–189. 1.4
52. D. Kőnig, *Über eine Schlussweise aus dem Endlichen ins Unendliche*, Acta Sci. Math. (Szeged) **3** (1927), no. 2-3, 121–130 (German). 1.3
53. I. Kříž, *Permutation groups in Euclidean Ramsey theory*, Proc. Amer. Math. Soc. **112** (1991), no. 3, 899–907. 1.4
54. C. Kuratowski, *Une méthode d'élimination des nombres transfinis des raisonnements mathématiques*, Fund. Math. **3** (1922), 76–108 (French). A.24

55. K. Leeb, *Vorlesungen über Pascaltheorie*, Lecture Notes, Universität Erlangen, 1973. 1.4
56. ______, *A full Ramsey-theorem for the Deuber-category*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II, North-Holland, Amsterdam, 1975, pp. 1043–1049. Colloq. Math. Soc. Janós Bolyai, Vol. 10. 1.4
57. Zs. Nagy, *A certain constructive estimate of the Ramsey number*, Mat. Lapok **23** (1972), 301–302 (1974). 2.4.2
58. J. Nešetřil, *For graphs there are only four types of hereditary Ramsey classes*, J. Combin. Theory Ser. B **46** (1989), no. 2, 127–132. 1.4
59. ______, *Ramsey classes and homogeneous structures*, Combin. Probab. Comput. **14** (2005), no. 1-2, 171–189. 1.4
60. ______, *Ramsey classes of topological and metric spaces*, Ann. Pure Appl. Logic **143** (2006), no. 1-3, 147–154. 1.4
61. ______, *Metric spaces are Ramsey*, European J. Combin. **28** (2007), no. 1, 457–468. 1.4
62. J. Nešetřil and V. Rödl, *Type theory of partition properties of graphs*, Recent advances in graph theory (Proc. Second Czechoslovak Sympos., Prague, 1974), Academia, Prague, 1975, pp. 405–412. 1.4
63. ______, *Partitions of finite relational and set systems*, J. Combinatorial Theory Ser. A **22** (1977), no. 3, 289–312. 1.4
64. ______, *A structural generalization of the Ramsey theorem*, Bull. Amer. Math. Soc. **83** (1977), no. 1, 127–128. 1.4
65. ______, *Ramsey classes of set systems*, J. Combin. Theory Ser. A **34** (1983), no. 2, 183–201. 1.4
66. ______, *Combinatorial partitions of finite posets and lattices—Ramsey lattices*, Algebra Universalis **19** (1984), no. 1, 106–119. 1.4
67. ______, *Partite construction and Ramsey space systems*, Mathematics of Ramsey theory, Algorithms Combin., vol. 5, Springer, Berlin, 1990, pp. 98–112. 1.4
68. A. Nilli, *Shelah's proof of the Hales-Jewett theorem*, Mathematics of Ramsey theory, Algorithms Combin., vol. 5, Springer, Berlin, 1990, pp. 150–151. 4, 4.2
69. R. Péter, *Konstruktion nichtrekursiver Funktionen*, Math. Ann. **111** (1935), no. 1, 42–60. 4.3
70. H. Poincaré, *Sur le probléme des trois corps et les équations de la dynamique*, Acta Math. **13** (1890), no. 1, 1–270. A.2.1
71. R. Rado, *Studien zur Kombinatorik*, Math. Z. **36** (1933), no. 1, 424–470. 1.1.2
72. ______, *Note on combinatorial analysis*, Proc. Lond. Math. Soc. (2) **48** (1943), 122–160. 1.1.3
73. ______, *Axiomatic treatment of rank in infinite sets*, Canadian J. Math. **1** (1949), 337–343. 1.3
74. S. P. Radziszowski, *Small Ramsey numbers*, Electron. J. Combin. **1** (1994-2009), Dynamic Survey 1, 72 pp. (electronic). 1.4
75. F. P. Ramsey, *On a problem in formal logic*, Proc. Lond. Math. Soc. (2) **30** (1930), 264–286. 1, 2.1, 2.5
76. R. M. Robinson, *Recursion and double recursion*, Bull. Amer. Math. Soc. **54** (1948), 987–993. 4.3
77. K. F. Roth, *Sur quelques ensembles d'entiers*, C. R. Acad. Sci. Paris **234** (1952), 388–390. 1.1.2, 3, 3.3, 3.3
78. ______, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109. 1.1.2, 3, 3.3, 3.3
79. I. Schur, *Über die Kongruenz $x^m + y^m \equiv z^m (mod\ p)$*, Jahresber. Deutsch. Math.-Verein. **25** (1916), 114–117. 1.1.2
80. S. Shelah, *Primitive recursive bounds for van der Waerden numbers*, J. Amer. Math. Soc. **1** (1988), no. 3, 683–697. 4, 4.2, 4.3
81. Th. Skolem, *Ein kombinatorischer Satz mit Anwendung auf ein logisches Entscheidungsproblem*, Fundam. Math. **20** (1933), 254–261 (German). 1.1.1, 1.4
82. J. Spencer, *Ramsey's theorem—a new lower bound*, J. Combinatorial Theory Ser. A **18** (1975), 108–115. 1.4, 2.4.2
83. J. H. Spencer, *Ramsey's theorem for spaces*, Trans. Amer. Math. Soc. **249** (1979), no. 2, 363–371. 1.1.3, 4.1.3
84. E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104. 3.3, 3.3
85. ______, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245, Collection of articles in memory of Juriĭ Vladimirovič Linnik. 1.1.2
86. T. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006. 1.1.2
87. A. Thomason, *An upper bound for some Ramsey numbers*, J. Graph Theory **12** (1988), no. 4, 509–517. 2.4.2

88. A. Tychonoff, *Über die topologische Erweiterung von Räumen*, Math. Ann. **102** (1930), no. 1, 544–561. 1.3, A.1.2
89. B. L. van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw Archief **15** (1927), 212–216. 1.1.2, 3.1
90. ———, *Einfall und Überlegung in der Mathematik. III*, Elemente der Math. **9** (1954), 49–56. 3.1
91. ———, *How the proof of Baudet's conjecture was found*, Studies in Pure Mathematics (Presented to Richard Rado), Academic Press, London, 1971, pp. 251–260. 3.1
92. E. Witt, *Ein kombinatorischer Satz der Elementargeometrie*, Math. Nachr. **6** (1952), 261–262. 1.1.3
93. M. Zorn, *A remark on method in transfinite algebra*, Bull. Amer. Math. Soc. **41** (1935), no. 10, 667–670. A.2.1