Ramsey Theory for Product Spaces

Pandelis Dodos

Vassilis Kanellopoulos

Preface

I. Ramsey theory is the general area of combinatorics devoted to the study of the pigeonhole principles that appear in mathematical practice. It originates from the works of Ramsey [**Ra**] and van der Waerden [**vdW**], and at the early stages of its development the focus was on structural properties of graphs and hypergraphs. However, the last 40 years or so, Ramsey theory has expanded significantly, both in scope and in depth, and is now constantly interacting with analysis, ergodic theory, logic, number theory, probability theory, theoretical computer science, and topological dynamics.

This book (which inherits, to some extent, the diversity of the field) is a detailed exposition of a number of Ramsey-type results concerning *product spaces* or, more accurately, finite Cartesian products $F_1 \times \cdots \times F_n$ where the factors $F_1, \dots, F_n$ may be equipped with an additional structure depending upon the context. Product spaces are ubiquitous in mathematics and are admittedly elementary objects, yet they exhibit a variety of Ramsey properties which depend on the *dimension* n and the *size* of each factor. Quantifying properly this dependence is one of the main goals of Ramsey theory, a goal which can sometimes be quite challenging.

I.1. The first example of a product space of interest to us in this book is the *discrete hypercube*

$$A^n \coloneqq \underbrace{A \times \cdots \times A}_{n-\text{times}}$$

where n is a positive integer and A is a nonempty finite set. In fact, we will be mostly interested in the *high-dimensional* case (that is, when the dimension n is large compared with the cardinality of A), but apart from this assumption no further constraints will be imposed on the set A.

A classical result concerning the structure of high-dimensional hypercubes was discovered in 1963 by Hales and Jewett [**HJ**]. It asserts that for every partition of A^n into, say, two pieces, one can always find a "sub-cube" of A^n which is entirely contained in one of the pieces of the partition. The Hales–Jewett theorem paved the way for a thorough study of the Ramsey properties of discrete hypercubes and related structures, and it triggered the development of several infinite-dimensional extensions. This material is the content of Chapters 2, 4 and 5.

Around 30 years after the work of Hales and Jewett, another fundamental result of Ramsey theory was proved by Furstenberg and Katznelson [**FK4**]. It is a natural, yet quite deep, refinement of the Hales–Jewett theorem and asserts that every *dense* subset of A^n (that is, every subset of A^n whose cardinality is proportional to that of A^n) must contain a "sub-cube" of A^n. Much more recently, the work of Furstenberg and Katznelson was revisited by several authors and a number of different proofs of this important result have been found. This line of research eventually led to a better understanding of the structure of dense subsets of hypercubes both in the finite and the infinite dimensional setting. We present these developments in Chapters 8 and 9.

I.2. A second example relevant to the theme of this book is the product space

$$T_1 \times \cdots \times T_d$$

where d is a positive integer and $T_1, \dots, T_d$ are nonempty trees. Partitions of products spaces of this form appear in the context of Ramsey theory for trees. However, in this case we are interested in the somewhat different regime where the dimension d is regarded as being fixed while the trees $T_1, \dots, T_d$ are assumed to be sufficiently large and even possibly infinite. Chapter 3 is devoted to this topic.

I.3. The last main example of a product space which we are considering in this book is of the form

$$\Omega_1 \times \cdots \times \Omega_n$$

where n is a positive integer and for each $i \in \{1, \dots, n\}$ the set Ω_i is the sample space of a probability space $(\Omega_i, \Sigma_i, \mu_i)$. We view, in this case, the set $\Omega_1 \times \cdots \times \Omega_n$ also as a probability space equipped with the product measure $\mu_1 \times \cdots \times \mu_n$.

A powerful result concerning products of probability spaces, with several consequences in Ramsey theory, was proved around 10 years ago. It asserts that for every finite family $\mathcal{F}$ of measurable events of $\Omega_1 \times \cdots \times \Omega_n$ whose joint probability is negligible, one can approximate the members of $\mathcal{F}$ by lower-complexity events (that is, by events which depend on fewer coordinates) whose intersection is empty. This result is known as the *removal lemma* and in this generality is due to Tao [**Tao1**], though closely related discrete analogues were obtained earlier by Gowers [**Go5**] and, independently, by Nagle, Rödl, Schacht and Skokan [**NRS, RSk**]. We present these results in Chapter 7.

Finally, in Chapter 6 we discuss certain aspects of the *regularity method*. It originates from the work of Szemerédi [**Sz1, Sz2**] and is used to show that dense subsets of discrete structures are inherently pseudorandom. We follow a probabilistic approach in the presentation of the method, emphasizing its relevance not only in the context of graphs and hypergraphs, but also in the analysis of high-dimensional product spaces.

II. This book is addressed to researchers in combinatorics, but also working mathematicians and advanced graduate students who are interested in this part of Ramsey theory. The prerequisites for reading this book are rather minimal; it only requires familiarity, at the graduate level, with probability theory and real analysis. Some familiarity with the basics of Ramsey theory (as exposed, for instance, in the book of Graham, Rothschild and Spencer [**GRS**]) would also be beneficial, though it is not necessary.

To assist the reader we have included six appendices, thus making this book essentially self-contained. In Appendix A we briefly discuss some properties of primitive recursive functions, while in Appendix B we present a classical estimate for the Ramsey numbers due to Erdős and Rado [**ER**]. In Appendix C we recall some results related to the Baire property which are needed in Section 3.2. Appendix D contains an exposition of a part of the theory of ultrafilters and idempotents in compact semigroups; we note that this material is used only in Section 4.1. Finally, in Appendix E we present the necessary background from probability theory, and in Appendix F we discuss open problems.

It is needless to say that this book is based on the work of many researchers who made Ramsey theory a rich and multifaceted area. Several new results are also included. Bibliographical information on the content of each chapter is contained in its final section named as "Notes and remarks".

Acknowledgments. During the preparation of this book we have been greatly helped from the comments and remarks of Thodoris Karageorgos and Kostas Tyros. We extend our warm thanks to both of them.

Athens — Pandelis Dodos
January 2015 — Vassilis Kanellopoulos

Contents

CHAPTER 1

Basic concepts

1.1. General notation

1.1.1. Throughout this book, by $\mathbb{N} = \{0, 1, \dots\}$ we shall denote the set of all natural numbers. Moreover, for every positive integer n we set $[n] \coloneqq \{1, \dots, n\}$.

For every set X by $|X|$ we shall denote its cardinality. If $k \in \mathbb{N}$ with $k \leqslant |X|$, then by $\binom{X}{k}$ we shall denote the set of all subsets of X of cardinality k, that is,

$$\binom{X}{k} = \{Y \subseteq X : |Y| = k\}. \tag{1.1}$$

On the other hand, if X is infinite, then $[X]^\infty$ stands for the set of all infinite subsets of X. The powerset of X will be denoted by $\mathcal{P}(X)$.

1.1.2. If X and Y are nonempty sets, then a map $c\colon X \to Y$ will be called a *Y-coloring of X*, or simply a *coloring* if X and Y are understood. A *finite coloring of X* is a coloring $c\colon X \to Y$ where Y is finite, and if $|Y| = r$ for some positive integer r, then c will be called an *r-coloring*. The nature of the set Y is irrelevant from a Ramsey theoretic perspective, and so we will view every r-coloring of X as a map $c\colon X \to [r]$.

Given a coloring $c\colon X \to Y$, a subset Z of X is said to be *monochromatic* (*with respect to the coloring* c) provided that $c(z_1) = c(z_2)$ for every $z_1, z_2 \in Z$, or equivalently, that $Z \subseteq c^{-1}(\{y\})$ for some $y \in Y$.

1.1.3. Let X be a (possibly infinite) nonempty set and Y a nonempty finite subset of X. For every $A \subseteq X$ the *density of A relative to Y* is defined by

$$\mathrm{dens}_Y(A) = \frac{|A \cap Y|}{|Y|}. \tag{1.2}$$

If it is clear from the context which set Y we are referring to (for instance, if Y coincides with X), then we shall drop the subscript Y and we shall denote the above quantity simply by $\mathrm{dens}(A)$. More generally, for every $f\colon X \to \mathbb{R}$ we set

$$\mathbb{E}_{y \in Y} f(y) = \frac{1}{|Y|} \sum_{y \in Y} f(y). \tag{1.3}$$

Notice that for every $A \subseteq X$ we have $\mathrm{dens}_Y(A) = \mathbb{E}_{y \in Y} \mathbf{1}_A(y)$ where $\mathbf{1}_A$ stands for the characteristic function of A, that is,

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{1.4}$$

The quantities $\mathrm{dens}_Y(A)$ and $\mathbb{E}_{y\in Y} f(y)$ have a natural probabilistic interpretation which is very important in the context of density Ramsey theory. Specifically, denoting by μ_Y the uniform probability measure on X concentrated on Y, we see that $\mathrm{dens}_Y(A) = \mu_Y(A)$ and $\mathbb{E}_{y\in Y} f(y) = \int f\, d\mu_Y$. A review of those tools from probability theory which are needed in this book can be found in Appendix E.

1.1.4. Recall that a *hypergraph* is a pair $\mathcal{H} = (V, E)$ where V is a nonempty set and $E \subseteq \mathcal{P}(V)$. The elements of V are called the *vertices* of $\mathcal{H}$ while the elements of E are called its *edges*. If E is a nonempty subset of $\binom{V}{r}$ for some $r \in \mathbb{N}$, then the hypergraph $\mathcal{H}$ will be called *r-uniform*. Thus, a 2-uniform hypergraph is just a graph with at least one edge.

1.1.5. For every function $f\colon \mathbb{N} \to \mathbb{N}$ and every $\ell \in \mathbb{N}$ by $f^{(\ell)}\colon \mathbb{N} \to \mathbb{N}$ we shall denote the *ℓ-th iteration of f* defined recursively by the rule

$$\begin{cases} f^{(0)}(n) = n, \\ f^{(\ell+1)}(n) = f\big(f^{(\ell)}(n)\big). \end{cases} \tag{1.5}$$

Note that this is a basic example of primitive recursion (see Appendix A).

1.2. Words over an alphabet

Let A be a *nonempty alphabet*, that is, a nonempty set. For every $n \in \mathbb{N}$ by A^n we shall denote the set of all sequences of length n having values in A. Precisely, A^0 contains just the empty sequence while if $n \geqslant 1$, then

$$A^n = \big\{(a_0, \dots, a_{n-1}) : a_i \in A \text{ for every } i \in \{0, \dots, n-1\}\big\}. \tag{1.6}$$

Also let

$$A^{<n+1} = \bigcup_{i=0}^{n} A^i \quad \text{and} \quad A^{<\mathbb{N}} = \bigcup_{n\in\mathbb{N}} A^n. \tag{1.7}$$

The elements of $A^{<\mathbb{N}}$ are called *words over A*, or simply *words* if A is understood. The *length* of a word w over A, denoted by $|w|$, is defined to be the unique natural number n such that $w \in A^n$. For every $i \in \mathbb{N}$ with $i \leqslant |w|$ by $w \upharpoonright i$ we shall denote the word of length i which is an initial segment of w. (In particular, we have that $w \upharpoonright 0$ is the empty word.) More generally, if X is a nonempty subset of $A^{<\mathbb{N}}$ such that for every $w \in X$ we have $i \leqslant |w|$, then we set

$$X \upharpoonright i = \{w \upharpoonright i : w \in X\}. \tag{1.8}$$

If w and u are two words over A, then the concatenation of w and u will be denoted by $w^\frown u$. Moreover, for every pair X, Y of nonempty subsets of $A^{<\mathbb{N}}$ we set

$$X^\frown Y = \{w^\frown u : w \in X \text{ and } u \in Y\}. \tag{1.9}$$

The *infimum* of w and u, denoted by $w \wedge u$, is defined to be the greatest common initial segment of w and u. Note that the infimum operation can be extended to nonempty sets of words. Specifically, for every nonempty subset X of $A^{<\mathbb{N}}$ the *infimum* of X, denoted by $\wedge X$, is the word over A of greatest length which is an initial segment of every $w \in X$. Observe that $w \wedge u = \wedge\{w, u\}$ for every $w, u \in A^{<\mathbb{N}}$.

If $<_A$ is a linear order on A, then for every distinct $w, u \in A^{<\mathbb{N}}$ we write $w <_{\mathrm{lex}} u$ provided that: (i) $|w| = |u| \geqslant 1$, and (ii) if $w = (w_0, \dots, w_{n-1})$, $u = (u_0, \dots, u_{n-1})$ and $i_0 = |w \wedge u|$, then $w_{i_0} <_A u_{i_0}$. Notice that for every positive integer n the partial order $<_{\mathrm{lex}}$ restricted on A^n is the usual lexicographical order.

1.2.1. Located words. Let A be a nonempty alphabet. For every (possibly empty) finite subset J of $\mathbb{N}$ by A^J we shall denote the set of all functions from J into A. An element of the set

$$\bigcup_{J \subseteq \mathbb{N} \text{ finite}} A^J \tag{1.10}$$

is called a *located word over A*. Clearly, every word over A is a located word over A. Indeed, notice that for every $n \in \mathbb{N}$ we have

$$A^{\{i \in \mathbb{N}:\, i<n\}} = A^n. \tag{1.11}$$

Conversely, we may identify located words over A with words over A as follows.

Definition 1.1. *Let A be a nonempty alphabet and let J be a nonempty finite subset of $\mathbb{N}$. Set $j = |J|$ and let $n_0 < \dots < n_{j-1}$ be the increasing enumeration of J. The* canonical isomorphism associated with J *is the bijection $\mathrm{I}_J \colon A^j \to A^J$ defined by the rule*

$$\mathrm{I}_J(w)(n_i) = w_i \tag{1.12}$$

for every $i \in \{0, \dots, j-1\}$ and every $w = (w_0, \dots, w_{j-1}) \in A^j$.

Moreover, observing that $A^{\emptyset} = A^0 = \{\emptyset\}$, we define the canonical isomorphism $\mathrm{I}_{\emptyset}$ associated with the empty set to be the identity.

If J, K are two finite subsets of $\mathbb{N}$ with $J \subseteq K$ and $w \in A^K$ is a located word over A, then by $w \upharpoonright J$ we shall denote the restriction of w on J. Notice that $w \upharpoonright J \in A^J$. Moreover, if I, J is a pair of finite subsets of $\mathbb{N}$ with $I \cap J = \emptyset$, then for every $u \in A^I$ and every $v \in A^J$ by (u, v) we shall denote the unique element z of $A^{I \cup J}$ such that $z \upharpoonright I = u$ and $z \upharpoonright J = v$.

1.2.2. Variable words. Let A be a nonempty alphabet and let n be a positive integer. We fix a set $\{x_0, \dots, x_{n-1}\}$ of distinct letters which is disjoint from A. We view $\{x_0, \dots, x_{n-1}\}$ as a set of variables. An *n-variable word over A* is a word v over the alphabet $A \cup \{x_0, \dots, x_{n-1}\}$ such that: (i) for every $i \in \{0, \dots, n-1\}$ the letter x_i appears in v at least once, and (ii) if $n \geqslant 2$, then for every $i, j \in \{0, \dots, n-1\}$ with $i < j$ all occurrences of x_i precede all occurrences of x_j.

If A is understood and $n \geqslant 2$, then n-variable words over A will be referred to as *n-variable words*. On the other hand, 1-variable words over A will be called simply as *variable words* and their variable will be denoted by x. A *left variable word* (*over A*) is a variable word whose leftmost letter is x.

Remark 1.1. The concept of a variable word is closely related to the notion of a parameter word introduced by Graham and Rothschild [**GR**]. Specifically, an *n-parameter word over A* is also a finite sequence having values in the alphabet $A \cup \{x_0, \dots, x_{n-1}\}$ which satisfies (i) above and such that: (ii)$'$ if $n \geqslant 2$, then for every $i, j \in \{0, \dots, n-1\}$ with $i < j$ the first occurrence of x_i precedes the first occurrence of x_j. In particular, every n-variable word is an n-parameter word. Of course, when $n=1$ the two notions coincide.

Now let A, B be nonempty alphabets. If v is a variable word over A and $b \in B$, then by $v(b)$ we shall denote the unique word over $A \cup B$ obtained by substituting in v all appearances of the variable x with b. Notice that if $b \in A$, then $v(b)$ is a word over A, while $v(x)=v$. More generally, let v be an n-variable word over A and let $b_0, \dots, b_{n-1} \in B$. By $v(b_0, \dots, b_{n-1})$ we shall denote the unique word over $A \cup B$ obtained by substituting in v all appearances of the letter x_i with b_i for every $i \in \{0, \dots, n-1\}$. Observe that if $B = A \cup \{x_0, \dots, x_{m-1}\}$ for some $m \in [n]$, then $v(b_0, \dots, b_{n-1})$ is a word over A if and only if $(b_0, \dots, b_{n-1})$ is a word over A; on the other hand, $v(b_0, \dots, b_{n-1})$ is an m-variable word over A if and only if $(b_0, \dots, b_{n-1})$ is an m-variable word over A. Taking into account these remarks, for every $m \in [n]$ we set

$$\mathrm{Subw}_m(v) = \big\{ v(b_0, \dots, b_{n-1}) : (b_0, \dots, b_{n-1}) \text{ is an } m\text{-variable word over } A \big\} \quad (1.13)$$

and we call an element of $\mathrm{Subw}_m(v)$ as an *m-variable subword of v*. Note that for every $u \in \mathrm{Subw}_m(v)$ and every $\ell \in [m]$ we have $\mathrm{Subw}_\ell(u) \subseteq \mathrm{Subw}_\ell(v)$.

1.3. Combinatorial spaces

Throughout this section, let A be a finite alphabet with $|A| \geqslant 2$. A *combinatorial space* of $A^{<\mathbb{N}}$ is a set of the form

$$V = \big\{ v(a_0, \dots, a_{n-1}) : a_0, \dots, a_{n-1} \in A \big\} \quad (1.14)$$

where n is a positive integer and v is a n-variable word over A. (Note that both n and v are unique since $|A| \geqslant 2$.) The positive integer n is called the *dimension* of V and is denoted by $\dim(V)$. The 1-dimensional combinatorial spaces will be called *combinatorial lines*.

Now let V be a combinatorial space of $A^{<\mathbb{N}}$ and set $n = \dim(V)$. Also let v be the (unique) n-variable word over A which generates V via formula (1.14) and notice that v induces a bijection between A^n and V. We will give this bijection a special name as follows.

Definition 1.2. *Let V be a combinatorial space of $A^{<\mathbb{N}}$. Set $n = \dim(V)$ and let v be the n-variable word which generates V via formula* (1.14)*. The* canonical isomorphism associated with V *is the bijection $\mathrm{I}_V : A^n \to V$ defined by the rule*

$$\mathrm{I}_V\big((a_0, \dots, a_{n-1})\big) = v(a_0, \dots, a_{n-1}) \quad (1.15)$$

for every $(a_0, \dots, a_{n-1}) \in A^n$.

We will view an n-dimensional combinatorial space V as a "copy" of A^n and, using the canonical isomorphism, we will identify V with A^n for most practical purposes. This identification is very convenient and will be constantly used throughout this book.

We proceed to discuss two alternative ways to define combinatorial spaces. First, for every nonempty finite sequence $(v_i)_{i=0}^{n-1}$ of variable words over A we set

$$V = \big\{v_0(a_0)^\frown\dots^\frown v_{n-1}(a_{n-1}) : a_0,\dots,a_{n-1}\in A\big\} \tag{1.16}$$

and we call V the *combinatorial space of* $A^{<\mathbb{N}}$ *generated by* $(v_i)_{i=0}^{n-1}$. Observe that two different finite sequences of variable words over A might generate the same combinatorial space of $A^{<\mathbb{N}}$.

Next, let V be a combinatorial space of $A^{<\mathbb{N}}$, set $n=\dim(V)$ and let v be the n-variable word over A which generates V via formula (1.14). Recall that v is a nonempty word over the alphabet $A\cup\{x_0,\dots,x_{n-1}\}$ and write $v=(v_0,\dots,v_{m-1})$ where $m=|v|$. For every $j\in\{0,\dots,n-1\}$ we set

$$X_j = \big\{i\in\{0,\dots,m-1\} : v_i = x_j\big\}. \tag{1.17}$$

Clearly, $X_0,\dots,X_{n-1}$ are nonempty subsets of $\{0,\dots,m-1\}$ and if $n\geqslant 2$, then $\max(X_i)<\min(X_{i+1})$ for every $i\in\{0,\dots,n-2\}$. The sets $X_0,\dots,X_{n-1}$ are called the *wildcard sets* of V. We also set

$$S = \{0,\dots,m-1\}\setminus\Big(\bigcup_{j=1}^{n-1} X_i\Big) \tag{1.18}$$

and we call S the *set of fixed coordinates* of V. Finally, the *constant part* of V is the located word $v\upharpoonright S\in A^S$. Note that the wildcard sets, the set of fixed coordinates and the constant part completely determine a combinatorial space.

1.3.1. Subspaces. If V and U are two combinatorial spaces of $A^{<\mathbb{N}}$, then we say that U is a *combinatorial subspace* of V if U is contained in V. For every combinatorial space V of $A^{<\mathbb{N}}$ and every $m\in[\dim(V)]$ by $\mathrm{Subsp}_m(V)$ we shall denote the set of all m-dimensional combinatorial subspaces of V.

We will present two different representations of the set $\mathrm{Subsp}_m(V)$ which are both straightforward consequences of the relevant definitions. The first representation relies on the canonical isomorphism I_V associated with V.

Fact 1.3. *Let V be a combinatorial space of $A^{<\mathbb{N}}$ and set $n=\dim(V)$. Then for every $m\in[n]$ the map*

$$\mathrm{Subsp}_m(A^n)\ni R\mapsto \mathrm{I}_V(R)\in\mathrm{Subsp}_m(V) \tag{1.19}$$

is a bijection.

The second representation will enable us to identify combinatorial subspaces with subwords. Specifically, we have the following fact.

Fact 1.4. *Let V be a combinatorial space of $A^{<\mathbb{N}}$. Set $n=\dim(V)$ and let v be the n-variable word over A which generates V via formula* (1.14)*. Then for every $m\in[n]$ the map*

$$\mathrm{Subw}_m(v)\ni u\mapsto\{u(a_0,\dots,a_{m-1}) : a_0,\dots,a_{m-1}\in A\}\in\mathrm{Subsp}_m(V) \tag{1.20}$$

is a bijection.

1.3.2. Restriction on smaller alphabets. Let V be a combinatorial space of $A^{<\mathbb{N}}$ and let I_V be the canonical isomorphism associated with V. For every $B \subseteq A$ with $|B| \geqslant 2$ we define the *restriction of V on B* by the rule

$$V \upharpoonright B = \big\{\mathrm{I}_V(u) : u \in B^{\dim(V)}\big\}. \tag{1.21}$$

Notice that the map $\mathrm{I}_V : B^{\dim(V)} \to V \upharpoonright B$ is a bijection, and so we may identify the restriction of V on B with a combinatorial space of $B^{<\mathbb{N}}$. Having this identification in mind, for every $m \in [\dim(V)]$ we set

$$\mathrm{Subsp}_m(V \upharpoonright B) = \big\{\mathrm{I}_V(X) : X \in \mathrm{Subsp}_m(B^{\dim(V)})\big\}. \tag{1.22}$$

By Definition 1.2 and (1.21), we have the following fact.

Fact 1.5. *Let V be a combinatorial space of $A^{<\mathbb{N}}$ and let $m \in [\dim(V)]$. Also let $B \subseteq A$ with $|B| \geqslant 2$. Then for every $R \in \mathrm{Subsp}_m(V \upharpoonright B)$ there exists a unique $U \in \mathrm{Subsp}_m(V)$ such that $R = U \upharpoonright B$.*

In particular, we have $\mathrm{Subsp}_m(V \upharpoonright B) \subseteq \{U \upharpoonright B : U \in \mathrm{Subsp}_m(V)\}$.

1.4. Reduced and extracted words

We are about to introduce two classes of combinatorial objects which are generated from sequences of variable words. In what follows, let A denote a finite alphabet with at least two letters.

1.4.1. Reduced words and variable words. Let $(w_i)_{i=0}^{n-1}$ be a nonempty finite sequence of variable words over A.

A *reduced word*[1] of $(w_i)_{i=0}^{n-1}$ is a word w over A of the form

$$w = w_0(a_0)^\frown \dots ^\frown w_{n-1}(a_{n-1}) \tag{1.23}$$

where $(a_0, \dots, a_{n-1})$ is a word over A. The set of all reduced words of $(w_i)_{i=0}^{n-1}$ will be denoted by $[(w_i)_{i=0}^{n-1}]$. Observe that $[(w_i)_{i=0}^{n-1}]$ coincides with the combinatorial space of $A^{<\mathbb{N}}$ generated by $(w_i)_{i=0}^{n-1}$.

A *reduced variable word* of $(w_i)_{i=0}^{n-1}$ is a variable word v over A of the form

$$v = w_0(\alpha_0)^\frown \dots ^\frown w_{n-1}(\alpha_{n-1}) \tag{1.24}$$

where $(\alpha_0, \dots, \alpha_{n-1})$ is a variable word over A. (Notice, in particular, that there exists $i \in \{0, \dots, n-1\}$ such that $\alpha_i = x$.) The set of all reduced variable words of $(w_i)_{i=0}^{n-1}$ will be denoted by $\mathrm{V}[(w_i)_{i=0}^{n}]$.

More generally, a finite sequence $(v_i)_{i=0}^{m-1}$ of variable words over A is said to be a *reduced subsequence* of $(w_i)_{i=0}^{n-1}$ if $m \in [n]$ and there exist a strictly increasing sequence $(n_i)_{i=0}^{m}$ in $\mathbb{N}$ with $n_0 = 0$ and $n_m = n$, and a sequence $(\alpha_j)_{j=0}^{n-1}$ in $A \cup \{x\}$ such that for every $i \in \{0, \dots, m-1\}$ we have $x \in \{\alpha_j : n_i \leqslant j \leqslant n_{i+1} - 1\}$ and

$$v_i = w_{n_i}(\alpha_{n_i})^\frown \dots ^\frown w_{n_{i+1}-1}(\alpha_{n_{i+1}-1}). \tag{1.25}$$

For every $m \in [n]$ by $\mathrm{V}_m[(w_i)_{i=0}^{n-1}]$ we shall denote the set of all reduced subsequences of $(w_i)_{i=0}^{n-1}$ of length m. Note that $\mathrm{V}_1[(w_i)_{i=0}^{n-1}] = \mathrm{V}[(w_i)_{i=0}^{n-1}]$.

---

[1] This terminology is, of course, group theoretic. The reader should have in mind though that it has somewhat different meaning in the present combinatorial context.

The above notions can be extended to infinite sequences of variable words. Specifically, let $\mathbf{w} = (w_i)$ be a sequence of variable words over A. For every positive integer n let $\mathbf{w} \upharpoonright n = (w_i)_{i=0}^{n-1}$ and set

$$[\mathbf{w}] = \bigcup_{n=1}^{\infty} [\mathbf{w} \upharpoonright n] \ \text{ and } \ \mathrm{V}[\mathbf{w}] = \bigcup_{n=1}^{\infty} \mathrm{V}[\mathbf{w} \upharpoonright n]. \tag{1.26}$$

An element of $[\mathbf{w}]$ will be called a *reduced word* of $\mathbf{w}$ while an element of $\mathrm{V}[\mathbf{w}]$ will be called a *reduced variable word* of $\mathbf{w}$. Moreover, for every positive integer m we define the set of all *reduced subsequences* of $\mathbf{w}$ of length m by the rule

$$\mathrm{V}_m[\mathbf{w}] = \bigcup_{n=m}^{\infty} \mathrm{V}_m[\mathbf{w} \upharpoonright n] \tag{1.27}$$

Finally, we say that an infinite sequence $\mathbf{v} = (v_i)$ of variable words over A is a *reduced subsequence* of $\mathbf{w}$ if for every integer $m \geqslant 1$ we have $\mathbf{v} \upharpoonright m \in \mathrm{V}_m[\mathbf{w}]$. The set of all reduced subsequences of $\mathbf{w}$ of infinite length will be denoted by $\mathrm{V}_\infty[\mathbf{w}]$.

We proceed to discuss some basic properties of reduced words and variable words. We first observe that if $(w_i)_{i=0}^{n-1}$ is a finite sequence of variable words over A, then every reduced subsequence of $(w_i)_{i=0}^{n-1}$ corresponds to a combinatorial subspace $[(w_i)_{i=0}^{n-1}]$. More precisely, we have the following fact.

Fact 1.6. *Let $(w_i)_{i=0}^{n-1}$ be a nonempty finite sequence of variable words over A and set $W = [(w_i)_{i=0}^{n-1}]$. Then for every $m \in [n]$ the map*

$$\mathrm{V}_m[(w_i)_{i=0}^{n-1}] \ni (v_i)_{i=0}^{m-1} \mapsto [(v_i)_{i=0}^{m-1}] \in \mathrm{Subsp}_m(W) \tag{1.28}$$

is onto. Moreover, this map is a bijection between $\mathrm{V}_1[(w_i)_{i=0}^{n-1}]$ and $\mathrm{Subsp}_1(W)$.

We also have the following coherence properties.

Fact 1.7. *Let $\mathbf{v}, \mathbf{w}$ be two nonempty sequences (of finite or infinite length) of variable words over A and assume that $\mathbf{v}$ is a reduced subsequence of $\mathbf{w}$. Then we have $[\mathbf{v}] \subseteq [\mathbf{w}]$, $\mathrm{V}[\mathbf{v}] \subseteq \mathrm{V}[\mathbf{w}]$ and $\mathrm{V}_m[\mathbf{v}] \subseteq \mathrm{V}_m[\mathbf{w}]$ for every positive integer m which is less than or equal to the length of $\mathbf{v}$. Moreover, if both $\mathbf{v}$ and $\mathbf{w}$ are infinite sequences, then we have $\mathrm{V}_\infty[\mathbf{v}] \subseteq \mathrm{V}_\infty[\mathbf{w}]$.*

1.4.2. Extracted words and variable words. As in the previous subsection, let $(w_i)_{i=0}^{n-1}$ be a nonempty finite sequence of variable words over A.

An *extracted word* of $(w_i)_{i=0}^{n-1}$ is a reduced word of a subsequence of $(w_i)_{i=0}^{n-1}$ while an *extracted variable word* of $(w_i)_{i=0}^{n-1}$ is a reduced variable word of a subsequence of $(w_i)_{i=0}^{n-1}$. (Thus, an extracted variable word of $(w_i)_{i=0}^{n-1}$ is of the form $w_{i_0}(\alpha_0)^\frown \dots ^\frown w_{i_\ell}(\alpha_\ell)$ where $\ell \in \mathbb{N}$, $0 \leqslant i_0 < \dots < i_\ell \leqslant n-1$ and $(\alpha_0, \dots, \alpha_\ell)$ is a variable word over A.) An *extracted subsequence* of $(w_i)_{i=0}^{n-1}$ is a reduced subsequence of a subsequence of $(w_i)_{i=0}^{n-1}$. By $\mathrm{E}[(w_i)_{i=0}^{n-1}]$ and $\mathrm{EV}[(w_i)_{i=0}^{n-1}]$ we shall denote the sets of all extracted words and all extracted variable words of $(w_i)_{i=0}^{n-1}$ respectively. Moreover, for every $m \in [n]$ the set of all extracted subsequences of $(w_i)_{i=0}^{n-1}$ of length m will be denoted by $\mathrm{EV}_m[(w_i)_{i=0}^{n-1}]$.

Next, let $\mathbf{w} = (w_i)$ be an infinite sequence of variable words over A. We set

$$\mathrm{E}[\mathbf{w}] = \bigcup_{n=1}^{\infty} \mathrm{E}[\mathbf{w} \upharpoonright n] \ \text{ and } \ \mathrm{EV}[\mathbf{w}] = \bigcup_{n=1}^{\infty} \mathrm{EV}[\mathbf{w} \upharpoonright n] \tag{1.29}$$

and for every positive integer m let

$$\mathrm{EV}_m[\mathbf{w}] = \bigcup_{n=m}^{\infty} \mathrm{EV}_m[\mathbf{w} \upharpoonright n]. \tag{1.30}$$

On the other hand, by $\mathrm{EV}_\infty[\mathbf{w}]$ we shall denote the set of all infinite extracted subsequences of $\mathbf{w}$, that is, the set of all (infinite) sequences of variable words over A which are reduced subsequences of a subsequence of $\mathbf{w}$.

We close this section with the following analogue of Fact 1.7.

FACT 1.8. *Let $\mathbf{v}, \mathbf{w}$ be two nonempty sequences (of finite or infinite length) of variable words over A and assume that $\mathbf{v}$ is an extracted subsequence of $\mathbf{w}$. Then we have $\mathrm{E}[\mathbf{v}] \subseteq \mathrm{E}[\mathbf{w}]$, $\mathrm{EV}[\mathbf{v}] \subseteq \mathrm{EV}[\mathbf{w}]$ and $\mathrm{EV}_m[\mathbf{v}] \subseteq \mathrm{EV}_m[\mathbf{w}]$ for every positive integer m which is less than or equal to the length of $\mathbf{v}$. Moreover, if both $\mathbf{v}$ and $\mathbf{w}$ are infinite sequences, then we have $\mathrm{EV}_\infty[\mathbf{v}] \subseteq \mathrm{EV}_\infty[\mathbf{w}]$.*

1.5. Carlson–Simpson spaces

Let A be a finite alphabet with $|A| \geqslant 2$. This alphabet will be fixed throughout this section. A *finite-dimensional Carlson–Simpson system over A* is a pair $\langle t, (w_i)_{i=0}^{d-1} \rangle$ where t is a word over A and $(w_i)_{i=0}^{d-1}$ a nonempty finite sequence of left variable words over A. The length d of the finite sequence $(w_i)_{i=0}^{d-1}$ will be called the *dimension* of the system.

A *finite-dimensional Carlson–Simpson space* of $A^{<\mathbb{N}}$ is a set of the form

$$W = \{t\} \cup \big\{ t^\frown w_0(a_0)^\frown \dots {}^\frown w_{m-1}(a_{m-1}) : m \in [d] \text{ and } a_0, \dots, a_{m-1} \in A \big\} \tag{1.31}$$

where $\langle t, (w_i)_{i=0}^{d-1} \rangle$ is a finite-dimensional Carlson–Simpson system over A. Note that the system $\langle t, (w_i)_{i=0}^{d-1} \rangle$ which generates W via formula (1.31) is unique; it will be called the *generating system* of W. The *dimension* of W, denoted by $\dim(W)$, is the dimension of its generating system (that is, the length of the finite sequence $(w_i)_{i=0}^{d-1}$). The 1-dimensional Carlson–Simpson spaces will be called *Carlson–Simpson lines*.

Let W be a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$, set $d = \dim(W)$ and let $\langle t, (w_i)_{i=0}^{d-1} \rangle$ be its generating system. For every $m \in \{0, \dots, d\}$ we define the *m-level* $W(m)$ of W by setting $W(0) = \{t\}$ and

$$W(m) = \big\{ t^\frown w_0(a_0)^\frown \dots {}^\frown w_{m-1}(a_{m-1}) : a_0, \dots, a_{m-1} \in A \big\} \tag{1.32}$$

if $m \in [d]$. Observe that $W = W(0) \cup \dots \cup W(d)$ and notice that for every $m \in [d]$ the m-level $W(m)$ of W is an m-dimensional combinatorial subspace of A^{n_m} where $n_m = |t| + \sum_{i=0}^{m-1} |w_i|$. The *level set* of W, denoted by $L(W)$, is defined by

$$L(W) = \big\{ n \in \mathbb{N} : W(m) \subseteq A^n \text{ for some } m \in \{0, \dots, d\} \big\}. \tag{1.33}$$

Equivalently, we have $L(W) = \big\{ |t| \big\} \cup \big\{ |t| + \sum_{i=0}^{m-1} |w_i| : m \in [d] \big\}$.

1.5.1. Subsystems and subspaces. Let d, m be positive integers, and let $\mathbf{w} = \langle t, (w_i)_{i=0}^{d-1} \rangle$ and $\mathbf{u} = \langle s, (v_i)_{i=0}^{m-1} \rangle$ be two Carlson–Simpson systems over A of dimensions d and m respectively. We say that $\mathbf{u}$ is a *subsystem* of $\mathbf{w}$ if $m \leqslant d$ and there exist a strictly increasing sequence $(n_i)_{i=0}^{m}$ in $\{0, \dots, d\}$ and a sequence $(a_j)_{j=0}^{n_m - 1}$ in $A \cup \{x\}$ such that the following conditions are satisfied.

(C1) If $n_0 = 0$, then $s = t$. Otherwise, we have $a_0, \dots, a_{n_0-1} \in A$ and

$$s = t^\frown w_0(a_0)^\frown \dots ^\frown w_{n_0-1}(a_{n_0-1}).$$

(C2) For every $i \in \{0, \dots, m-1\}$ we have $a_{n_i} = x$ and

$$v_i = w_{n_i}(a_{n_i})^\frown \dots ^\frown w_{n_{i+1}-1}(a_{n_{i+1}-1}).$$

The set of all m-dimensional subsystems of $\mathbf{w}$ will be denoted by $\mathrm{Subsys}_m(\mathbf{w})$.

On the other hand, if W and U are two finite-dimensional Carlson–Simpson spaces of $A^{<\mathbb{N}}$, then we say that U is a (*Carlson–Simpson*) *subspace* of W if U is contained in W. (This implies, in particular, that $\dim(U) \leqslant \dim(W)$.) For every $m \in [\dim(W)]$ by $\mathrm{SubCS}_m(W)$ we shall denote the set of all m-dimensional Carlson–Simpson subspaces of W. Notice that, setting

$$W \upharpoonright m+1 = W(0) \cup \dots \cup W(m), \tag{1.34}$$

we have $W \upharpoonright m+1 \in \mathrm{SubCS}_m(W)$.

There is a natural correspondence between subsystems and subspaces. Indeed, let W and U be two finite-dimensional Carlson–Simpson spaces of $A^{<\mathbb{N}}$ generated by the systems $\mathbf{w}$ and $\mathbf{u}$ respectively, and observe that U is a subspace of W if and only if $\mathbf{u}$ is a subsystem of $\mathbf{w}$. More precisely, we have the following fact.

FACT 1.9. *Let W be a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ and let $\mathbf{w}$ be its generating system. For every Carlson–Simpson subspace U of W let $\mathbf{w}_U$ be its generating system. Then for every $m \in [\dim(W)]$ the map*

$$\mathrm{SubCS}_m(W) \ni U \mapsto \mathbf{w}_U \in \mathrm{Subsys}_m(\mathbf{w}) \tag{1.35}$$

is a bijection.

1.5.2. Canonical isomorphisms. Let d be a positive integer and note that the archetypical example of a d-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ is the set $A^{<d+1}$ of all finite sequences in A of length less than or equal to d. In fact, every d-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ can be viewed as a "copy" of $A^{<d+1}$. The philosophy is identical to that in Section 1.3.

DEFINITION 1.10. *Let W be a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$, set $d = \dim(W)$ and let $\langle t, (w_i)_{i=0}^{d-1} \rangle$ be its generating system. The* canonical isomorphism associated with W *is the bijection $\mathrm{I}_W \colon A^{<d+1} \to W$ defined by setting $\mathrm{I}_W(\emptyset) = t$ and*

$$\mathrm{I}_W\big((a_0, \dots, a_{m-1})\big) = t^\frown w_0(a_0)^\frown \dots ^\frown w_{m-1}(a_{m-1}) \tag{1.36}$$

for every $m \in [d]$ and every $(a_0, \dots, a_{m-1}) \in A^m$.

The canonical isomorphism preserves all structural properties one is interested in while working in the category of Carlson–Simpson spaces. In particular, we have the following analogue of Fact 1.3.

FACT 1.11. *Let W be a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ and set $d = \dim(W)$. Then for every $m \in [d]$ the map*

$$\mathrm{SubCS}_m(A^{<d+1}) \ni R \mapsto \mathrm{I}_W(R) \in \mathrm{SubCS}_m(W) \tag{1.37}$$

is a bijection.

Another basic property of canonical isomorphisms is that they preserve infima.

FACT 1.12. *Let W be a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$. Then for every nonempty subset F of $A^{<\dim(W)+1}$ we have $\mathrm{I}_W(\wedge F) = \wedge \mathrm{I}_W(F)$.*

We continue by presenting a method to produce Carlson–Simpson spaces from combinatorial spaces. The method is based on canonical isomorphisms. Specifically, let W be a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ and let $m \in [\dim(W)]$. Recall that for every $U \in \mathrm{SubCS}_m(W)$ the m-level $U(m)$ of U is an m-dimensional combinatorial space of $A^{<\mathbb{N}}$. Let $d = \dim(W)$, and set

$$\mathrm{SubCS}^{\max}_m(W) = \big\{U \in \mathrm{SubCS}_m(W) : U(m) \subseteq W(d)\big\}. \tag{1.38}$$

That is, $\mathrm{SubCS}^{\max}_m(W)$ is the set of all m-dimensional Carlson–Simpson subspaces of W whose last level is contained in the last level of W.

LEMMA 1.13. *Let W a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ and set $d = \dim(W)$. Then for every $m \in [d]$ the map*

$$\mathrm{SubCS}^{\max}_m(W) \ni U \mapsto U(m) \in \mathrm{Subsp}_m\big(W(d)\big) \tag{1.39}$$

is a bijection.

PROOF. Notice that every $V \in \mathrm{Subsp}_m(A^d)$ is of the form $V = R(m)$ for some unique $R \in \mathrm{SubCS}^{\max}_m(A^{<d+1})$. By Fact 1.11 and taking into account this remark, the result follows. □

1.5.3. Restriction on smaller alphabets. Let W be a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ and let I_W be the canonical isomorphism associated with W. As in Subsection 1.3.2, for every $B \subseteq A$ with $|B| \geqslant 2$ we define the *restriction of W on B* by the rule

$$W \upharpoonright B = \big\{\mathrm{I}_W(u) : u \in B^{<\dim(W)+1}\big\} \tag{1.40}$$

and for every $m \in [\dim(W)]$ we set

$$\mathrm{SubCS}_m(W \upharpoonright B) = \big\{\mathrm{I}_W(X) : X \in \mathrm{SubCS}_m(B^{<\dim(W)+1})\big\}. \tag{1.41}$$

Observe that the map $\mathrm{I}_W \colon B^{<\dim(W)+1} \to W \upharpoonright B$ is a bijection. Moreover, we have the following fact.

FACT 1.14. *Let W be a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$. Also let $m \in [\dim(W)]$ and $B \subseteq A$ with $|B| \geqslant 2$. Then for every $R \in \mathrm{SubCS}_m(W \upharpoonright B)$ there exists a unique $U \in \mathrm{SubCS}_m(W)$ such that $R = U \upharpoonright B$.*

In particular, we have $\mathrm{SubCS}_m(W \upharpoonright B) \subseteq \{U \upharpoonright B : U \in \mathrm{SubCS}_m(W)\}$.

1.5.4. Infinite-dimensional Carlson–Simpson spaces. We proceed to discuss the infinite versions of the concepts introduced in this section so far.

An *infinite-dimensional Carlson–Simpson system over* A is a pair $\langle t, (w_i)\rangle$ where t is a word over A and (w_i) is a sequence of left variable words over A. On the other hand, an *infinite-dimensional Carlson–Simpson space* of $A^{<\mathbb{N}}$ is a set of the form

$$W = \{t\} \cup \big\{t^\frown w_0(a_0)^\frown \dots ^\frown w_m(a_m) : m \in \mathbb{N} \text{ and } a_0, \dots, a_m \in A\big\} \tag{1.42}$$

where $\langle t, (w_i)\rangle$ is an infinite-dimensional Carlson–Simpson system over A. This system is clearly unique and will also be called the *generating system* of W. Respectively, for every $m \in \mathbb{N}$ the *m-level* $W(m)$ of W is defined by setting $W(0) = t$ and $W(m) = \{t^\frown w_0(a_0)^\frown \dots ^\frown w_{m-1}(a_{m-1}) : a_0, \dots, a_{m-1} \in A\}$ if $m \geqslant 1$, while the *level set* of W is defined by

$$L(W) = \{n \in \mathbb{N} : W(m) \subseteq A^n \text{ for some } m \in \mathbb{N}\}. \tag{1.43}$$

We have the following analogue of Definition 1.10.

Definition 1.15. *Let W be an infinite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ and let $\langle t, (w_i)\rangle$ be its generating system. The* canonical isomorphism associated with W *is the bijection* $\mathrm{I}_W : A^{<\mathbb{N}} \to W$ *defined by setting* $\mathrm{I}_W(\emptyset) = t$ *and*

$$\mathrm{I}_W\big((a_0, \dots, a_m)\big) = t^\frown w_0(a_0)^\frown \dots ^\frown w_m(a_m) \tag{1.44}$$

for every $m \in \mathbb{N}$ and every $a_0, \dots, a_m \in A$.

Now let W be an infinite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$. A *Carlson–Simpson subspace* of W is a finite or infinite dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ which is contained in W. For every positive integer m by $\mathrm{SubCS}_m(W)$ we shall denote the set of all m-dimensional Carlson–Simpson subspaces of W while $\mathrm{SubCS}_\infty(W)$ stands for the set of all infinite-dimensional Carlson–Simpson subspaces of W. We close this section with the following analogue of Fact 1.11.

Fact 1.16. *Let W be an infinite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$. Also let m be a positive integer. Then the maps*

$$\mathrm{SubCS}_m(A^{<\mathbb{N}}) \ni R \mapsto \mathrm{I}_W(R) \in \mathrm{SubCS}_m(W) \tag{1.45}$$

and

$$\mathrm{SubCS}_\infty(A^{<\mathbb{N}}) \ni R \mapsto \mathrm{I}_W(R) \in \mathrm{SubCS}_\infty(W) \tag{1.46}$$

are both bijections.

1.6. Trees

By the term *tree* we mean a (possibly empty) partially ordered set $(T, <_T)$ such that the set $\{s \in T : s <_T t\}$ is finite and linearly ordered under $<_T$ for every $t \in T$. The cardinality of this set is defined to be the *length of* t *in* T and will be denoted by $|t|_T$. For every $n \in \mathbb{N}$ the *n-level* of T, denoted by $T(n)$, is defined to be the set

$$T(n) = \{t \in T : |t|_T = n\}. \tag{1.47}$$

The *height* of T, denoted by $h(T)$, is defined as follows. First, we set $h(T)=0$ if T is empty. If T is nonempty and there exists $n\in\mathbb{N}$ such that $T(n)=\emptyset$, then we set

$$h(T)=\max\{n\in\mathbb{N}:T(n)\neq\emptyset\}+1;$$

otherwise, we set $h(T)=\infty$. Notice that the height of a nonempty finite tree T is the cardinality of the set of all nonempty levels of T.

An element of a tree T is called a *node* of T. For every node t of T by $\mathrm{Succ}_T(t)$ we shall denote the set of all *successors of t in T*, that is,

$$\mathrm{Succ}_T(t)=\{s\in T:t=s \text{ or } t<_T s\}. \tag{1.48}$$

The set of *immediate successors of t in T* is the subset of $\mathrm{Succ}_T(t)$ defined by

$$\mathrm{ImmSucc}_T(t)=\{s\in T:t<_T s \text{ and } |s|_T=|t|_T+1\}. \tag{1.49}$$

A node t of T is called *maximal* if $\mathrm{ImmSucc}_T(t)$ is empty.

A nonempty tree T is said to be *finitely branching* (respectively, *pruned*) if for every $t\in T$ the set of immediate successors of t in T is finite (respectively, nonempty). It is said to be *rooted* if $T(0)$ is a singleton; in this case, the unique node of $T(0)$ is called the *root* of T.

A *chain* of a tree T is a subset of T which is linearly ordered under $<_T$. A maximal (with respect to inclusion) chain of T is called a *branch* of T. The tree T is said to be *balanced* if all branches of T have the same cardinality. Note that a tree of infinite height is balanced if and only if it is pruned.

Now let T be a tree and let D be a subset of T. The *level set of D in T*, denoted by $L_T(D)$, is defined to be the set

$$L_T(D)=\{n\in\mathbb{N}:D\cap T(n)\neq\emptyset\}. \tag{1.50}$$

Moreover, for every $n\in\mathbb{N}$ let

$$D\upharpoonright n=\bigcup_{\{m\in\mathbb{N}:m<n\}} D\cap T(m). \tag{1.51}$$

(In particular, $D\upharpoonright 0$ is the empty tree.) Note that if $D=T$ and $n\geqslant 1$, then

$$T\upharpoonright n=T(0)\cup\cdots\cup T(n-1). \tag{1.52}$$

More generally, for every $M\subseteq\mathbb{N}$ we set

$$D\upharpoonright M=\bigcup_{m\in M} D\cap T(m). \tag{1.53}$$

Finally, if D is finite, then we define the *depth of D in T*, denoted by $\mathrm{depth}_T(D)$, to be the least $n\in\mathbb{N}$ such that $D\subseteq T\upharpoonright n$. Observe that for every nonempty finite subset D of T we have $\mathrm{depth}_T(D)=\max\big(L_T(D)\big)+1$.

1.6.1. Strong subtrees. A *subtree* of a tree $(T,<_T)$ is a subset of T viewed as a tree equipped with the induced partial ordering. An *initial subtree* of T is a subtree of T of the form $T\upharpoonright n$ for some $n\in\mathbb{N}$ with $n<h(T)$. The following class of subtrees is of particular importance in the context of Ramsey theory.

DEFINITION 1.17. *A subtree S of a tree T is said to be* strong *if either S is empty, or S is nonempty and satisfies the following conditions.*

(a) *The tree S is rooted and balanced.*
(b) *Every level of S is a subset of some level of T, that is, for every $n \in \mathbb{N}$ with $n < h(S)$ there exists $m \in \mathbb{N}$ such that $S(n) \subseteq T(m)$.*
(c) *For every non-maximal node $s \in S$ and every $t \in \mathrm{ImmSucc}_T(s)$ the set $\mathrm{ImmSucc}_S(s) \cap \mathrm{Succ}_T(t)$ is a singleton.*

For every $k \in \mathbb{N}$ with $k < h(T)$ by $\mathrm{Str}_k(T)$ we shall denote the set of all strong subtrees of T of height k. Notice that $\mathrm{Str}_0(T)$ contains only the empty tree; on the other hand, we have $\mathrm{Str}_1(T) = \big\{\{t\} : t \in T\big\}$ and so we may identify the set $\mathrm{Str}_1(T)$ with T. If T has infinite height, then by $\mathrm{Str}_{<\infty}(T)$ and $\mathrm{Str}_\infty(T)$ we shall denote the set of all strong subtrees of T of finite and infinite height respectively.

We isolate below some elementary (though basic) properties of strong subtrees.

Fact 1.18. *Let T be a tree and let S be a strong subtree of T.*

(a) *Every strong subtree of S is also a strong subtree of T.*
(b) *If T is pruned and $S \in \mathrm{Str}_k(T)$ for some $k \in \mathbb{N}$, then there is $R \in \mathrm{Str}_\infty(T)$ (not necessarily unique) such that $S = R \upharpoonright k$.*

1.6.2. Homogeneous trees. Let $b \in \mathbb{N}$ with $b \geqslant 2$ and recall that $[b]^{<\mathbb{N}}$ stands for the set of all finite sequences having values in $[b]$. We view $[b]^{<\mathbb{N}}$ as a tree equipped with the (strict) partial order $\sqsubset$ of end-extension. In particular, for every positive integer n the set $[b]^{<n}$ is the initial subtree of $[b]^{<\mathbb{N}}$ of height n.

A *homogeneous tree* is a nonempty strong subtree T of $[b]^{<\mathbb{N}}$ for some $b \in \mathbb{N}$ with $b \geqslant 2$. The (unique) integer b is called the *branching number* of T and is denoted by b_T. Note that a homogeneous tree T of height k is just a "copy" of $[b_T]^{<k}$ inside $[b_T]^{<\mathbb{N}}$. More precisely, we have the following definition.

Definition 1.19. *Let T be a homogeneous tree of finite height. The* canonical isomorphism associated with *T is the unique bijection $\mathrm{I}_T \colon [b_T]^{<h(T)} \to T$ such that for every $t, t' \in [b_T]^{<h(T)}$ we have*

(P1) *$|t| = |t'|$ if and only if $|\mathrm{I}_T(t)|_T = |\mathrm{I}_T(t')|_T$,*
(P2) *$t \sqsubset t'$ if and only if $\mathrm{I}_T(t) \sqsubset \mathrm{I}_T(t')$, and*
(P3) *$t <_{\mathrm{lex}} t'$ if and only if $\mathrm{I}_T(t) <_{\mathrm{lex}} \mathrm{I}_T(t')$.*

Respectively, the canonical isomorphism associated with *a homogeneous tree T of infinite height is the unique bijection $\mathrm{I}_T \colon [b_T]^{<\mathbb{N}} \to T$ satisfying* (P1), (P2) *and* (P3) *for every $t, t' \in [b_T]^{<\mathbb{N}}$.*

1.6.3. Vector trees. A *vector tree* is a finite sequence $\mathbf{T} = (T_1, \dots, T_d)$ of trees having common height. This common height is defined to be the *height* of $\mathbf{T}$ and will be denoted by $h(\mathbf{T})$. A vector tree $\mathbf{T} = (T_1, \dots, T_d)$ is said to be *finitely branching* (respectively, *pruned*, *rooted*, *balanced*) if for every $i \in [d]$ the tree T_i is finitely branching (respectively, pruned, rooted, balanced).

If $\mathbf{T} = (T_1, \dots, T_d)$ is a vector tree, then a *vector subset* of $\mathbf{T}$ is a finite sequence $\mathbf{D} = (D_1, \dots, D_d)$ where $D_i \subseteq T_i$ for every $i \in [d]$. As in (1.51), for every $n \in \mathbb{N}$ let

$$\mathbf{D} \upharpoonright n = (D_1 \upharpoonright n, \dots, D_d \upharpoonright n) \tag{1.54}$$

and, more generally, for every $M \subseteq \mathbb{N}$ let

$$\mathbf{D} \upharpoonright M = (D_1 \upharpoonright M, \dots, D_d \upharpoonright M). \tag{1.55}$$

In particular, if $\mathbf{D} = \mathbf{T}$ and $n \geqslant 1$, then we have $\mathbf{T} \upharpoonright n = (T_1 \upharpoonright n, \dots, T_d \upharpoonright n)$.

Now let $\mathbf{D} = (D_1, \dots, D_d)$ be a vector subset of $\mathbf{T} = (T_1, \dots, T_d)$. If D_i is finite for every $i \in [d]$, then the *depth of* $\mathbf{D}$ *in* $\mathbf{T}$, denoted by $\mathrm{depth}_{\mathbf{T}}(\mathbf{D})$, is defined by

$$\mathrm{depth}_{\mathbf{T}}(\mathbf{D}) = \min\{n \in \mathbb{N} : \mathbf{D} \text{ is a vector subset of } \mathbf{T} \upharpoonright n\}. \tag{1.56}$$

On the other hand, we say that $\mathbf{D}$ is *level compatible* if there exists $L \subseteq \mathbb{N}$ such that $L_{T_i}(D_i) = L$ for every $i \in [d]$. The (unique) set L will be denoted by $L_{\mathbf{T}}(\mathbf{D})$ and will be called the *level set of* $\mathbf{D}$ *in* $\mathbf{T}$. Moreover, for every $n \in L_{\mathbf{T}}(\mathbf{D})$ we set

$$\otimes\mathbf{D}(n) = \big(D_1 \cap T_1(n)\big) \times \cdots \times \big(D_d \cap T_d(n)\big) \tag{1.57}$$

and we define the *level product* of $\mathbf{D}$ by the rule

$$\otimes\mathbf{D} = \bigcup_{n \in L_{\mathbf{T}}(\mathbf{D})} \otimes\mathbf{D}(n). \tag{1.58}$$

In particular, for every $n \in \mathbb{N}$ with $n < h(\mathbf{T})$ we have

$$\otimes\mathbf{T}(n) = T_1(n) \times \cdots \times T_d(n)$$

and

$$\otimes\mathbf{T} = \bigcup_{n < h(\mathbf{T})} \otimes\mathbf{T}(n).$$

Finally, for every $\mathbf{t} = (t_1, \dots, t_d) \in \otimes\mathbf{T}$ by $|\mathbf{t}|_{\mathbf{T}}$ we shall denote the unique natural number n such that $\mathbf{t} \in \otimes\mathbf{T}(n)$.

1.6.4. Vector strong subtrees. The concept of a strong subtree is naturally extended to vector trees. Specifically, we have the following definition.

Definition 1.20. *Let $\mathbf{T} = (T_1, \dots, T_d)$ be a vector tree. A* vector strong subtree *of $\mathbf{T}$ is a vector subset $\mathbf{S} = (S_1, \dots, S_d)$ of $\mathbf{T}$ which is level compatible (that is, there exists $L \subseteq \mathbb{N}$ with $L_{T_i}(S_i) = L$ for every $i \in [d]$) and such that S_i is a strong subtree of T_i for every $i \in [d]$.*

Notice that every vector strong subtree $\mathbf{S}$ of a vector tree $\mathbf{T}$ is a vector tree on its own, and observe that its height $h(\mathbf{S})$ coincides with the common height of $S_1, \dots, S_d$. For every $k \in \mathbb{N}$ with $k < h(\mathbf{T})$ by $\mathrm{Str}_k(\mathbf{T})$ we shall denote the set of all vector strong subtrees of $\mathbf{T}$ of height k. If, in addition, $\mathbf{T}$ is of infinite height, then by $\mathrm{Str}_{<\infty}(\mathbf{T})$ and $\mathrm{Str}_\infty(\mathbf{T})$ we shall denote the set of all vector strong subtrees of $\mathbf{T}$ of finite and infinite height respectively.

We close this subsection with the following analogue of Fact 1.18.

Fact 1.21. *Let $\mathbf{T}$ be a vector tree and let $\mathbf{S}$ be a vector strong subtree of $\mathbf{T}$.*

(a) *Every vector strong subtree of $\mathbf{S}$ is also a vector strong subtree of $\mathbf{T}$.*
(b) *If $\mathbf{T}$ is pruned and $\mathbf{S} \in \mathrm{Str}_k(\mathbf{T})$ for some $k \in \mathbb{N}$, then there is $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{T})$ (not necessarily unique) such that $\mathbf{S} = \mathbf{R} \upharpoonright k$.*

1.6.5. Vector homogeneous trees. A *vector homogeneous tree* is a vector tree $\mathbf{T} = (T_1, \dots, T_d)$ such that T_i is homogeneous for every $i \in [d]$. Observe that a vector homogeneous tree $\mathbf{T} = (T_1, \dots, T_d)$ is a vector strong subtree of $([b_{T_1}]^{<\mathbb{N}}, \dots, [b_{T_d}]^{<\mathbb{N}})$ with $h(\mathbf{T}) \geqslant 1$.

1.7. Notes and remarks

1.7.1. The notion of a combinatorial line originates from the classical paper of Hales and Jewett [**HJ**]. On the other hand, the concepts of a reduced and an extracted word appeared first in the work of Carlson [**C**]. Carlson–Simpson spaces were introduced in [**CS**]; however, our exposition follows later presentations (see, e.g., [**DKT3, McC1**]).

1.7.2. There are several (essentially) equivalent ways to define trees. We followed the set theoretic approach mainly for historical reasons (see, in particular, the discussion in Section 3.4). We also note that the notion of a strong subtree was introduced by Laver in the late 1960s (see also [**M2, M3**]).

Part 1

Coloring theory

CHAPTER 2

Combinatorial spaces

2.1. The Hales–Jewett theorem

The following theorem is due to Hales and Jewett [**HJ**] and the corresponding bounds are due to Shelah [**Sh1**].

THEOREM 2.1. *For every pair k, r of positive integers with $k \geqslant 2$ there exists a positive integer N with the following property. If $n \geqslant N$, then for every alphabet A with $|A| = k$ and every r-coloring of A^n there exists a variable word w over A of length n such that the set $\{w(a) : a \in A\}$ is monochromatic. The least positive integer with this property will be denoted by* $\mathrm{HJ}(k, r)$.

Moreover, the numbers $\mathrm{HJ}(k, r)$ *are upper bounded by a primitive recursive function belonging to the class* $\mathcal{E}^5$.

The Hales–Jewett theorem is considered to be one of the cornerstones of modern Ramsey theory, for several good reasons. We will single out two of these reasons which appear to be the most substantial.

First, the Hales–Jewett theorem is at the right level of generality, and as such, it is applicable to a wide range of problems. This is ultimately related to the rich combinatorial nature of the hypercube A^n which can encode both algebraic and geometric information. For example, looking at the expansion of the natural numbers in base ℓ (where $\ell \geqslant 2$ is a fixed integer), one sees that the Hales–Jewett theorem implies the theorem of van der Waerden on arithmetic progressions [**vdW**]. In fact, with a bit more effort (see [**GRS**] for details) one sees that the Hales–Jewett theorem also implies the higher-dimensional analogue[1] of this classical result. The penetrating power of the Hales–Jewett theorem is unique and there are numerous more examples some of which we will encounter later on in this book.

Beyond the scope of applications, the Hales–Jewett theorem is a constant source of inspiration in Ramsey theory. In particular, there are several different proofs (as well as extensions) of this result. We will follow Shelah's proof [**Sh1**] which proceeds by induction on the cardinality of the finite alphabet A. The general inductive step splits into two parts. First, given a finite coloring c of A^n, one finds a combinatorial subspace W of A^n of large dimension such that the coloring c restricted on W is "simple". Once the coloring has been made "simple", the proof is completed with an application of the inductive assumptions. This method is very fruitful and many of the results that we present in this book are proved following this general scheme.

[1]The higher-dimensional version of the van der Waerden theorem is known as *Gallai's theorem* (see [**GRS**]).

Of course, to implement this strategy, one has to define what a "simple" coloring actually is, and this is usually the most interesting part of the proof. In the context of the Hales–Jewett theorem, the definition of the proper concept of "simplicity" was undoubtedly a significant conceptual breakthrough in the work of Shelah and has proven to be very influential. This is the notion of an *insensitive coloring* which we are about to introduce.

2.1.1. Insensitive sets and insensitive colorings. Let A be a finite alphabet with $|A| \geqslant 2$ and $a, b \in A$ with $a \neq b$. Also let z, y be two words over A. We say that z and y are (a,b)-*equivalent* provided that: (i) $|z| = |y|$, and (ii) if $z = (z_0, \dots, z_{n-1})$ and $y = (y_0, \dots, y_{n-1})$ with $n \geqslant 1$, then for every $i \in \{0, \dots, n-1\}$ and every $\gamma \in A \setminus \{a, b\}$ we have

$$z_i = \gamma \text{ if and only if } y_i = \gamma.$$

That is, the words z and y are (a,b)-*equivalent* if they have the same length and possibly differ only in the coordinates taking values in $\{a, b\}$.

Definition 2.2. *Let A be a finite alphabet with $|A| \geqslant 2$ and $a, b \in A$ with $a \neq b$. Also let S be a set of words over A.*

(a) *We say that S is* (a,b)-insensitive *provided that for every $z \in S$ and every $y \in A^{<\mathbb{N}}$ if z and y are (a,b)-equivalent, then $y \in S$.*
(b) *We say that S is* (a,b)-insensitive in a combinatorial space *W of $A^{<\mathbb{N}}$ if $\mathrm{I}_W^{-1}(S \cap W)$ is an (a,b)-insensitive subset of $A^{<\mathbb{N}}$ where I_W is the canonical isomorphism associated with W (see Definition 1.2).*

Notice that the family of all (a,b)-insensitive subsets of $A^{<\mathbb{N}}$ is an algebra of sets. In particular, it is closed under intersections, unions and complements. The same remark applies to the family of all (a,b)-insensitive sets in a combinatorial space W of $A^{<\mathbb{N}}$.

The notion of an insensitive set is naturally extended to colorings as follows.

Definition 2.3. *Let A be a finite alphabet with $|A| \geqslant 2$ and $a, b \in A$ with $a \neq b$. Also let r be a positive integer, W a combinatorial space of $A^{<\mathbb{N}}$ and c an r-coloring of W. We say that the coloring c is* (a,b)-insensitive in *W if for every $p \in [r]$ the set $c^{-1}(\{p\})$ is (a,b)-insensitive in W.*

Notice that if W is a combinatorial space of $A^{<\mathbb{N}}$ of dimension d and c is an r-coloring of W, then for every $z \in A^d$ and every $p \in [r]$ we have $c\big(\mathrm{I}_W(z)\big) = p$ if and only if $z \in \mathrm{I}_W^{-1}\big(c^{-1}(\{p\}) \cap W\big)$. Using this observation we obtain the following characterization of insensitive colorings.

Fact 2.4. *Let A be a finite alphabet with $|A| \geqslant 2$ and $a, b \in A$ with $a \neq b$. Also let W be a combinatorial space of $A^{<\mathbb{N}}$ of dimension d and c a finite coloring of W. Then the coloring c is (a,b)-insensitive in W if and only if $c\big(\mathrm{I}_W(z)\big) = c\big(\mathrm{I}_W(y)\big)$ for every $z, y \in A^d$ which are (a,b)-equivalent.*

2.1.2. Shelah's insensitivity lemma. We are now in a position to state the key result in Shelah's proof of the Hales–Jewett theorem.

LEMMA 2.5 (Shelah's insensitivity lemma). *For every triple k, d, r of positive integers there exists a positive integer N with the following property. If $n \geqslant N$, then for every alphabet A with $|A| = k+1$, every $a, b \in A$ with $a \neq b$ and every coloring $c\colon A^n \to [r]$ there exists a d-dimensional combinatorial subspace W of A^n such that the coloring c is (a,b)-insensitive in W. The least positive integer with this property will be denoted by $\mathrm{Sh}(k,d,r)$.*

Moreover, the numbers $\mathrm{Sh}(k,d,r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^4$.

The first step towards the proof of Lemma 2.5 is an extension of Fact 2.4. It will enable us to reduce Lemma 2.5 to the construction of a d-dimensional combinatorial subspace W on which the coloring c satisfies a property seemingly weaker than insensitivity.

SUBLEMMA 2.6. *Let A be an alphabet with $|A| \geqslant 2$ and $a, b \in A$ with $a \neq b$. Also let W be a combinatorial space of $A^{<\mathbb{N}}$ of dimension d and c a finite coloring of W. Then the coloring c is (a,b)-insensitive in W if and only if*

$$c\big(\mathrm{I}_W(v^\frown a^\frown u)\big) = c\big(\mathrm{I}_W(v^\frown b^\frown u)\big)$$

for every $i \in \{0, \dots, d-1\}$, every $v \in A^i$ and every $u \in A^{d-i-1}$.

PROOF. First we notice that for every $i \in \{0, \dots, d-1\}$, every $v \in A^i$ and every $u \in A^{d-i-1}$ the words $v^\frown a^\frown u$ and $v^\frown b^\frown u$ are (a,b)-equivalent. Therefore the "only if" part follows readily by Fact 2.4.

Conversely, let $z, y \in A^d$ and assume that z and y are (a,b)-equivalent. By Fact 2.4, it is enough to show that $c\big(\mathrm{I}_W(z)\big) = c\big(\mathrm{I}_W(y)\big)$. To this end, let ℓ be the number of coordinates where z and y differ. We select a finite sequence $(z_0, \dots, z_\ell)$ in A^d such that: (i) $z_0 = z$, (ii) $z_\ell = y$, and (iii) for every $i \in [\ell-1]$ the words z_{i-1} and z_i differ in exactly one coordinate at which one of them takes the value a and the other the value b. Invoking our assumption we see that $c\big(\mathrm{I}_W(z_0)\big) = \dots = c\big(\mathrm{I}_W(z_\ell)\big)$, and since $z_0 = z$ and $z_\ell = y$, we conclude that $c\big(\mathrm{I}_W(z)\big) = c\big(\mathrm{I}_W(y)\big)$. The proof of Sublemma 2.6 is completed. □

It is convenient to introduce the following notation. For every nonempty alphabet A, every $a \in A$ and every positive integer i we set

$$a^i = (\underbrace{a, \dots, a}_{i-\text{times}}). \tag{2.1}$$

Also let a^0 denote the empty word.

The next result is the combinatorial core of Shelah's insensitivity lemma and deals with the first nontrivial case, namely when "$d = 1$".

SUBLEMMA 2.7. *For every pair k, r of positive integers we have $\mathrm{Sh}(k,1,r) = r$.*

Proof. First we observe that $\mathrm{Sh}(k,1,r) \geqslant r$. This is obvious if $r=1$, and so, we may assume that $r \geqslant 2$. Let A be a finite alphabet with $|A| \geqslant 2$ and fix $a,b \in A$ with $a \neq b$. Define $c\colon A^{r-1} \to \{0,\dots,r-1\}$ by

$$c\big((a_0,\dots,a_{r-2})\big) = |\{i \in \{0,\dots,r-2\} : a_i = a\}|$$

and notice that $c\big(w(a)\big) \neq c\big(w(b)\big)$ for every variable word w over A of length $r-1$. This implies, of course, that $\mathrm{Sh}(k,1,r) \geqslant r$.

We proceed to show that $\mathrm{Sh}(k,1,r) \leqslant r$. Fix an integer $n \geqslant r$, an alphabet A with $|A| = k+1$ and a coloring $c\colon A^n \to [r]$. We also fix $a,b \in A$ with $a \neq b$. We need to find a variable word w over A of length n such that $c\big(w(a)\big) = c\big(w(b)\big)$. To this end we define

$$D(a,b,n) = \{a^{i}{}^\frown b^{n-i} : 0 \leqslant i \leqslant n\} \subseteq A^n. \tag{2.2}$$

The set $D(a,b,n)$ satisfies the following crucial property: every pair of distinct elements of $D(a,b,n)$ forms a combinatorial line over the alphabet $\{a,b\}$ of length n. Indeed, let $z,y \in D(a,b,n)$ with $z \neq y$ and write $z = a^{i}{}^\frown b^{n-i}$ and $y = a^{j}{}^\frown b^{n-j}$ where $0 \leqslant i < j \leqslant n$. Setting $w = a^{i}{}^\frown x^{j-i}{}^\frown b^{n-j}$, we see that w is a variable word over $\{a,b\}$ of length n such that $w(a) = y$ and $w(b) = z$.

Now observe that $|D(a,b,n)| = n+1 > r$. Therefore, by the classical pigeonhole principle, there exist $z_1, z_2 \in D(a,b,n)$ with $z_1 \neq z_2$ and such that $c(z_1) = c(z_2)$. By the previous discussion, there exists a variable word w over $\{a,b\}$ of length n such that $\{w(a), w(b)\} = \{z_1, z_2\}$. Clearly, w is as desired. The proof of Sublemma 2.7 is completed. $\square$

We need to introduce some numerical invariants. Specifically, let $f\colon \mathbb{N}^5 \to \mathbb{N}$ be defined by

$$f(k,d,r,i,n) = \begin{cases} r^{(k+1)^{n+d-i-1}} & \text{if } n+d-i-1 \geqslant 0,\\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

and define $g\colon \mathbb{N}^4 \to \mathbb{N}$ recursively by the rule

$$\begin{cases} g(k,d,r,0) = 0,\\ g(k,d,r,i+1) = g(k,d,r,i) + f\big(k,d,r,i,g(k,d,r,i)\big). \end{cases} \tag{2.4}$$

Finally, we define $\phi\colon \mathbb{N}^3 \to \mathbb{N}$ by

$$\phi(k,d,r) = g(k,d,r,d). \tag{2.5}$$

The function f has double exponential growth and so it is majorized by a function belonging to the class $\mathcal{E}^3$. It follows that both g and ϕ are upper bounded by primitive recursive functions belonging to the class $\mathcal{E}^4$. Moreover, notice that

$$\phi(k,1,r) = r. \tag{2.6}$$

We are ready to give the proof of Lemma 2.5.

Proof of Lemma 2.5. We fix a pair k,r of positive integers. It is enough to show that for every positive integer d we have

$$\mathrm{Sh}(k,d,r) \leqslant \phi(k,d,r). \tag{2.7}$$

Notice, first, that the case "$d = 1$" follows from Sublemma 2.7 and (2.6). So assume that $d \geqslant 2$. For every $i \in \{0, \dots, d\}$ we set $N_i = g(k, d, r, i)$. Observe that $N_0 = 0$, $N_d = g(k, d, r, d) = \phi(k, d, r)$ and

$$N_{i+1} = N_i + f(k, d, r, i, N_i) = N_i + r^{(k+1)^{N_i+d-i-1}} \tag{2.8}$$

for every $i \in \{0, \dots, d-1\}$. In particular, the estimate in (2.7) will follow once we show that $\mathrm{Sh}(k, d, r) \leqslant N_d$.

To this end, let $n \geqslant N_d$ and A an alphabet with $|A| = k+1$. Also let $c \colon A^n \to [r]$ be a coloring and fix $a, b \in A$ with $a \neq b$. First we claim that we may assume that $n = N_d$. Indeed, we select an element $z_0 \in A^{n-N_d}$ and we define $c' \colon A^{N_d} \to [r]$ by the rule $c'(z) = c(z^\frown z_0)$ for every $z \in A^{N_d}$. If W is a d-dimensional combinatorial subspace of A^{N_d} such that the coloring c' is (a, b)-insensitive in W, then so is the coloring c in the d-dimensional combinatorial subspace $W^\frown z_0$ of A^n.

The desired d-dimensional combinatorial subspace W of A^{N_d} will be generated by a sequence $(w_i)_{i=0}^{d-1}$ of variable words over A (actually over the smaller alphabet $\{a, b\}$). This sequence of variable words will be selected by backwards induction subject to the following conditions.

(C1) For every $i \in \{0, \dots, d-1\}$ the length of w_i is n_i where

$$n_i = N_{i+1} - N_i = r^{(k+1)^{N_i+d-i-1}}. \tag{2.9}$$

(C2) For every $z \in A^{N_{d-1}}$ we have $c\big(z^\frown w_{d-1}(a)\big) = c\big(z^\frown w_{d-1}(b)\big)$.

(C3) For every $i \in \{0, \dots, d-2\}$, every $z \in A^{N_i}$ and every $y \in A^{d-i-1}$ we have

$$c\big(z^\frown w_i(a)^\frown \mathrm{I}_{W_{i+1}}(y)\big) = c\big(z^\frown w_i(b)^\frown \mathrm{I}_{W_{i+1}}(y)\big)$$

where W_{i+1} is the combinatorial subspace of $A^{N_d - N_{i+1}}$ generated by the finite sequence $(w_j)_{j=i+1}^{d-1}$ via formula (1.16) and $\mathrm{I}_{W_{i+1}}$ is the canonical isomorphism associated with W_{i+1}.

The first step is identical to the general one, and so let $i \in \{0, \dots, d-2\}$ and assume that the variable words $w_{i+1}, \dots, w_{d-1}$ have been selected so that the above conditions are satisfied. By (2.9), we may identify $[r]^{A^{N_i} \times A^{d-i-1}}$ with $[n_i]$. We define a coloring $C \colon A^{n_i} \to [n_i]$ by the rule

$$C(u) = \big\langle c\big(z^\frown u^\frown \mathrm{I}_{W_{i+1}}(y)\big) : (z, y) \in A^{N_i} \times A^{d-i-1} \big\rangle$$

for every $u \in A^{n_i}$. (For the first step we set $C(u) = \langle c(z^\frown u) : z \in A^{N_{d-1}} \rangle$.) By Sublemma 2.7 applied to the coloring C, there exists a variable word w over A of length n_i such that $C\big(w(a)\big) = C\big(w(b)\big)$. We set $w_i = w$ and we observe that with this choice the above conditions are satisfied. The selection of the sequence $(w_i)_{i=0}^{d-1}$ is thus completed.

It remains to check that the coloring c is (a, b)-insensitive in the combinatorial subspace W of A^{N_d} generated by the sequence $(w_i)_{i=0}^{d-1}$. Indeed, by conditions (C2) and (C3), we see that $c\big(\mathrm{I}_W(v^\frown a^\frown u)\big) = c\big(\mathrm{I}_W(v^\frown b^\frown u)\big)$ for every $i \in \{0, \dots, d-1\}$, every $v \in A^i$ and every $u \in A^{d-i-1}$. By Sublemma 2.6, we conclude that the coloring c is (a, b)-insensitive in W and the proof of Lemma 2.5 is completed. $\square$

2.1.3. Proof of Theorem 2.1. As we have already mentioned, the proof proceeds by induction on k. The initial case "$k = 2$" follows from Sublemma 2.7. Indeed, for every positive integer r we have $\mathrm{HJ}(2, r) = \mathrm{Sh}(1, 1, r)$ and so

$$\mathrm{HJ}(2, r) = r. \tag{2.10}$$

Next, let $k \geqslant 2$ and assume that the result has been proved up to k. The following lemma is the second main step of the proof of Theorem 2.1.

Lemma 2.8. *Let k, r be positive integers with $k \geqslant 2$ and assume that the number $\mathrm{HJ}(k, r)$ has been estimated. Also let A be an alphabet with $|A| = k+1$ and $a, b \in A$ with $a \neq b$. Finally, let n, d be positive integers with $n \geqslant d \geqslant \mathrm{HJ}(k, r)$, c an r-coloring of A^n and W a d-dimensional combinatorial subspace of A^n. If the coloring c is (a, b)-insensitive in W, then there exists a combinatorial line of A^n which is monochromatic with respect to c.*

Proof. Set $B = A \setminus \{b\}$ and define $c' \colon B^d \to [r]$ by $c'(z) = c\big(\mathrm{I}_W(z)\big)$ where I_W is the canonical isomorphism associated with W (see Definition 1.2). Since $|B| = k$ and $d \geqslant \mathrm{HJ}(k, r)$, there exists a variable word w over B of length d such that the combinatorial line $\{w(\beta) : \beta \in B\}$ of B^d is monochromatic with respect to c'. Therefore, the set

$$\big\{\mathrm{I}_W\big(w(\beta)\big) : \beta \in B\big\} \tag{2.11}$$

is contained in W and is monochromatic with respect to c. Next observe that $w(a)$ and $w(b)$ are (a, b)-equivalent words of A^d and recall that the coloring c is (a, b)-insensitive in W. By Fact 2.4, we obtain that $c\big(\mathrm{I}_W(w(a))\big) = c\big(\mathrm{I}_W(w(b))\big)$. It follows from the previous discussion that the set

$$U = \big\{\mathrm{I}_W\big(w(\beta)\big) : \beta \in B\big\} \cup \big\{\mathrm{I}_W\big(w(b)\big)\big\} \tag{2.12}$$

is a combinatorial line of A^n which is monochromatic with respect to c. The proof of Lemma 2.8 is completed. $\square$

We are now ready to estimate the numbers $\mathrm{HJ}(k+1, r)$. Specifically, by Lemmas 2.5 and 2.8, we see that

$$\mathrm{HJ}(k+1, r) \leqslant \mathrm{Sh}\big(k, \mathrm{HJ}(k, r), r\big) \tag{2.13}$$

for every positive integer r. This completes, of course, the proof of the general inductive step.

Finally, the fact that the Hales–Jewett numbers are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^5$ is an immediate consequence of (2.10) and (2.13), and the upper bounds for the numbers $\mathrm{Sh}(k, d, r)$ obtained by Lemma 2.5. The proof of Theorem 2.1 is completed.

2.2. The multidimensional Hales–Jewett theorem

The following result is known as the *multidimensional Hales–Jewett theorem* and is a natural refinement of Theorem 2.1.

THEOREM 2.9. *For every triple k, d, r of positive integers with $k \geqslant 2$ there exists a positive integer N with the following property. If A is an alphabet with $|A| = k$, then for every combinatorial space W of $A^{<\mathbb{N}}$ of dimension at least N and every r-coloring of W there exists a monochromatic d-dimensional combinatorial subspace of W. The least positive integer with this property will be denoted by* $\mathrm{MHJ}(k, d, r)$.

Moreover, the numbers $\mathrm{MHJ}(k, d, r)$ *are upper bounded by a primitive recursive function belonging to the class* $\mathcal{E}^5$.

We will present two proofs of Theorem 2.9. The first proof is a modification of Shelah's proof of the Hales–Jewett theorem.

FIRST PROOF OF THEOREM 2.9. It is enough to observe that for every triple k, d, r of positive integers with $k \geqslant 2$ we have $\mathrm{MHJ}(2, d, r) = \mathrm{Sh}(1, d, r)$ and

$$\mathrm{MHJ}(k+1, d, r) \leqslant \mathrm{Sh}\big(k, \mathrm{MHJ}(k, d, r), r\big). \tag{2.14}$$

The estimate in (2.14) is, of course, the analogue of (2.13) and can be easily proved using Shelah's insensitivity lemma and arguing as in Lemma 2.8. The first proof of Theorem 2.9 is completed. □

The second proof of Theorem 2.9 is more general and relies on a direct application of the Hales–Jewett theorem. In particular, given an alphabet A, the idea is to use words over an appropriately selected finite Cartesian product A^d of A. We will see, later on, more applications of this technique.

We proceed to the details. Let A be a finite alphabet with $|A| \geqslant 2$ and let d, ℓ be positive integers. Notice that the finite alphabets $(A^d)^\ell$ and $A^{d\cdot\ell}$ have the same cardinality, and so, they can be identified in many ways. In the following definition we fix a convenient, for our purposes, identification.

DEFINITION 2.10. *Let A be a finite alphabet with $|A| \geqslant 2$ and d, ℓ positive integers. We define a map $\mathrm{T}\colon (A^d)^\ell \to A^{d\cdot\ell}$ as follows. Let $b_0, \dots, b_{\ell-1} \in A^d$ be arbitrary. For every $i \in \{0, \dots, d-1\}$ and every $j \in \{0, \dots, \ell-1\}$ denote by $b_{i,j}$ the i-th coordinate of b_j and define*

$$\mathrm{T}\big((b_0, \dots, b_{\ell-1})\big) = (b_{0,0}, \dots, b_{0,\ell-1})^\frown \dots ^\frown(b_{d-1,0}, \dots, b_{d-1,\ell-1}). \tag{2.15}$$

The next fact is straightforward.

FACT 2.11. *Let A, d, ℓ and T be as in Definition 2.10. If $m \in \{0, \dots, d\cdot\ell-1\}$, then let $i_m \in \{0, \dots, d-1\}$ and $j_m \in \{0, \dots, \ell-1\}$ be the unique integers such that $m = i_m \cdot \ell + j_m$. Finally, let $b_0, \dots, b_{\ell-1} \in A^d$ and $a_0, \dots, a_{d\cdot\ell-1} \in A$. Then $\mathrm{T}\big((b_0, \dots, b_{\ell-1})\big) = (a_0, \dots, a_{d\cdot\ell-1})$ if and only if a_m is the i_m-th coordinate of b_{j_m} for every $m \in \{0, \dots, d\cdot\ell-1\}$.*

The main property of the map T is described in the following lemma.

LEMMA 2.12. *Let A be a finite alphabet with $|A| \geqslant 2$ and d, ℓ positive integers, and set $B = A^d$ and $N = d\cdot\ell$. Let $\mathrm{T}\colon B^\ell \to A^N$ be as in Definition 2.10. Then the image under the map T of a combinatorial line of B^ℓ is a d-dimensional combinatorial subspace of A^N.*

Proof. Fix a combinatorial line L of B^ℓ. Let X be the wildcard set of L, S the set of its fixed coordinates and $(f_j)_{j\in S}\in B^S$ its constant part. For every $i\in\{0,\dots,d-1\}$ let $X_i=\{x+i\cdot\ell : x\in X\}$ and observe that $\max(X_i)<\min(X_{i+1})$. Moreover, for every $a\in A$ set

$$S_a=\{i\cdot\ell+j : j\in S \text{ and the } i\text{-th coordinate of } f_j \text{ is } a\}.$$

Now let $w=(w_0,\dots,w_{N-1})\in A^N$ be arbitrary. By Fact 2.11, we see that $w\in \mathrm{T}(L)$ if and only if: (i) for every $i\in\{0,\dots,d-1\}$ the located word $w\upharpoonright X_i$ is constant, and (ii) for every $a\in A$ and every $m\in S_a$ we have $w_m=a$. Therefore, the set $\mathrm{T}(L)$ is a d-dimensional combinatorial subspace of A^N with wildcard sets $X_0,\dots,X_{d-1}$. The proof of Lemma 2.12 is completed. □

We are ready to give the second proof of Theorem 2.9.

Second proof of Theorem 2.9. Fix a triple k,d,r of positive integers with $k\geqslant 2$. We will show that

$$\mathrm{MHJ}(k,d,r)\leqslant d\cdot \mathrm{HJ}(k^d,r). \tag{2.16}$$

To this end, set $\ell=\mathrm{HJ}(k^d,r)$ and $N=d\cdot\ell$. Also fix an alphabet A with $|A|=k$ and set $B=A^d$. Finally, let W be an N-dimensional combinatorial space of $A^{<\mathbb{N}}$ and $c\colon W\to[r]$ an r-coloring of W. We set $c'=c\circ \mathrm{I}_W\circ \mathrm{T}$ where $\mathrm{I}_W\colon A^N\to W$ is the canonical isomorphism associated with W (see Definition 1.2) and $\mathrm{T}\colon B^\ell\to A^N$ is as in Definition 2.10. Notice that c' is an r-coloring of B^ℓ. Since $|B|=k^d$ and $\ell=\mathrm{HJ}(k^d,r)$, there exists a combinatorial line L of B^ℓ which is monochromatic with respect to c'. We set $V=\mathrm{I}_W\big(\mathrm{T}(L)\big)$. By Lemma 2.12 and Fact 1.3, we see that V is a d-dimensional combinatorial subspace of W which is monochromatic with respect to c. This shows that the estimate in (2.16) is satisfied and the second proof of Theorem 2.9 is completed. □

We close this section with the following proposition which provides significantly better upper bounds for the multidimensional Hales–Jewett numbers when “$k=2$”.

Proposition 2.13. *For every pair d,r of positive integers we have*

$$\mathrm{MHJ}(2,d,r)\leqslant d\cdot r^{3^{d-1}}. \tag{2.17}$$

Proof. It is similar to the proof of Sublemma 2.7. The choice of the alphabet is irrelevant, and so we may assume that $A=\{0,1\}$. It is also convenient to introduce the following terminology. We say that a variable word w over $\{0,1\}$ is *simple* if there exist $i,j,k\in\mathbb{N}$ with $j\neq 0$ such that $w=0^i{}^\frown x^j{}^\frown 1^k$. Respectively, we say that a combinatorial line over $\{0,1\}$ is *simple* if it is generated by a simple variable word. Finally, for every $n\in\mathbb{N}$ let

$$D(n)=\{0^i{}^\frown 1^{n-i} : 0\leqslant i\leqslant n\}\subseteq\{0,1\}^n. \tag{2.18}$$

Notice that $|D(n)|=n+1$. Arguing as in the proof of Sublemma 2.7, it is easy to see that there exists a bijection between $\binom{D(n)}{2}$ and the set of all simple combinatorial lines of $\{0,1\}^n$. In particular, for every positive integer n there are exactly $\binom{n+1}{2}$ simple combinatorial lines of $\{0,1\}^n$.

CLAIM 2.14. *Let d, r be positive integers and define a sequence (n_i) in $\mathbb{N}$ recursively by the rule*

$$\begin{cases} n_0 = r, \\ n_{i+1} = r \prod_{j=0}^{i} \binom{n_j+1}{2}. \end{cases} \tag{2.19}$$

If $N_d = \sum_{i=0}^{d-1} n_i$, then for every coloring $c\colon \{0,1\}^{N_d} \to [r]$ there exists a finite sequence $(w_i)_{i=0}^{d-1}$ of simple variable words over $\{0,1\}$ such that: (i) *w_i has length n_i for every $i \in \{0, \dots, d-1\}$, and* (ii) *the combinatorial subspace of A^{N_d} generated by $(w_i)_{i=0}^{d-1}$ is monochromatic.*

PROOF OF CLAIM 2.14. By induction on d. The case "$d = 1$" follows from Sublemma 2.7. Let $d \geqslant 1$ and assume that the result has been proved up to d. Write $N_{d+1} = \sum_{i=0}^{d} n_i = N_d + n_d$ and let $c\colon \{0,1\}^{N_{d+1}} \to [r]$ be an arbitrary coloring. For every $s \in D(n_d)$ we define $c_s\colon \{0,1\}^{N_d} \to [r]$ by the rule $c_s(y) = c(y^\frown s)$ for every $y \in \{0,1\}^{N_d}$. Notice that the cardinality of the set of all finite sequences $(v_i)_{i=0}^{d-1}$ such that v_i is a simple variable word over $\{0,1\}$ of length n_i for every $i \in \{0, \dots, d-1\}$, is equal to

$$\binom{n_0+1}{2} \cdots \binom{n_{d-1}+1}{2} = \prod_{i=0}^{d-1} \binom{n_i+1}{2}.$$

Moreover,

$$|D(n_d)| = n_d + 1 \overset{(2.19)}{=} r \prod_{i=0}^{d-1} \binom{n_i+1}{2} + 1.$$

Taking into account the above remarks, applying our inductive assumption to each coloring in the family $\{c_s : s \in D(n_d)\}$ and using the classical pigeonhole principle, we select $s, t \in D(n_d)$ with $s \neq t$ and a sequence $(v_0, \dots, v_{d-1})$ of simple variable words over $\{0,1\}$ such that v_i has length n_i for every $i \in \{0, \dots, d-1\}$ and satisfying the following property. If V is the d-dimensional combinatorial subspace of $\{0,1\}^{N_d}$ generated by the sequence $(v_0, \dots, v_{d-1})$, then $c(y^\frown s) = c(z^\frown t)$ for every $y, z \in V$. Finally, let v_d be the unique simple variable word over $\{0,1\}$ of length n_d such that $\{v_d(0), v_d(1)\} = \{s, t\}$. Clearly, the finite sequence $(v_0, \dots, v_{d-1}, v_d)$ is as desired. The proof of Claim 2.14 is completed. $\square$

Fix a pair d, r of positive integers. By (2.19), we have $n_{i+1} = n_i \binom{n_i+1}{2} \leqslant n_i^3$ and so $n_i \leqslant r^{3^i}$ for every $i \in \mathbb{N}$. Therefore, by Claim 2.14, we conclude that

$$\mathrm{MHJ}(2, d, r) \leqslant \sum_{i=0}^{d-1} n_i \leqslant \sum_{i=0}^{d-1} r^{3^i} \leqslant d \cdot r^{3^{d-1}}$$

and the proof of Proposition 2.13 is completed. $\square$

2.3. Colorings of combinatorial spaces

So far we have been dealing with colorings of points of combinatorial spaces. We will now change our perspective and we will consider colorings of combinatorial spaces of a fixed dimension. Specifically, this section is devoted to the proof of the following result which is a significant extension of the Hales–Jewett theorem.

THEOREM 2.15. *For every quadruple k, d, m, r of positive integers with $k \geqslant 2$ and $d \geqslant m$ there exists a positive integer N with the following property. If $n \geqslant N$ and A is an alphabet with $|A| = k$, then for every n-variable word w over A and every r-coloring of $\mathrm{Subw}_m(w)$ there exists $v \in \mathrm{Subw}_d(w)$ such that the set $\mathrm{Subw}_m(v)$ is monochromatic. The least positive integer with this property will be denoted by $\mathrm{GR}(k, d, m, r)$.*

Moreover, the numbers $\mathrm{GR}(k, d, m, r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^6$.

We remark that Theorem 2.15 is a variant[2] of the Graham–Rothschild theorem [**GR**] which refers to m-parameter words instead of m-variable words. We will not use the Graham–Rothschild theorem in this book, and so our choice of the acronym "GR" will not cause confusion. However, the reader should have in mind that these two results refer to different types of structures.

Also note that, taking into account the correspondence between m-variable words and m-dimensional combinatorial spaces (see Fact 1.4), Theorem 2.15 is equivalently formulated as follows.

THEOREM 2.15′. *Let k, d, m, r be positive integers with $k \geqslant 2$ and $d \geqslant m$. If A is an alphabet with $|A| = k$, then for every combinatorial space W of $A^{<\mathbb{N}}$ of dimension at least $\mathrm{GR}(k, d, m, r)$ and every r-coloring of $\mathrm{Subsp}_m(W)$ there exists $V \in \mathrm{Subsp}_d(W)$ such that the set $\mathrm{Subsp}_m(V)$ is monochromatic.*

The proof of Theorem 2.15 will be given in Subsection 2.3.3. It follows the general scheme we discussed in Section 2.1. More precisely, given a finite coloring c of $\mathrm{Subw}_m(w)$, the strategy is to find $u \in \mathrm{Subw}_\ell(w)$, where ℓ is sufficiently large, such that for every $v \in \mathrm{Subw}_m(u)$ the color $c(v)$ of v depends only on the position of its variables. In this way, Theorem 2.15 is effectively reduced to a simpler statement which is a finite version of the Milliken–Taylor theorem [**M1, Tay1**]. The finite version of the Milliken–Taylor theorem, as well as some related results of independent interest, are presented in Subsections 2.3.1 and 2.3.2.

2.3.1. The disjoint unions theorem. A *disjoint sequence* is a nonempty finite sequence $\mathcal{F} = (F_0, \dots, F_{n-1})$ of nonempty finite subsets of $\mathbb{N}$ with the property that $F_i \cap F_j = \emptyset$ for every $i, j \in \{0, \dots, n-1\}$ with $i \neq j$. A disjoint sequence $\mathcal{F} = (F_0, \dots, F_{n-1})$ is said to be *block* if $\max(F_i) < \min(F_j)$ for every $i, j \in \{0, \dots, n-1\}$ with $i < j$. The *set of nonempty unions* of a disjoint sequence $\mathcal{F} = (F_0, \dots, F_{n-1})$ is defined by

$$\mathrm{NU}(\mathcal{F}) = \Big\{ \bigcup_{s \in S} F_s : S \text{ is a nonempty subset of } \{0, \dots, n-1\} \Big\}. \tag{2.20}$$

[2]In fact, not only are Theorem 2.15 and the Graham–Rothschild theorem similar statements but also the corresponding known bounds are of the same order of magnitude. Specifically, Shelah has shown (see [**Sh1**, page 687]) that the original Graham–Rothschild numbers are also upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^6$.

The following result is known as the *disjoint unions theorem*[3] and appeared first in [**GR**]. The corresponding primitive recursive bounds are taken from [**Tay2**].

THEOREM 2.16. *For every pair d,r of positive integers there exists a positive integer N with the following property. If $n \geqslant N$, then for every disjoint sequence $\mathcal{F} = (F_0,\dots,F_{n-1})$ and every r-coloring of $\mathrm{NU}(\mathcal{F})$ there exists a disjoint sequence $\mathcal{G} = (G_0,\dots,G_{d-1})$ in $\mathrm{NU}(\mathcal{F})$ such that the set $\mathrm{NU}(\mathcal{G})$ is monochromatic. The least positive integer with this property will be denoted by $\mathrm{T}(d,r)$.*

Moreover, the numbers $\mathrm{T}(d,r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^4$.

The proof of the disjoint unions theorem is based on the following lemma which is an appropriate interpretation of the multidimensional Hales–Jewett theorem for the alphabet $A = \{0,1\}$.

LEMMA 2.17. *Let n,r be positive integers and set $N = \mathrm{MHJ}(2,n+1,r)$. Also let $\mathcal{F} = (F_0,\dots,F_{N-1})$ be a disjoint sequence and $c\colon \mathrm{NU}(\mathcal{F}) \to [r]$ a coloring. Then there exist a set $E \in \mathrm{NU}(\mathcal{F})$ and a disjoint sequence $\mathcal{G} = (G_0,\dots,G_{n-1})$ in $\mathrm{NU}(\mathcal{F})$ with $E \cap \big(\bigcup_{i=0}^{n-1} G_i\big) = \emptyset$ and such that the set $\{E\} \cup \{E \cup H : H \in \mathrm{NU}(\mathcal{G})\}$ is monochromatic.*

PROOF. For every set X let $X^{(0)} = \emptyset$ and $X^{(1)} = X$. We identify $\{0,1\}^N$ with the set of all unions (not necessarily nonempty) of $\mathcal{F}$ via the map

$$\{0,1\}^N \ni (\varepsilon_0,\dots,\varepsilon_{N-1}) \mapsto \bigcup_{j=0}^{N-1} F_j^{(\varepsilon_j)}$$

and we extend the coloring c to an r-coloring of $\{0,1\}^N$. By the choice of N, there exists a combinatorial subspace V of $\{0,1\}^N$ of dimension $n+1$ which is monochromatic. Let $X_0,\dots,X_n$ be the wildcard sets of V, S the set of its fixed coordinates and $(f_j)_{j\in S} \in \{0,1\}^S$ its constant part. We set

$$E = \Big(\bigcup_{j\in S} F_j^{(f_j)}\Big) \cup \Big(\bigcup_{j\in X_0} F_j\Big).$$

Also for every $i \in \{0,\dots,n-1\}$ let

$$G_i = \bigcup_{j\in X_{i+1}} F_j.$$

It is easy to check that the set E and the disjoint sequence $\mathcal{G} = (G_0,\dots,G_{n-1})$ are as desired. The proof of Lemma 2.17 is completed. $\square$

We are ready to give the proof of Theorem 2.16.

PROOF OF THEOREM 2.16. We define $f\colon \mathbb{N}^2 \to \mathbb{N}$ recursively by the rule

$$\begin{cases} f(r,0) = 1, \\ f(r,i+1) = \mathrm{MHJ}\big(2, f(r,i)+1, r\big). \end{cases} \tag{2.21}$$

[3]The disjoint unions theorem has a number theoretic counterpart which is known as the *non-repeating sums theorem* and is attributed to Rado [**Rado1, Rado2**], Folkman (unpublished) and J. H. Sanders [**Sa1**].

By Proposition 2.13, we see that f is upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^4$. Hence, the proof will be completed once we show that

$$\mathrm{T}(d,r) \leqslant f(r,(d-1)\cdot r+1) \tag{2.22}$$

for every pair d,r of positive integers. To this end we need the following claim.

Claim 2.18. *Let $r,\ell \in \mathbb{N}$ with $r \geqslant 1$ and $\ell \geqslant 2$, and set $N = f(r,\ell)$. Also let $\mathcal{F} = (F_0,\dots,F_{N-1})$ be a disjoint sequence and $c\colon \mathrm{NU}(\mathcal{F}) \to [r]$ a coloring. Then there exists a disjoint sequence $\mathcal{E} = (E_0,\dots,E_{\ell-1})$ in $\mathrm{NU}(\mathcal{F})$ such that for every $i \in \{0,\dots,\ell-2\}$ the set $\{E_i\} \cup \big\{E_i \cup H : H \in \mathrm{NU}\big((E_{i+1},\dots,E_{\ell-1})\big)\big\}$ is monochromatic.*

Granting Claim 2.18, the estimate in (2.22) follows from the classical pigeonhole principle.

It remains to prove Claim 2.18. Fix the parameters r and ℓ and for every $i \in \{0,\dots,\ell-1\}$ set $n_i = f(r,i)$. By (2.21), we have $n_{i+1} = \mathrm{MHJ}(2,n_i+1,r)$ for every $i \in \{0,\dots,\ell-2\}$. Hence, by Lemma 2.17 and backwards induction, we may select a disjoint sequence $(E_i)_{i=0}^{\ell-1}$ in $\mathrm{NU}(\mathcal{F})$ and a finite sequence $(\mathcal{G}_i)_{i=0}^{\ell}$ of disjoint sequences with $\mathcal{G}_\ell = \mathcal{F}$ and satisfying the following conditions for every $i \in [\ell-1]$.

(C1) The set E_i belongs to $\mathrm{NU}(\mathcal{G}_{i+1})$. Moreover, the sequence $\mathcal{G}_i$ is a disjoint sequence in $\mathrm{NU}(\mathcal{G}_{i+1})$ of length n_i such that $E_i \cap \big(\bigcup \mathcal{G}_i\big) = \emptyset$.

(C2) The set $\{E_i\} \cup \{E_i \cup H : H \in \mathrm{NU}(\mathcal{G}_i)\}$ is monochromatic.

We set $\mathcal{E} = (E_0,\dots,E_{\ell-1})$. Using conditions (C1) and (C2), we see that $\mathcal{E}$ is as desired. This completes the proof of Claim 2.18, and as we have indicated, the proof of Theorem 2.16 is also completed. □

The second result in this subsection is the following variant of Theorem 2.16. It refers to block, instead of disjoint, sequences and is the finite analogue of Hindman's theorem [**H**].

Proposition 2.19. *For every pair d,r of positive integers there exists a positive integer N with the following property. If $n \geqslant N$, then for every block sequence $\mathcal{F} = (F_0,\dots,F_{n-1})$ and every r-coloring of $\mathrm{NU}(\mathcal{F})$ there exists a block sequence $\mathcal{G} = (G_0,\dots,G_{d-1})$ in $\mathrm{NU}(\mathcal{F})$ such that the set $\mathrm{NU}(\mathcal{G})$ is monochromatic. The least positive integer with this property will be denoted by $\mathrm{H}(d,r)$.*

Moreover, the numbers $\mathrm{H}(d,r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^4$.

As in Appendix B, for every triple d,m,r of positive integers by $\mathrm{R}(d,m,r)$ we denote the corresponding Ramsey number. Recall that $\mathrm{R}(d,m,r)$ is defined to be the least integer $n \geqslant d$ such that for every n-element set X and every r-coloring of $\binom{X}{m}$ there exists $Z \in \binom{X}{d}$ such that the set $\binom{Z}{m}$ is monochromatic. For the proof of Proposition 2.19 we need the following consequence of Ramsey's theorem.

Fact 2.20. *Let m,r be positive integers. Also let X be a set with*

$$|X| \geqslant \mathrm{R}(2m,m,r^m) \tag{2.23}$$

and $c\colon \{F \subseteq X : 1 \leqslant |F| \leqslant m\} \to [r]$ *an* r*-coloring of the family of all nonempty subsets of* X *of cardinality at most* m*. Then there exists* $Y \in \binom{X}{m}$ *such that for every* $i \in [m]$ *the set* $\binom{Y}{i}$ *is monochromatic.*

Proof. Clearly we may assume that $X \subseteq \mathbb{N}$. For every nonempty subset F of X and every $i \in \{1, \dots, |F|\}$ we denote by $F \upharpoonright i$ the set of the first i elements of F. We define $C\colon \binom{X}{m} \to [r]^m$ by the rule $C(F) = \langle c(F \upharpoonright i) : i \in [m]\rangle$. By (2.23), there exists a subset Z of X with $|Z| = 2m$ such that the set $\binom{Z}{m}$ is monochromatic with respect to C. We set $Y = Z \upharpoonright m$. It is easy to see that Y is as desired. The proof of Fact 2.20 is completed. $\square$

We proceed to the proof of Proposition 2.19.

Proof of Proposition 2.19. By Theorems 2.16 and B.1, it is enough to show that for every pair d, r of positive integers we have

$$\mathrm{H}(d,r) \leqslant \mathrm{R}\big(2\mathrm{T}(d,r), \mathrm{T}(d,r), r^{\mathrm{T}(d,r)}\big). \tag{2.24}$$

To this end, fix the parameters d, r and set $m = \mathrm{T}(d,r)$ and $N = \mathrm{R}(2m, m, r^m)$. Let $n \geqslant N$ and $\mathcal{F} = (F_0, \dots, F_{n-1})$ a block sequence. Also let $c\colon \mathrm{NU}(\mathcal{F}) \to [r]$ be a coloring. Notice that the map

$$\{S \subseteq \{0, \dots, n-1\} : 1 \leqslant |S| \leqslant m\} \ni S \mapsto \bigcup_{s \in S} F_s \in \mathrm{NU}(\mathcal{F})$$

is an injection. Hence, by Fact 2.20, there is a block sequence $\mathcal{E} = (E_0, \dots, E_{m-1})$ in $\mathrm{NU}(\mathcal{F})$ such that $c\big(\bigcup_{s\in S} E_s\big) = c\big(\bigcup_{t \in T} E_t\big)$ for every pair S, T of nonempty subsets of $\{0, \dots, m-1\}$ with $|S| = |T|$. By the choice of m, there is a disjoint sequence $\mathcal{H} = (H_0, \dots, H_{d-1})$ in $\mathrm{NU}(\mathcal{E})$ such that the set $\mathrm{NU}(\mathcal{H})$ is monochromatic. For every $j \in \{0, \dots, d-1\}$ let S_j be the unique subset of $\{0, \dots, m-1\}$ such that $H_j = \bigcup_{s \in S_j} E_s$. Notice that $(S_0, \dots, S_{d-1})$ is a disjoint sequence of nonempty subsets of $\{0, \dots, m-1\}$. Therefore, we may select a block sequence $(T_0, \dots, T_{d-1})$ such that $T_j \subseteq \{0, \dots, m-1\}$ and $|T_j| = |S_j|$ for every $j \in \{0, \dots, d-1\}$. We set $\mathcal{G} = (G_0, \dots, G_{d-1})$ where $G_j = \bigcup_{t \in T_j} E_t$ for every $j \in \{0, \dots, d-1\}$. Clearly $\mathcal{G}$ is as desired. The proof of Proposition 2.19 is thus completed. $\square$

2.3.2. The finite version of the Milliken–Taylor theorem. First we introduce some pieces of notation and some terminology. Given two block sequences $\mathcal{F} = (F_0, \dots, F_{n-1})$ and $\mathcal{G} = (G_0, \dots, G_{m-1})$, we say that $\mathcal{G}$ is a *block subsequence* of $\mathcal{F}$ if for every $i \in \{0, \dots, m-1\}$ we have $G_i \in \mathrm{NU}(\mathcal{F})$. For every block sequence $\mathcal{F}$ of length n and every integer $m \in [n]$ we denote by $\mathrm{Block}_m(\mathcal{F})$ the set of all block subsequences of $\mathcal{F}$ of length m. Also let $\mathcal{F} \upharpoonright m = (F_0, \dots, F_{m-1})$ and notice that $\mathcal{F} \upharpoonright m \in \mathrm{Block}_m(\mathcal{F})$. If $\mathcal{G}$ is a block subsequence of $\mathcal{F}$, then the *depth of* $\mathcal{G}$ *in* $\mathcal{F}$, denoted by $\mathrm{depth}_{\mathcal{F}}(\mathcal{G})$, is defined to be the least integer $d \in [n]$ such that $\mathcal{G}$ is a block subsequence of $\mathcal{F} \upharpoonright d$. Finally, for every $m \in [n]$ we define

$$\mathrm{Block}_m^{\max}(\mathcal{F}) = \{\mathcal{G} \in \mathrm{Block}_m(\mathcal{F}) : \mathrm{depth}_{\mathcal{F}}(\mathcal{G}) = n\}. \tag{2.25}$$

That is, $\mathrm{Block}_m^{\max}(\mathcal{F})$ is the set of all block subsequences $\mathcal{F}$ of length m and of maximal depth.

This subsection is devoted to the proof of the following theorem which is the finite version of the Milliken–Taylor theorem [**M1, Tay1**].

THEOREM 2.21. *For every triple d, m, r of positive integers with $d \geqslant m$ there exists a positive integer N with the following property. For every block sequence $\mathcal{F}$ of length at least N and every r-coloring of $\mathrm{Block}_m(\mathcal{F})$ there exists $\mathcal{G} \in \mathrm{Block}_d(\mathcal{F})$ such that the set $\mathrm{Block}_m(\mathcal{G})$ is monochromatic. The least positive integer with this property will be denoted by $\mathrm{MT}(d, m, r)$.*

Moreover, the numbers $\mathrm{MT}(d, m, r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^6$.

For the proof of Theorem 2.21 we need to do some preparatory work. Specifically, we define $h\colon \mathbb{N}^4 \to \mathbb{N}$ recursively by the rule

$$\begin{cases} h(\ell, m, r, 0) = 1, \\ h(\ell, m, r, i+1) = \mathrm{H}\big(h(\ell, m, r, i), r^{2^{\ell m}}\big) + 1 \end{cases} \tag{2.26}$$

where ℓ, m and r vary over all positive integers. If some of the parameters ℓ, m, r happens to be zero, then we set $h(\ell, m, r, i) = 0$. By Proposition 2.19, we see that the function h is upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^5$. The following lemma is the main step of the proof of Theorem 2.21.

LEMMA 2.22. *Let ℓ, m, N, r be positive integers with $\ell \geqslant m+1$ and such that*

$$N = m - 1 + h(\ell, m, r, \ell - m). \tag{2.27}$$

Also let $\mathcal{F} = (F_0, \dots, F_{N-1})$ be a block sequence and $c\colon \mathrm{Block}_{m+1}(\mathcal{F}) \to [r]$ a coloring. Then there exists $\mathcal{G} \in \mathrm{Block}_\ell(\mathcal{F})$ such that for every $\mathcal{X}, \mathcal{Y} \in \mathrm{Block}_{m+1}(\mathcal{G})$ with $\mathcal{X} \upharpoonright m = \mathcal{Y} \upharpoonright m$ we have $c(\mathcal{X}) = c(\mathcal{Y})$.

PROOF. For every $i \in \{m-1, \dots, \ell-1\}$ set $N_i = h(\ell, m, r, \ell - 1 - i)$ and observe that $1 = N_{\ell-1} \leqslant N_i \leqslant N_{m-1} = N - (m-1)$. Also notice that

$$N_{i-1} = \mathrm{H}\big(N_i, r^{2^{\ell m}}\big) + 1 \tag{2.28}$$

for every $i \in \{m, \dots, \ell-1\}$.

Recursively we will select two sequences $(\mathcal{G}_i)_{i=m-1}^{\ell-1}$ and $(\mathcal{H}_i)_{i=m-1}^{\ell-1}$ of block subsequences of $\mathcal{F}$ and a sequence $(E_i)_{i=m-1}^{\ell-1}$ in $\mathrm{NU}(\mathcal{F})$ such that the following conditions are satisfied for every $i \in \{m-1, \dots, \ell-1\}$.

(C1) We have $|\mathcal{G}_i| = i$, $|\mathcal{H}_i| = N_i - 1$ and

$$\max\Big(\bigcup \mathcal{G}_i\Big) < \min(E_i) \leqslant \max(E_i) < \min\Big(\bigcup \mathcal{H}_i\Big).$$

(C2) If $i \geqslant m$, then $\mathcal{G}_i = \mathcal{G}_{i-1}{}^\frown E_{i-1}$.

(C3) If $i \geqslant m$, then $E_i{}^\frown \mathcal{H}_i$ is a block subsequence of $\mathcal{H}_{i-1}$.

(C4) If $i \geqslant m$, then $c\big(\mathcal{Z}^\frown X\big) = c\big(\mathcal{Z}^\frown Y\big)$ for every $\mathcal{Z} \in \mathrm{Block}_m^{\max}(\mathcal{G}_i)$ and every $X, Y \in \mathrm{NU}\big(E_i{}^\frown \mathcal{H}_i\big)$.

For the first step of the recursive selection we set $\mathcal{G}_{m-1} = \mathcal{F} \upharpoonright (m-1)$, $E_{m-1} = F_{m-1}$ and $\mathcal{H}_{m-1} = (F_m, \dots, F_{N-1})$. By (2.27), we see that with these choices condition (C1) is satisfied. The other conditions are superfluous in this case, and so the first step of the selection is completed.

Let $i \in \{m, \dots, \ell-1\}$ and assume that the block sequences $\mathcal{G}_{m-1}, \dots, \mathcal{G}_{i-1}$ and $\mathcal{H}_{m-1}, \dots, \mathcal{H}_{i-1}$ as well as the sets $E_{m-1}, \dots, E_{i-1}$ have been selected. First we set $\mathcal{G}_i = \mathcal{G}_{i-1}{}^\frown E_{i-1}$. For notational convenience let $\mathcal{B} = \mathrm{Block}_m^{\max}(\mathcal{G}_i)$. We define a coloring $C\colon \mathrm{NU}(\mathcal{H}_{i-1}) \to [r]^{\mathcal{B}}$ by the rule $C(X) = \langle c(\mathcal{Z}^\frown X) : \mathcal{Z} \in \mathcal{B}\rangle$. Notice that $|\mathcal{B}| < 2^{im} < 2^{\ell m}$. Moreover, by our inductive assumptions, we have that $|\mathcal{H}_{i-1}| = N_{i-1} - 1$ and so, by (2.28), the length of $\mathcal{H}_{i-1}$ is $\mathrm{H}\big(N_i, r^{2^{\ell m}}\big)$. By Proposition 2.19, there exists a block subsequence $\mathcal{V} = (V_0, \dots, V_{N_i-1})$ of $\mathcal{H}_{i-1}$ such that $\mathrm{NU}(\mathcal{V})$ is monochromatic with respect to the coloring C. We set $E_i = V_0$ and $\mathcal{H}_i = (V_1, \dots, V_{N_i-1})$. It is easy to check that with these choices conditions (C1) up to (C4) are fulfilled. The recursive selection is thus completed.

We set $\mathcal{G} = \mathcal{G}_{\ell-1}$ and we claim that $\mathcal{G}$ is as desired. Indeed, notice first that $\mathcal{G}$ is a block subsequence of $\mathcal{F}$ of length ℓ. Let $\mathcal{X}, \mathcal{Y} \in \mathrm{Block}_{m+1}(\mathcal{G})$ with $\mathcal{X} \upharpoonright m = \mathcal{Y} \upharpoonright m$ and write $\mathcal{X} = \mathcal{Z}^\frown X$ and $\mathcal{Y} = \mathcal{Z}^\frown Y$ where $\mathcal{Z} = \mathcal{X} \upharpoonright m = \mathcal{Y} \upharpoonright m$. There exists a unique $i \in \{m, \dots, \ell-1\}$ such that $\mathcal{Z} \in \mathrm{Block}_m^{\max}(\mathcal{G}_i)$ and $X, Y \in \mathrm{NU}(E_i{}^\frown \mathcal{H}_i)$. Therefore, by condition (C4), we conclude that $c(\mathcal{X}) = c(\mathcal{Z}^\frown X) = c(\mathcal{Z}^\frown Y) = c(\mathcal{Y})$. The proof of Lemma 2.22 is completed. □

We are ready to give the proof of Theorem 2.21.

Proof of Theorem 2.21. Notice first that

$$\mathrm{MT}(d, 1, r) = \mathrm{H}(d, r). \tag{2.29}$$

On the other hand, by Lemma 2.22, we see that

$$\mathrm{MT}(d, m+1, r) \leqslant m - 1 + h\big(\mathrm{MT}(d, m, r), m, r, \mathrm{MT}(d, m, r) - m\big) \tag{2.30}$$

for every triple d, m, r of positive integers with $d \geqslant m+1$.

Finally, recall that the function h is upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^5$. Therefore, by (2.29), (2.30) and Proposition 2.19, we see that the numbers $\mathrm{MT}(d, m, r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^6$. The proof of Theorem 2.21 is completed. □

2.3.3. Proof of Theorem 2.15.

We begin by introducing some numerical invariants. First, let $f\colon \mathbb{N}^6 \to \mathbb{N}$ be defined by

$$f(k, \ell, m, r, i, n) = \mathrm{HJ}\big(k, r^{(k+m)^{n+\ell-i-1}}\big) \tag{2.31}$$

if $k \geqslant 2$, $r \geqslant 1$ and $n + \ell - i - 1 \geqslant 0$; otherwise, we set $f(k, \ell, m, r, i, n) = 0$. Next, we define $g\colon \mathbb{N}^5 \to \mathbb{N}$ recursively by the rule

$$\begin{cases} g(k, \ell, m, r, 0) = 0, \\ g(k, \ell, m, r, i+1) = g(k, \ell, m, r, i) + f\big(k, \ell, m, r, i, g(k, \ell, m, r, i)\big). \end{cases} \tag{2.32}$$

By Theorem 2.1, we see that the function g is upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^6$.

We also need to introduce some pieces of notation. Specifically, let A be a finite alphabet with $|A| \geqslant 2$ and $n, m \in \mathbb{N}$ with $n \geqslant m \geqslant 1$. For every m-variable word $w = (w_0, \dots, w_{n-1})$ over A of length n and every $j \in \{0, \dots, m-1\}$ let

$$X_j = \big\{ i \in \{0, \dots, n-1\} : w_i = x_j \big\}. \tag{2.33}$$

That is, X_j is the set of coordinates where the j-th variable x_j appears in w. Set

$$\mathcal{X}(w) = (X_0, \dots, X_{m-1}) \tag{2.34}$$

and notice that $\mathcal{X}(w)$ is a block sequence.

As we have already pointed out, the strategy of the proof is to reduce Theorem 2.15 to Theorem 2.21. This is achieved with the following lemma.

Lemma 2.23. *Let k, ℓ, m, r be positive integers with $k \geqslant 2$ and $\ell \geqslant m$, and set*

$$N = g(k, \ell, m, r, \ell).$$

Then for every alphabet A with $|A| = k$ and every r-coloring c of the set of all m-variable words over A of length N there exists an ℓ-variable word w over A of length N such that $c(y) = c(z)$ for every $y, z \in \mathrm{Subw}_m(w)$ with $\mathcal{X}(y) = \mathcal{X}(z)$.

The following result is the analogue of Sublemma 2.6 and its proof is identical to that of Sublemma 2.6.

Sublemma 2.24. *Let ℓ, m be positive integers with $\ell \geqslant m$ and A a finite alphabet with $|A| \geqslant 2$. Also let w be an ℓ-variable word over A and c a finite coloring of $\mathrm{Subw}_m(w)$. Then the following are equivalent.*

(a) *We have $c(y) = c(z)$ for every $y, z \in \mathrm{Subw}_m(w)$ with $\mathcal{X}(y) = \mathcal{X}(z)$.*

(b) *For every m-variable word $(\alpha_0, \dots, \alpha_{\ell-1})$ over A, if $i \in \{0, \dots, \ell-1\}$ is such that $\alpha_i \in A$, then*

$$c\big(w(\alpha_0, \dots, \alpha_{i-1}, a, \alpha_{i+1}, \dots, \alpha_{\ell-1})\big) = c\big(w(\alpha_0, \dots, \alpha_{i-1}, b, \alpha_{i+1}, \dots, \alpha_{\ell-1})\big)$$

for every $a, b \in A$.

We proceed to the proof of Lemma 2.23.

Proof of Lemma 2.23. Clearly, we may assume that $\ell \geqslant m+1$. For every $i \in \{0, \dots, \ell\}$ set $N_i = g(k, \ell, m, r, i)$. Moreover, for every $i \in \{0, \dots, \ell-1\}$ let $M_i = N_i + \ell - i - 1$. Notice that $N_0 = 0$, $M_0 = \ell - 1$, $N_\ell = N$, $M_{\ell-1} = N_{\ell-1}$ and

$$N_{i+1} = N_i + \mathrm{HJ}\big(k, r^{(k+m)^{M_i}}\big) \tag{2.35}$$

for every $i \in \{0, \dots, \ell-1\}$.

Let A be an alphabet with $|A| = k$ and c an r-coloring of the set of all m-variable words over A of length N. By backwards induction, we will select a sequence $(w_i)_{i=0}^{\ell-1}$ of variable words over A such that the following conditions are satisfied.

(C1) For every $i \in \{0, \dots, \ell-1\}$ the variable word w_i has length n_i where

$$n_i = N_{i+1} - N_i \overset{(2.35)}{=} \mathrm{HJ}\big(k, r^{(k+m)^{M_i}}\big). \tag{2.36}$$

(C2) For every m-variable word v over A of length $M_{\ell-1}$ and every $a, b \in A$ we have $c\big(v^\frown w_{\ell-1}(a)\big) = c\big(v^\frown w_{\ell-1}(b)\big)$.

(C3) For every $i \in \{0, \dots, \ell-2\}$ and every m-variable word $v = (v_0, \dots, v_{M_i-1})$ over A of length M_i, setting

$$v^{(i)} = v \restriction N_i \ \text{ and } \ v^{(i+1)} = w_{i+1}(v_{N_i})^\frown w_{i+2}(v_{N_i+1})^\frown \dots ^\frown w_{\ell-1}(v_{M_i-1}), \tag{2.37}$$

we have $c\big(v^{(i)\frown} w_i(a)^\frown v^{(i+1)}\big) = c\big(v^{(i)\frown} w_i(b)^\frown v^{(i+1)}\big)$ for every $a, b \in A$.

The first step is identical to the general one, and so let $i \in \{0, \dots, \ell-2\}$ and assume that the variable words $w_{i+1}, \dots, w_{\ell-1}$ have been selected so that the above conditions are satisfied. For every m-variable word $v = (v_0, \dots, v_{M_i-1})$ over A of length M_i let $v^{(i)}$ and $v^{(i+1)}$ be as in (2.37) and observe that $v^{(i)\frown} z^\frown v^{(i+1)}$ is an m-variable word over A of length N for every $z \in A^{n_i}$. We define a coloring C of A^{n_i} by the rule

$$C(z) = \big\langle c(v^{(i)\frown} z^\frown v^{(i+1)}) : v \text{ is an } m\text{-variable word over } A \text{ of length } M_i \big\rangle.$$

(We set $C(z) = \langle c(v^\frown z) : v$ is an m-variable word over A of length $M_{\ell-1}\rangle$ for the case "$i = \ell-1$".) Notice that the number of all m-variable words over A of length M_i is less than $(k+m)^{M_i}$. Hence, by (2.36), there exists a variable word w over A of length n_i such that the combinatorial line $\{w(a) : a \in A\}$ of A^{n_i} is monochromatic with respect to C. We set $w_i = w$ and we observe that with this choice the above conditions are satisfied. The selection of the sequence $(w_i)_{i=0}^{d-1}$ is thus completed.

We set $w = w_0(x_0)^\frown \dots ^\frown w_{\ell-1}(x_{\ell-1})$. It is clear that w is an ℓ-variable word over A of length $N_\ell = N$. Moreover, by conditions (C1) and (C2), it satisfies the following property. For every $i \in \{0, \dots, \ell-1\}$ and every pair $(\alpha_0, \dots, \alpha_{\ell-1})$ and $(\beta_0, \dots, \beta_{\ell-1})$ of m-variable words over A of length ℓ, if $\alpha_j = \beta_j$ for every $j \in \{0, \dots, \ell-1\} \setminus \{i\}$ and $\alpha_i, \beta_i \in A$, then $c\big(w(\alpha_0, \dots, \alpha_{\ell-1})\big) = c\big(w(\beta_0, \dots, \beta_{\ell-1})\big)$. By Sublemma 2.24 and taking into account the previous remarks, we see that w is as desired. The proof of Lemma 2.23 is completed. □

We are now ready for the last step of the proof of Theorem 2.15. By Lemma 2.23 and Theorem 2.21, we see that

$$\mathrm{GR}(k,d,m,r) \leqslant g\big(k, \mathrm{MT}(d,m,r), m, r, \mathrm{MT}(d,m,r)\big). \tag{2.38}$$

Hence, by (2.38), Theorem 2.21 and the fact that g is dominated by a function belonging to the class $\mathcal{E}^6$, we conclude that the numbers $\mathrm{GR}(k,d,m,r)$ are upper bounded by a primitive recursive function also belonging to the class $\mathcal{E}^6$. The proof of Theorem 2.15 is completed.

2.3.4. Colorings of combinatorial lines. We close this section with the following result due to Tyros [**Ty**] which provides better upper bounds for the Graham–Rothschild numbers for the important special case "$m = 1$".

PROPOSITION 2.25. *There exists a primitive recursive function $\psi \colon \mathbb{N}^3 \to \mathbb{N}$ belonging to the class $\mathcal{E}^5$ such that*

$$\mathrm{GR}(k,d,1,r) \leqslant \psi(k,d,r) \tag{2.39}$$

for every triple k,d,r of positive integers with $k \geqslant 2$.

The proof of Proposition 2.25 is a modification of Shelah's proof of the Hales–Jewett theorem and relies on the following analogue of Definition 2.2 in the context of variable words.

DEFINITION 2.26. *Let A be a finite alphabet with $|A| \geqslant 2$ and $a, b \in A$ with $a \neq b$. Also let n, d be positive integers with $n \geqslant d$ and c a finite coloring of the set of all variable words over A of length n. Finally, let w be a d-variable word* *over A of length n. We say that the coloring c is (a,b)-*insensitive in $\mathrm{Subw}_1(w)$ *if $c(u)=c(v)$ for every $u,v\in \mathrm{Subw}_1(w)$ which are (a,b)-equivalent when viewed as words over the alphabet $A\cup\{x\}$.*

We have the following analogue of Shelah's insensitivity lemma.

LEMMA 2.27. *For every triple k,d,r of positive integers there exists a positive integer N with the following property. If $n\geqslant N$, then for every alphabet A with $|A|=k+1$, every $a,b\in A$ with $a\neq b$ and every r-coloring c of the set of all variable words over A of length n, there exists a d-variable word w over A of length n such that the coloring c is (a,b)-insensitive in* $\mathrm{Subw}_1(w)$. *The least positive integer with this property will be denoted by* $\mathrm{Sh_v}(k,d,r)$.

Moreover, the numbers $\mathrm{Sh_v}(k,d,r)$ *are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^4$.*

For the proof of Lemma 2.27 we need to introduce some slight variants of the functions f,g and ϕ defined in (2.3), (2.4) and (2.5) respectively. Specifically, we define $f'\colon \mathbb{N}^5\to\mathbb{N}$ by

$$f'(k,d,r,i,n)=\begin{cases} r^{(k+2)^{n+d-i-1}} & \text{if } n+d-i-1>0,\\ 1 & \text{otherwise.}\end{cases} \tag{2.40}$$

Also let $g'\colon \mathbb{N}^4\to\mathbb{N}$ be defined recursively by the rule

$$\begin{cases} g'(k,d,r,0)=0,\\ g'(k,d,r,i+1)=g'(k,d,r,i)+f'\big(k,d,r,i,g'(k,d,r,i)\big)\end{cases} \tag{2.41}$$

and define $\phi'\colon\mathbb{N}^3\to\mathbb{N}$ by setting

$$\phi'(k,d,r)=g'(k,d,r,d). \tag{2.42}$$

Notice that g' and ϕ' are both upper bounded by primitive recursive functions belonging to the class $\mathcal{E}^4$.

PROOF OF LEMMA 2.27. We will show that for every triple k,d,r of positive integers we have

$$\mathrm{Sh_v}(k,d,r)\leqslant\phi'(k,d,r). \tag{2.43}$$

Notice, first, that $\mathrm{Sh_v}(k,1,r)=1=\phi'(k,1,r)$, and so we may assume that $d\geqslant 2$. For every $i\in\{0,\dots,d\}$ let $N_i=g'(k,d,r,i)$ and $M_i=N_i+d-i-1$. Observe that $N_0=0$, $M_0=d-1$ and

$$N_{i+1}=N_i+r^{(k+2)^{M_i}} \tag{2.44}$$

for every $i\in\{0,\dots,d-1\}$. Therefore, it is enough to show that $\mathrm{Sh_v}(k,d,r)\leqslant N_d$. To this end let $n\geqslant N_d$, A an alphabet with $|A|=k+1$ and $a,b\in A$ with $a\neq b$. Also let c be an r-coloring of the set of all variable words over A of length n. Clearly, we may assume that $n=N_d$. By backwards induction and arguing as in the proof of Lemma 2.5, we select a sequence $(w_i)_{i=0}^{d-1}$ of variable words over A such that the following conditions are satisfied.

(C1) For every $i \in \{0, \dots, d-1\}$ the variable word w_i has length n_i where
$$n_i = N_{i+1} - N_i \overset{(2.44)}{=} r^{(k+2)^{M_i}}.$$

(C2) We have $c\big(v^\frown w_{d-1}(a)\big) = c\big(v^\frown w_{d-1}(b)\big)$ for every variable word v over A of length $N_{\ell-1}$.

(C3) For every $i \in \{0, \dots, d-2\}$ and every variable word v over A of length M_i, writing $v = (v_0, \dots, v_{M_i - 1})$ and setting
$$v^{(i)} = v \upharpoonright N_i \ \text{ and } \ v^{(i+1)} = w_{i+1}(v_{N_i})^\frown w_{i+2}(v_{N_i+1})^\frown \dots ^\frown w_{d-1}(v_{M_i-1}),$$
we have $c\big(v^{(i)\frown} w_i(a)^\frown v^{(i+1)}\big) = c\big(v^{(i)\frown} w_i(b)^\frown v^{(i+1)}\big)$.

We define $w = w_0(x_0)^\frown \dots ^\frown w_{d-1}(x_{d-1})$. By conditions (C1)–(C3), it is clear that w is as desired. The proof of Lemma 2.27 is completed. □

We are ready to complete the proof of Proposition 2.25.

Proof of Proposition 2.25. First observe that, by Proposition 2.19 and Lemma 2.27, for every pair d, r of positive integers we have
$$\mathrm{GR}(2, d, 1, r) \leqslant \mathrm{Sh_v}\big(1, \mathrm{H}(d, r), r\big). \tag{2.45}$$
On the other hand, invoking Lemma 2.27 once again, we see that
$$\mathrm{GR}(k+1, d, 1, r) \leqslant \mathrm{Sh_v}\big(k, \mathrm{GR}(k, d, 1, r), r\big) \tag{2.46}$$
for every triple k, d, r of positive integers with $k \geqslant 2$.

Now recall that, by Proposition 2.19, the numbers $\mathrm{H}(d, r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^4$. Therefore, by (2.45), (2.46) and Lemma 2.27, we conclude that the numbers $\mathrm{GR}(k, d, 1, r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^5$. The proof of Proposition 2.25 is completed. □

2.4. Notes and remarks

2.4.1. We have already pointed out that there are several different proofs of the Hales–Jewett theorem. The original proof was combinatorial in nature and was based on a *color focusing argument*, a method invented by van der Waerden [**vdW**]. The color focusing argument is very flexible, but has the drawback that it yields upper bounds of Ackermann type. Nevertheless, it is still influential and there are some interesting recent results which are proved using this method (see, e.g., [**W**]).

There is a second approach which utilizes the structure of the *Stone–Čech compactification βX of a discrete topological space* X, a classical construction which can also be identified with the set of all ultrafilters on X. The use of ultrafilters in the context of the Hales–Jewett theorem was first implemented by Carlson [**C**] and further developed by several authors (see, e.g., [**BBH, B2, HM1, McC2**]). This is a very fruitful approach which in addition leads to elegant proofs.

A third approach (closely related to the work of Carlson) was developed by Furstenberg and Katznelson in [**FK3**]. It uses tools from *topological dynamics* (the branch of the theory of dynamical systems which studies the behavior of iterations of continuous transformations acting on sufficiently regular spaces) and was motivated by Furstenberg's proof [**F**] of Szemerédi's theorem on arithmetic progressions [**Sz1**]. As such, it is naturally placed in the general context of ergodic Ramsey theory (see, e.g., [**McC1**]).

In spite of their diversity, the aforementioned proofs shed no light on the behavior of the Hales–Jewett numbers. In particular, the best known general upper bounds for these invariants are the ones obtained by Shelah, but still they are huge when compared to the known lower bounds. It is one of the central open problems of Ramsey theory to obtain tight estimates for the numbers $\mathrm{HJ}(k,r)$ and any significant improvement on Shelah's bounds would be of fundamental importance.

2.4.2. We note that the proof of Theorem 2.15 that we presented, follows the method developed by Shelah in his proof of the Graham–Rothschild theorem (see [**Sh1**, Theorem 2.2]). Working with m-variable words instead of m-parameter words makes the argument slightly more involved, but the overall strategy is identical. We also note that Theorem 2.15 has several infinite-dimensional extensions[4]. We will discuss in detail these extensions in Chapter 4.

[4]Theorem 2.15 can be derived, of course, from these extensions via a standard compactness argument. However, this reduction is ineffective and gives no quantitative information for the numbers $\mathrm{GR}(k,d,m,r)$.

CHAPTER 3

Strong subtrees

3.1. The Halpern–Läuchli theorem

The topic of this section is the study of partitions of finite Cartesian products of trees. Specifically, given a coloring of the product of a finite tuple $(T_1,\dots,T_d)$ of rooted, pruned and finitely branching trees, the goal is to find for each $i\in[d]$ a "structured" subset S_i of the tree T_i such that the product $S_1\times\cdots\times S_d$ is monochromatic. This problem is, of course, interesting on its own, but is also essential for the development of Ramsey theory for trees.

It is easy to see that some restriction has to be imposed on the colorings under consideration. Indeed, let T and S be two, say, dyadic trees of infinite height and color red an element (t,s) of $T\times S$ if $|t|_T\geqslant|s|_S$; otherwise, color it blue. Clearly, if A and B are infinite subsets of T and S respectively, then $A\times B$ contains elements of both colors. To avoid this pathological behavior, it is thus necessary to restrict our attention to colorings of certain subsets of products of trees. In this context, the most natural (and practically useful) choice is to consider colorings of level products. It turns out that once this restriction is imposed, one can obtain a very satisfactory positive answer to the aforementioned problem. This is the content of the following theorem. General notation and terminology about trees can be found in Section 1.6.

Theorem 3.1. *Let $\mathbf{T}=(T_1,\dots,T_d)$ be a rooted, pruned and finitely branching vector tree. Then for every finite coloring of the level product $\otimes\mathbf{T}$ of $\mathbf{T}$ there exists a vector strong subtree $\mathbf{S}$ of $\mathbf{T}$ of infinite height whose level product is monochromatic.*

Theorem 3.1 is known as the *strong subtree version of the Halpern–Läuchli theorem* and its formulation is due to Laver. It is a consequence of a slightly more general result due to Halpern and Läuchli [**HL**]. However, from a combinatorial perspective, Theorem 3.1 is the most important result of this kind.

The aforementioned result of Halpern and Läuchli can be stated in several equivalent ways. We will state the "dominating set version" which is quite close to the original formulation of Halpern and Läuchli. It also deals with colorings of level products of vector trees but it does not refer to vector strong subtrees. Instead it refers to dominating sets, a concept which we are about to introduce.

Let $\mathbf{T}=(T_1,\dots,T_d)$ be a rooted, pruned and finitely branching vector tree. Also let $\mathbf{D}=(D_1,\dots,D_d)$ be a vector subset of $\mathbf{T}$ and $\mathbf{t}=(t_1,\dots,t_d)\in\otimes\mathbf{T}$. We say that $\mathbf{D}$ is $\mathbf{t}$-*dominating* provided that: (i) $\mathbf{D}$ is level compatible (that is, the sets $D_1,\dots,D_d$ have a common level set), and (ii) for every $n\in\mathbb{N}$ with $n\geqslant|\mathbf{t}|_{\mathbf{T}}$

there exists $m \in \mathbb{N}$ such that for every $i \in [d]$ and every $s \in \mathrm{Succ}_{T_i}(t_i) \cap T_i(n)$ there exists $w \in D_i \cap T_i(m)$ with $s \leqslant_{T_i} w$. If $\mathbf{D}$ is $\mathbf{T}(0)$-dominating, then it will be referred to simply as *dominating*.

We are now ready to state the *dominating set version of the Halpern–Läuchli theorem*.

THEOREM 3.2. *Let d be a positive integer and $\mathbf{T} = (T_1, \dots, T_d)$ a rooted, pruned and finitely branching vector tree. Also let $\mathbf{D}$ be a dominating vector subset of $\mathbf{T}$ and $\mathcal{P}$ a subset of the level product $\otimes\mathbf{D}$ of $\mathbf{D}$. Then, either*

(a) *there exists a vector subset $\mathbf{X}$ of $\mathbf{D}$ which is dominating and whose level product is contained in $\mathcal{P}$, or*

(b) *there exists a vector subset $\mathbf{Y}$ of $\mathbf{D}$ which is $\mathbf{t}$-dominating for some $\mathbf{t} \in \otimes\mathbf{T}$ and whose level product is contained in the complement of $\mathcal{P}$.*

The proof of Theorem 3.2 will be given in Subsection 3.1.1. As we have already indicated, Theorem 3.1 follows from Theorem 3.2. The argument is simple and we will present it at this point. To this end, notice the following consequence of Theorem 3.2: if c is a finite coloring of the level product of a rooted, pruned and finitely branching vector tree $\mathbf{T}$, then one of the colors contains the level product of a vector subset $\mathbf{D}$ of $\mathbf{T}$ which is $\mathbf{t}$-dominating for some $\mathbf{t} \in \otimes\mathbf{T}$. Using this observation, Theorem 3.1 follows from the following general fact.

FACT 3.3. *Let $\mathbf{T}$ be a rooted, pruned and finitely branching vector tree, and $\mathbf{t} \in \otimes\mathbf{T}$. Also let $\mathbf{D} = (D_1, \dots, D_d)$ be a $\mathbf{t}$-dominating vector subset of $\mathbf{T}$. Then for every $\mathbf{s} = (s_1, \dots, s_d) \in \otimes\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}) \cap \otimes\mathbf{D}$ there exists $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ with $\mathbf{S}(0) = \mathbf{s}$ and such that $S_i \subseteq D_i$ for every $i \in [d]$.*

PROOF. Recursively and invoking the definition of a $\mathbf{t}$-dominating vector set, we select a strictly increasing sequence (m_n) in $L_{\mathbf{T}}(\mathbf{D})$ with $m_0 = |\mathbf{s}|_{\mathbf{T}}$ and such that for every $n \in \mathbb{N}$, every $i \in [d]$ and every $s \in \mathrm{Succ}_{T_i}(s_i) \cap T_i(m_n + 1)$ there exists $w \in D_i \cap T_i(m_{n+1})$ with $s \leqslant_{T_i} w$. It is then easy to construct for each $i \in [d]$ a strong subtree S_i of T_i such that $S_i(n) \subseteq D_i \cap T_i(m_n)$ for every $n \in \mathbb{N}$. □

3.1.1. Proof of Theorem 3.2. We follow the proof from [**AFK**] which proceeds by induction on the number of trees. First, we need to introduce some pieces of notation concerning dominating sets.

Let $\mathbf{T} = (T_1, \dots, T_d)$ be a rooted, pruned and finitely branching vector tree. If $\mathbf{A} = (A_1, \dots, A_d)$ and $\mathbf{B} = (B_1, \dots, B_d)$ are vector subsets of $\mathbf{T}$, then we say that $\mathbf{B}$ *dominates* $\mathbf{A}$ if for every $i \in [d]$ and every $a \in A_i$ there exists $b \in B_i$ with $a_i \leqslant_{T_i} b_i$. For every vector subset $\mathbf{D} = (D_1, \dots, D_d)$ of $\mathbf{T}$ and every $m \in \mathbb{N}$ let

$$\mathbf{D}(m) = \big(D_1 \cap T_1(m), \dots, D_d \cap T_d(m)\big). \tag{3.1}$$

(In particular, if D is a subset of a pruned tree T, then $D(m)$ stands for the set $D \cap T(m)$ for every $m \in \mathbb{N}$.) Moreover, for every $\mathbf{t} = (t_1, \dots, t_d) \in \otimes\mathbf{T}$ we set

$$\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{D}) = \big(\mathrm{Succ}_{T_1}(t_1) \cap D_1, \dots, \mathrm{Succ}_{T_d}(t_d) \cap D_d\big). \tag{3.2}$$

Notice that $\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{D})$ is a vector subset of $\mathrm{Succ}_{\mathbf{T}}(\mathbf{t})$. The following fact is straightforward.

FACT 3.4. *Let* $\mathbf{T}$ *be a rooted, pruned and finitely branching vector tree, and* $\mathbf{t} \in \otimes\mathbf{T}$. *Also let* $\mathbf{D}$ *be a vector subset of* $\mathbf{T}$.

(a) *The vector set* $\mathbf{D}$ *is* $\mathbf{t}$*-dominating if and only if* $\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{D})$ *is dominating in the vector tree* $\mathrm{Succ}_{\mathbf{T}}(\mathbf{t})$.

(b) *If* $\mathbf{D}$ *is* $\mathbf{t}$*-dominating, then there is a dominating vector subset* $\mathbf{E}$ *of* $\mathbf{T}$ *such that* $\mathbf{D}$ *is a vector subset of* $\mathbf{E}$, $L_{\mathbf{T}}(\mathbf{E}) = L_{\mathbf{T}}(\mathbf{D})$ *and satisfying* $\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{E}) = \mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{D})$.

We will also need the following fact.

FACT 3.5. *Let* $\mathbf{T}$ *be a rooted, pruned and finitely branching vector tree, and* $\mathbf{t} \in \otimes\mathbf{T}$. *Also let* $\mathbf{D}$ *be a* $\mathbf{t}$*-dominating vector subset of* $\mathbf{T}$. *Then there exists an infinite subset* L *of its level set* $L_{\mathbf{T}}(\mathbf{D})$ *such that for every infinite subset* M *of* L *the restriction* $\mathbf{D} \upharpoonright M$ *is* $\mathbf{t}$*-dominating.*

PROOF. Clearly, we may assume that $\mathbf{D}$ is dominating. We select a strictly increasing sequence (ℓ_n) in $L_{\mathbf{T}}(\mathbf{D})$ such that $\mathbf{D}(\ell_n)$ dominates $\mathbf{T}(n)$ for every $n \in \mathbb{N}$, and we set $L = \{\ell_n : n \in \mathbb{N}\}$. The proof of Fact 3.5 is completed. $\square$

We now proceed to the details of the proof of Theorem 3.2. The initial case "$d = 1$" is the content of the following lemma.

LEMMA 3.6. *Let* T *be a rooted, pruned and finitely branching tree, and* D *a dominating subset of* T. *If* P *is a subset of* D, *then either* P *is dominating, or there exists* $t \in T$ *such that* $D \setminus P$ *is* t*-dominating.*

PROOF. Assume that P is not dominating. Then there exists $n_0 \in \mathbb{N}$ such that $T(n_0)$ is not dominated by $P(m)$ for every $m \in \mathbb{N}$. Let $L \subseteq L_T(D)$ be as in Fact 3.5. The set $T(n_0)$ is finite since the tree T is finitely branching. Therefore, by the classical pigeonhole principle, there exist a node $t_0 \in T(n_0)$ and an infinite subset M of L such that $P \cap T(m) \cap \mathrm{Succ}_T(t_0) = \emptyset$ for every $m \in M$.

We will show that $D \setminus P$ is t_0-dominating. Indeed, let $n \geqslant |t_0|_T$ be arbitrary. Since $D \upharpoonright M$ is dominating, there exists $m \in M$ such that $D(m)$ dominates $\mathrm{Succ}_T(t_0) \cap T(n)$. On the other hand, by the previous discussion, we see that $D \cap T(m) \cap \mathrm{Succ}_T(t_0) = (D \setminus P) \cap T(m) \cap \mathrm{Succ}_T(t_0)$. This implies that $(D \setminus P) \cap T(m)$ dominates $\mathrm{Succ}_T(t_0) \cap T(n)$. Since n was arbitrary, we conclude that $D \setminus P$ is t_0-dominating and the proof of Lemma 3.6 is completed. $\square$

The rest of this subsection is devoted to the proof of the general inductive step. Specifically, let d be a positive integer and assume that the result has been proved for any d-tuple $(T_1, \dots, T_d)$ of rooted, pruned and finitely branching trees. *We emphasize that, in what follows, this positive integer* d *will be fixed.* Also it is convenient to denote a $(d+1)$-tuple of trees as $(\mathbf{T}, W)$ where $\mathbf{T} = (T_1, \dots, T_d)$ is a d-tuple of trees and W is a tree. Respectively, a vector subset of $(\mathbf{T}, W)$ will be denoted as $(\mathbf{D}, E)$ where $\mathbf{D}$ is a vector subset of $\mathbf{T}$ and E is a subset of W. Notice, in particular, that $(\mathbf{D}, E)$ is dominating if and only if both $\mathbf{D}$ and E are dominating and have a common level set.

Lemma 3.7. *Let $T_1, \dots, T_d, W$ be rooted, pruned and finitely branching trees and set $\mathbf{T} = (T_1, \dots, T_d)$. Also let $\mathbf{D}$ be a dominating vector subset of $\mathbf{T}$ and C an infinite chain of W with $L_{\mathbf{T}}(\mathbf{D}) = L_W(C)$. Then for every subset $\mathcal{P}$ of $\otimes(\mathbf{T}, W)$ one of the following is satisfied.*

(a) *There exist a vector subset $\mathbf{X}$ of $\mathbf{D}$ and an infinite subchain A of C such that $\mathbf{X}$ is dominating, $L_{\mathbf{T}}(\mathbf{X}) = L_W(A)$ and $\otimes(\mathbf{X}, A) \subseteq \mathcal{P}$.*

(b) *There exist a vector subset $\mathbf{Y}$ of $\mathbf{D}$, an infinite subchain B of C and $\mathbf{t} \in \otimes\mathbf{T}$ such that $\mathbf{Y}$ is dominating, $L_{\mathbf{T}}\big(\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{Y})\big) = L_W(B)$ and $\otimes\big(\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{Y}), B\big) \cap \mathcal{P} = \emptyset$.*

Proof. Let (w_n) be the $\leqslant_W$-increasing enumeration of C and set

$$\mathcal{R} = \bigcup_{n \in \mathbb{N}} \big\{ \mathbf{t} \in \otimes\mathbf{D} : |\mathbf{t}|_{\mathbf{T}} = |w_n|_W \text{ and } (\mathbf{t}, w_n) \in \mathcal{P} \big\}.$$

By our inductive assumptions, one of the following alternatives is satisfied.

(A1) There exists a vector subset $\mathbf{X}$ of $\mathbf{D}$ which is dominating and whose level product is contained in $\mathcal{R}$.

(A2) There exist $\mathbf{t} \in \otimes\mathbf{T}$ and a vector subset $\mathbf{Z}$ of $\mathbf{D}$ which is $\mathbf{t}$-dominating and whose level product is contained in $\otimes\mathbf{D} \setminus \mathcal{R}$.

If alternative (A1) holds true, then we set $A = C \upharpoonright L_{\mathbf{T}}(\mathbf{X})$. It is easy to check that the first part of the lemma is satisfied for $\mathbf{X}$ and A. Otherwise, by Fact 3.4, we may select a dominating vector subset $\mathbf{Y}$ of $\mathbf{T}$ such that $L_{\mathbf{T}}(\mathbf{Y}) = L_{\mathbf{T}}(\mathbf{Z})$ and $\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{Y}) = \mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{Z})$. Therefore, setting $B = C \upharpoonright L_{\mathbf{T}}(\mathbf{Y})$, we see that $\otimes(\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{Y}), B) = \otimes(\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{Z}), B) \subseteq \otimes(\mathbf{Z}, B) \subseteq \otimes(\mathbf{T}, W) \setminus \mathcal{P}$. The proof of Lemma 3.7 is completed. $\square$

We are about to introduce a family of sets which plays a crucial role in the proof of Theorem 3.2.

Definition 3.8. *Let $\mathbf{T}$ and W be as in Lemma 3.7. Also let $\mathbf{t} \in \otimes\mathbf{T}$, $w \in W$ and $\mathcal{P} \subseteq \otimes(\mathbf{T}, W)$. By $\mathcal{D}(\mathbf{t}, w, \mathcal{P})$ we denote the family of all dominating vector subsets $(\mathbf{D}, E)$ of $(\mathbf{T}, W)$ satisfying the following property. For every dominating vector subset $\mathbf{Y}$ of $\mathbf{D}$ and every $v \in \mathrm{Succ}_W(w) \cap E$ there exist a dominating vector subset $\mathbf{X}$ of $\mathbf{Y}$ and an infinite chain $C \subseteq \mathrm{Succ}_W(v) \cap E$ such that*

$$L_{\mathbf{T}}\big(\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{X})\big) = L_W(C) \quad \text{and} \quad \otimes\big(\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}, \mathbf{X}), C\big) \subseteq \mathcal{P}. \tag{3.3}$$

Note that in Definition 3.8 we do not demand that $\mathbf{t}$ and w have necessarily the same length. This will be crucial in the following lemma.

Lemma 3.9. *Let $\mathbf{T}$ and W be as in Lemma 3.7. Also let $(\mathbf{D}, E)$ be a dominating vector subset of $(\mathbf{T}, W)$ and $\mathcal{P} \subseteq \otimes(\mathbf{D}, E)$. Then one of the following is satisfied.*

(a) *There exists an infinite subset L of $L_{(\mathbf{T}, W)}(\mathbf{D}, E)$ such that $(\mathbf{D} \upharpoonright L, E \upharpoonright L)$ belongs to $\mathcal{D}\big(\mathbf{T}(0), W(0), \mathcal{P}\big)$.*

(b) *There exist $\mathbf{t}' \in \otimes\mathbf{T}$, $w' \in W$ and a vector subset $(\mathbf{D}', E')$ of $(\mathbf{D}, E)$ such that $(\mathbf{D}', E')$ belongs to $\mathcal{D}(\mathbf{t}', w', \mathcal{Q})$ where $\mathcal{Q} = \otimes(\mathbf{D}, E) \setminus \mathcal{P}$.*

Proof. Assume that neither (a) nor (b) is satisfied. We will derive a contradiction using Lemma 3.7. To this end fix an enumeration $(\mathbf{t}_n)$ of the level product $\otimes\mathbf{T}$ of $\mathbf{T}$. By Fact 3.5, we also fix an infinite subset M of $L_{(\mathbf{T},W)}(\mathbf{D},E)$ such that $(\mathbf{D}\upharpoonright N, E\upharpoonright N)$ is dominating for every infinite subset N of M. Recursively, we will select two sequences (w^*_n) and (w_n) in $E\upharpoonright M$, a sequence $(\mathbf{D}_n)$ of dominating vector subsets of $\mathbf{D}\upharpoonright M$ and two strictly increasing sequences (ℓ_n) and (m_n) of natural numbers such that the following conditions are satisfied.

- (C1) If $\mathbf{X}$ is a dominating vector subset $\mathbf{D}_0$ and $C\subseteq \mathrm{Succ}_W(w^*_0)\cap(E\upharpoonright M)$ is an infinite chain with $L_W(C)=L_{\mathbf{T}}(\mathbf{X})$, then $\otimes(\mathbf{X},C)\nsubseteq\mathcal{P}$.
- (C2) For every $n\in\mathbb{N}$ we have $|w^*_n|_W=\ell_n<m_n$. Moreover, $\big(\mathbf{D}_n(m_n),E(m_n)\big)$ dominates $\big(\mathbf{T}(\ell_n),W(\ell_n)\big)$ and $w_n\in E(m_n)\cap\mathrm{Succ}_W(w^*_n)$.
- (C3) For every $n\in\mathbb{N}$ we have that $\mathbf{D}_{n+1}$ is a dominating vector subset of $\mathbf{D}_n$ and $w^*_{n+1}\in\mathrm{Succ}_W(w_n)\cap E_n$ where $E_n=E\upharpoonright L_{\mathbf{T}}(\mathbf{D}_n)$. Moreover, for every dominating vector subset $\mathbf{X}$ of $\mathbf{D}_{n+1}$ and every infinite chain $C\subseteq\mathrm{Succ}_W(w^*_{n+1})\cap E_n$ with $L_W(C)=L_{\mathbf{T}}(\mathbf{X})$ we have that $\otimes\big(\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}_n,\mathbf{X}),C\big)\nsubseteq\mathcal{Q}$.

The first step of the recursive selection (i.e., the choice of $\mathbf{D}_0$, w^*_0, w_0, ℓ_0 and m_0) follows from our assumption that part (a) is not satisfied. The next steps are carried out using the negation of part (b).

We set

$$\mathbf{D}_\infty=\bigcup_{n\in\mathbb{N}}\mathbf{D}_n(m_n)\ \text{ and }\ C_\infty=\{w_n:n\in\mathbb{N}\}.$$

By conditions (C2) and (C3), we see that $\mathbf{D}_\infty$ is a dominating vector subset of $\mathbf{D}_0$ and C_∞ is an infinite chain of $\mathrm{Succ}_W(w^*_0)\cap(E\upharpoonright M)$ with $L_W(C_\infty)=L_{\mathbf{T}}(\mathbf{D}_\infty)$. Therefore, by (C1) and Lemma 3.7, there exist a dominating vector subset $\mathbf{Y}$ of $\mathbf{D}_\infty$, an infinite subchain B of C_∞ and $\mathbf{t}\in\otimes\mathbf{T}$ such that $L_{\mathbf{T}}\big(\mathrm{Succ}_{\mathbf{T}}(\mathbf{t},\mathbf{Y})\big)=L_W(B)$ and $\otimes\big(\mathrm{Succ}_{\mathbf{T}}(\mathbf{t},\mathbf{Y}),B\big)\subseteq\mathcal{Q}$.

We are now in a position to derive the contradiction. Let $n_0\in\mathbb{N}$ be such that $\mathbf{t}_{n_0}=\mathbf{t}$ and set $N=\{m_n:n\geqslant n_0+1\}$, $C=B\upharpoonright N$ and $\mathbf{X}=\mathbf{Y}\upharpoonright N$. First observe that $\otimes\big(\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}_{n_0},\mathbf{X}),C\big)\subseteq\otimes\big(\mathrm{Succ}_{\mathbf{T}}(\mathbf{t},\mathbf{Y}),B\big)\subseteq\mathcal{Q}$. Next, by (C2) and (C3), notice that $\mathbf{X}$ is a dominating vector subset of $\mathbf{D}_{n_0+1}$ and $C\subseteq\mathrm{Succ}_W(w^*_{n_0+1})\cap E_{n_0}$ is an infinite chain with $L_W(C)=L_{\mathbf{T}}(\mathbf{X})=N$. Therefore, invoking condition (C3) once again, we conclude that $\otimes\big(\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}_{n_0},\mathbf{X}),C\big)\nsubseteq\mathcal{Q}$. This is clearly a contradiction. The proof of Lemma 3.9 is thus completed. □

The following lemma is the last step of the proof.

Lemma 3.10. *Let $\mathbf{T}$ and W be as in Lemma 3.7, $\mathbf{t}'\in\otimes\mathbf{T}$ and $w'\in W$. Also let $\mathcal{Q}\subseteq\otimes(\mathbf{T},W)$ and $(\mathbf{D}',E')\in\mathcal{D}(\mathbf{t}',w',\mathcal{Q})$. Then for every $(\mathbf{s},v)\in\otimes(\mathbf{T},W)$ with $\mathbf{s}\in\otimes\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}')$ and $v\in\mathrm{Succ}_W(w')$ there exists a vector subset $(\mathbf{D}'',E'')$ of $(\mathbf{D}',E')$ which is $(\mathbf{s},v)$-dominating and such that $\otimes(\mathbf{D}'',E'')\subseteq\mathcal{Q}$.*

Proof. We fix $(\mathbf{s},v)\in\otimes(\mathbf{T},W)$ with $\mathbf{s}\in\otimes\mathrm{Succ}_{\mathbf{T}}(\mathbf{t}')$ and $v\in\mathrm{Succ}_W(w')$, and we set $\ell=|\mathbf{s}|_{\mathbf{T}}=|v|_W$. The proof is based on the following claim.

Claim 3.11. *For every integer $n \geqslant \ell$ there exist $m \in \mathbb{N}$, a dominating vector subset $\mathbf{X}$ of $\mathbf{D}'$ and a subset H of E' such that $\big(\mathbf{X}(m), H(m)\big)$ dominates $\big(\mathbf{T}(n), \mathrm{Succ}_W(v) \cap W(n)\big)$ and $\otimes\big(\mathrm{Succ}_{\mathbf{T}}\big(\mathbf{s}, \mathbf{X}(m)\big), H(m)\big) \subseteq \mathcal{Q}$.*

Granting Claim 3.11 the proof of the lemma is completed as follows. Recursively and using Claim 3.11, we may select a strictly increasing sequence (m_n) in $\mathbb{N}$ with $m_0 \geqslant \ell$, a sequence $(\mathbf{X}_n)$ of vector subsets of $\mathbf{D}'$ and a sequence (H_n) of subsets of E' such that for every $n \in \mathbb{N}$ we have that $\big(\mathbf{X}_n(m_n), H_n(m_n)\big)$ dominates $\big(\mathbf{T}(n), \mathrm{Succ}_W(v) \cap W(n)\big)$ and $\otimes\big(\mathrm{Succ}_{\mathbf{T}}\big(\mathbf{s}, \mathbf{X}(m_n)\big), H(m_n)\big) \subseteq \mathcal{Q}$. Hence, setting

$$\mathbf{D}'' = \bigcup_{n\in\mathbb{N}} \mathrm{Succ}_{\mathbf{T}}\big(\mathbf{s}, \mathbf{X}(m_n)\big) \quad \text{and} \quad E'' = \bigcup_{n\in\mathbb{N}} H_n(m_n),$$

we see that $\mathbf{D}''$ and E'' satisfy the requirements of the lemma.

It remains to prove Claim 3.11. Notice first that $(\mathbf{D}', E') \in \mathcal{D}(\mathbf{s}, v, \mathcal{Q})$. Fix an integer $n \geqslant \ell$ and let $(w_k)_{k=1}^r$ be an enumeration of the set $\mathrm{Succ}_W(v) \cap W(n)$. By repeated applications of Definition 3.8, we obtain a sequence $(\mathbf{X}_k)_{k=1}^r$ of dominating vector subsets of $\mathbf{D}'$ and a sequence $(C_k)_{k=1}^r$ of infinite chains of E' such that the following conditions are satisfied.

- (C1) For every $k \in [r]$ we have that $C_k \subseteq \mathrm{Succ}_W(w_k) \cap E'$, $L_{\mathbf{T}}(\mathbf{X}_k) = L_W(C_k)$ and $\otimes\big(\mathrm{Succ}_{\mathbf{T}}(\mathbf{s}, \mathbf{X}_k), C_k\big) \subseteq \mathcal{P}$.
- (C2) For every $k \in \{2, \dots, r\}$ we have that $\mathbf{X}_k$ is a vector subset of $\mathbf{X}_{k-1}$.

Let $L = L_{\mathbf{T}}(\mathbf{X}_r) = L_W(C_r)$. We set $m = \min(L)$, $\mathbf{X} = \mathbf{X}_r$ and $H = \bigcup_{k=1}^r (C_k \upharpoonright L)$. It is easy to check that with these choices the result follows. This completes the proof of Claim 3.11, and as we have already indicated, the proof of Lemma 3.10 is also completed. □

We are in a position to complete the proof of the general inductive step of the theorem. Let $T_1, \dots, T_d, W$ be rooted, pruned and finitely branching trees and set $\mathbf{T} = (T_1, \dots, T_d)$. Also let $(\mathbf{D}, E)$ be a dominating subset of $(\mathbf{T}, W)$ and $\mathcal{P}$ a subset of $\otimes(\mathbf{D}, E)$, and assume that $\mathcal{P}$ does not contain the level product of a dominating vector subset of $(\mathbf{D}, E)$. This assumption implies, in particular, that for every infinite subset L of $L_{(\mathbf{T},W)}(\mathbf{D}, E)$ we have that $(\mathbf{D} \upharpoonright L, E \upharpoonright L) \notin \mathcal{D}\big(\mathbf{T}(0), W(0), \mathcal{P}\big)$. By Lemmas 3.9 and 3.10, we see that the complement of $\mathcal{P}$ must contain the level product of a vector subset $(\mathbf{D}'', E'')$ of $(\mathbf{D}, E)$ which is $(\mathbf{s}, v)$-dominating for some $(\mathbf{s}, v) \in \otimes(\mathbf{T}, W)$. This completes the proof of the inductive step and so the entire proof of Theorem 3.2 is completed.

3.2. Milliken's tree theorem

We now turn our attention to finite colorings of strong subtrees. These questions were investigated in detail by Milliken in [**M2, M3**]. His results, collectively known as *Milliken's tree theorem*, are naturally categorized according to the height of the strong subtrees that we color. To state them, it is convenient to introduce some pieces of notation.

Let $\mathbf{T}$ be a rooted, balanced and finitely branching vector tree. For every $k \in \mathbb{N}$, every $\mathbf{A} \in \mathrm{Str}_k(\mathbf{T})$ and every integer $m \geqslant k$ let

$$\mathrm{Str}_m(\mathbf{A},\mathbf{T}) = \{\mathbf{S} \in \mathrm{Str}_m(\mathbf{T}) : \mathbf{S} \upharpoonright k = \mathbf{A}\}. \tag{3.4}$$

If, in addition, $\mathbf{T}$ has infinite height[1], then we set

$$\mathrm{Str}_{<\infty}(\mathbf{A},\mathbf{T}) = \{\mathbf{S} \in \mathrm{Str}_{<\infty}(\mathbf{T}) : \mathbf{S} \upharpoonright k = \mathbf{A}\} \tag{3.5}$$

and

$$\mathrm{Str}_{\infty}(\mathbf{A},\mathbf{T}) = \{\mathbf{S} \in \mathrm{Str}_{\infty}(\mathbf{T}) : \mathbf{S} \upharpoonright k = \mathbf{A}\}. \tag{3.6}$$

In the special case where $\mathbf{A} = \mathbf{T} \upharpoonright k$, the above sets will be denoted simply by $\mathrm{Str}_m(k,\mathbf{T})$, $\mathrm{Str}_{<\infty}(k,\mathbf{T})$ and $\mathrm{Str}_{\infty}(k,\mathbf{T})$ respectively. Notice, in particular, that $\mathrm{Str}_m(0,\mathbf{T}) = \mathrm{Str}_m(\mathbf{T})$, $\mathrm{Str}_{<\infty}(0,\mathbf{T}) = \mathrm{Str}_{<\infty}(\mathbf{T})$ and $\mathrm{Str}_{\infty}(0,\mathbf{T}) = \mathrm{Str}_{\infty}(\mathbf{T})$. Finally, recall that for every vector subset $\mathbf{X}$ of $\mathbf{T}$ the *depth* of $\mathbf{X}$ in $\mathbf{T}$ is the least $n \in \mathbb{N}$ such that $\mathbf{X}$ is a vector subset of $\mathbf{T} \upharpoonright n$; it is denoted by $\mathrm{depth}_{\mathbf{T}}(\mathbf{X})$.

3.2.1. Colorings of strong subtrees of finite height. The first instance of the circle of results that we present in this section deals with colorings of strong subtrees of a fixed finite height. Specifically, we have the following theorem.

THEOREM 3.12. *For every rooted, pruned and finitely branching vector tree* $\mathbf{T}$*, every positive integer* k *and every finite coloring of* $\mathrm{Str}_k(\mathbf{T})$ *there is* $\mathbf{S} \in \mathrm{Str}_{\infty}(\mathbf{T})$ *such that the set* $\mathrm{Str}_k(\mathbf{S})$ *is monochromatic.*

Note that there is also a finite version of Theorem 3.12 which is obtained with a standard compactness argument. A quantitative refinement of this finite version will be presented in Section 3.3.

The proof of Theorem 3.12 is based on the following lemma.

LEMMA 3.13. *Let* $\mathbf{T} = (T_1, \dots, T_d)$ *be a rooted, pruned and finitely branching vector tree. Also let* $k \in \mathbb{N}$ *and* $\mathbf{A} \in \mathrm{Str}_k(\mathbf{T})$*, and set* $n = \mathrm{depth}_{\mathbf{T}}(\mathbf{A})$*. Then for every* $\mathcal{F} \subseteq \mathrm{Str}_{k+1}(\mathbf{T})$ *there exists* $\mathbf{S} \in \mathrm{Str}_{\infty}(n,\mathbf{T})$ *such that either* $\mathrm{Str}_{k+1}(\mathbf{A},\mathbf{S}) \subseteq \mathcal{F}$ *or* $\mathrm{Str}_{k+1}(\mathbf{A},\mathbf{S}) \cap \mathcal{F} = \emptyset$*.*

PROOF. Notice that the case "$k = 0$" is a restatement of the strong subtree version of the Halpern–Läuchli theorem. Therefore, in what follows, we may assume that $k \geqslant 1$.

First we will deal with the case "$d = 1$". Specifically, let T be a rooted, pruned and finitely branching tree, k a positive integer and $A \in \mathrm{Str}_k(T)$. Set $n = \mathrm{depth}_T(A)$ and fix $\mathcal{F} \subseteq \mathrm{Str}_{k+1}(T)$. Let $\{t_1, \dots, t_\ell\}$ be an enumeration of the n-level $T(n)$ of T. For every $i \in [\ell]$ let $S_i = \mathrm{Succ}_T(t_i)$ and set $\mathbf{S} = (S_1, \dots, S_\ell)$. For every $\mathbf{s} \in \otimes\mathbf{S}$ we define $E(A,\mathbf{s}) \in \mathrm{Str}_{k+1}(A,T)$ as follows. Let $A(k-1)$ be the $(k-1)$-level of A and notice that $A(k-1) \subseteq T(n-1)$. For every $t \in A(k-1)$ set

$$I(t) = \big\{ i \in [\ell] : t_i \in \mathrm{ImmSucc}_T(t) \big\}$$

and let

$$I(A) = \bigcup_{t \in A(k-1)} I(t).$$

[1]Notice that a balanced vector tree $\mathbf{T}$ has infinite height if and only if it is pruned.

Finally, for every $\mathbf{s} = (s_1, \dots, s_\ell) \in \otimes\mathbf{S}$ we define

$$E(A, \mathbf{s}) = A \cup \{s_i : i \in I(A)\}.$$

It is easy to see that the map

$$\otimes\mathbf{S} \ni \mathbf{s} \mapsto E(A, \mathbf{s}) \in \mathrm{Str}_{k+1}(A, T)$$

is a surjection. Therefore, by Theorem 3.1, there exists $\mathbf{R} = (R_1, \dots, R_\ell) \in \mathrm{Str}_\infty(\mathbf{S})$ such that either $\{E(A, \mathbf{s}) : \mathbf{s} \in \otimes\mathbf{R}\} \subseteq \mathcal{F}$ or $\{E(A, \mathbf{s}) : \mathbf{s} \in \otimes\mathbf{R}\} \cap \mathcal{F} = \emptyset$. We set

$$S = (T \upharpoonright n) \cup \bigcup_{i=1}^{\ell} R_i.$$

Clearly, we have $S \in \mathrm{Str}(n, T)$ and either $\mathrm{Str}_{k+1}(A, S) \subseteq \mathcal{F}$ or $\mathrm{Str}_{k+1}(A, S) \cap \mathcal{F} = \emptyset$.

We now proceed to the general case. Let $d \geqslant 2$ and let $\mathbf{T} = (T_1, \dots, T_d)$, k, $\mathbf{A}$, n and $\mathcal{F}$ be as in the statement of the lemma. We fix an element τ with $\tau \notin T_i$ for every $i \in [d]$ and for every vector subset $\mathbf{W} = (W_1, \dots, W_d)$ of $\mathbf{T}$ we set $\tau(\mathbf{W}) = \{\tau\} \cup \bigcup_{i=1}^{d} W_i$. We set $T = \tau(\mathbf{T})$ and $A = \tau(\mathbf{A})$. Notice that T is naturally viewed as a pruned and finitely branching tree with root τ. Moreover, A is a strong subtree of T of height $k+1$ and of depth $n+1$. Finally, observe that the sets $\mathrm{Str}_{k+1}(\mathbf{A}, \mathbf{T})$ and $\mathrm{Str}_{k+2}(A, T)$, as well as the sets $\mathrm{Str}_\infty(n, \mathbf{T})$ and $\mathrm{Str}_\infty(n+1, T)$, can be identified via the map $\mathbf{W} \mapsto \tau(\mathbf{W})$. Taking into account these remarks, we see that the general case is reduced to the case "$d = 1$". The proof of Lemma 3.13 is thus completed. □

Notice that for every rooted, pruned and finitely branching vector tree $\mathbf{T}$ and every $n \in \mathbb{N}$ the set of all vector strong subtrees of $\mathbf{T}$ of depth n is finite. Therefore, by repeated applications of Lemma 3.13, we obtain the following corollary.

Corollary 3.14. *Let $\mathbf{T}$ be a rooted, pruned and finitely branching vector tree. Then for every $n \in \mathbb{N}$ and every finite coloring of $\mathrm{Str}_{<\infty}(\mathbf{T})$ there is $\mathbf{S} \in \mathrm{Str}_\infty(n, \mathbf{T})$ such that for every vector strong subtree $\mathbf{A}$ of $\mathbf{T}$ with $\mathrm{depth}_{\mathbf{T}}(\mathbf{A}) = n$ the set $\mathrm{Str}_{h(\mathbf{A})+1}(\mathbf{A}, \mathbf{S})$ is monochromatic.*

We are now ready to give the proof of Theorem 3.12.

Proof of Theorem 3.12. By induction on k. The case "$k = 1$" follows from the strong subtree version of the Halpern–Läuchli theorem, and so let $k \geqslant 1$ and assume that the result has been proved up to k. Fix a rooted, pruned and finitely branching vector tree $\mathbf{T}$ and let c be a finite coloring of $\mathrm{Str}_{k+1}(\mathbf{T})$. Recursively and using Corollary 3.14, we select a sequence $(\mathbf{T}_n)$ in $\mathrm{Str}_\infty(\mathbf{T})$ such that: (i) $\mathbf{T}_0 = \mathbf{T}$, (ii) $\mathbf{T}_{n+1} \in \mathrm{Str}_\infty(n+k, \mathbf{T}_n)$ for every $n \in \mathbb{N}$, and (iii) the family $\mathrm{Str}_{k+1}(\mathbf{A}, \mathbf{T}_{n+1})$ is monochromatic for every $\mathbf{A} \in \mathrm{Str}_k(\mathbf{T}_n)$ with $\mathrm{depth}_{\mathbf{T}_n}(\mathbf{A}) = n+k$. For every $n \in \mathbb{N}$ write $\mathbf{T}_n = (T_1^{(n)}, \dots, T_d^{(n)})$ and for every $i \in [d]$ set

$$R_i = (T_i^{(0)} \upharpoonright k) \cup \bigcup_{n=1}^{\infty} T_i^{(n)}(n+k-1).$$

Notice that $\mathbf{R} = (R_1, \dots, R_d)$ is a vector strong subtree of $\mathbf{T}$ of infinite height. Also observe that $c(\mathbf{B}) = c(\mathbf{C})$ for every $\mathbf{B}, \mathbf{C} \in \mathrm{Str}_{k+1}(\mathbf{R})$ with $\mathbf{B} \upharpoonright k = \mathbf{C} \upharpoonright k$.

In other words, the coloring of $\mathrm{Str}_{k+1}(\mathbf{R})$ is reduced to a coloring of $\mathrm{Str}_k(\mathbf{R})$. By our inductive assumptions, there exists $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{R}) \subseteq \mathrm{Str}_\infty(\mathbf{T})$ such that the set $\mathrm{Str}_{k+1}(\mathbf{S})$ is monochromatic. The proof of Theorem 3.12 is completed. $\square$

3.2.2. Colorings of strong subtrees of infinite height. Let $\mathbf{T}$ be a rooted, pruned and finitely branching vector tree. A subset $\mathcal{F}$ of $\mathrm{Str}_\infty(\mathbf{T})$ is called *Ramsey* if for every $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ there exists $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{S})$ such that either $\mathrm{Str}_\infty(\mathbf{R}) \subseteq \mathcal{F}$ or $\mathrm{Str}_\infty(\mathbf{R}) \cap \mathcal{F} = \emptyset$. It is called *completely Ramsey* if for every $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ and every $\mathbf{A} \in \mathrm{Str}_{<\infty}(\mathbf{S})$ there exists $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A}, \mathbf{S})$ such that either $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \subseteq \mathcal{F}$ or $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \cap \mathcal{F} = \emptyset$. Clearly, if $\mathcal{F}$ is completely Ramsey, then it is Ramsey.

Using the axiom of choice one can easily construct subsets of $\mathrm{Str}_\infty(\mathbf{T})$ which are not Ramsey. However, these examples are by no means "canonical" and one expects to be able to prove that "simple" subsets of $\mathrm{Str}_\infty(\mathbf{T})$ are not only Ramsey but in fact completely Ramsey. This basic intuition turns out to be correct. The proper concept of "simplicity" in this context is related to the complexity of a given subset of $\mathrm{Str}_\infty(\mathbf{T})$ with respect to an appropriate topology on $\mathrm{Str}_\infty(\mathbf{T})$ which we are about to introduce.

Let

$$\mathcal{E} = \big\{ \mathrm{Str}_\infty(\mathbf{A}, \mathbf{S}) : \mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T}) \text{ and } \mathbf{A} \in \mathrm{Str}_{<\infty}(\mathbf{S}) \big\} \tag{3.7}$$

and define the *Ellentuck topology* on $\mathrm{Str}_\infty(\mathbf{T})$ to be the topology generated by $\mathcal{E}$, that is, the smallest topology on $\mathrm{Str}_\infty(\mathbf{T})$ that contains every member of $\mathcal{E}$. It is easy to see that the family $\mathcal{E}$ is actually a basis for the Ellentuck topology.

We are ready to state the main result of this section. General facts about the Baire property can be found in Appendix C.

THEOREM 3.15. *Let $\mathbf{T}$ be a rooted, pruned and finitely branching vector tree. Then a subset of $\mathrm{Str}_\infty(\mathbf{T})$ is completely Ramsey if and only if it has the Baire property in the Ellentuck topology.*

Although the Ellentuck topology on $\mathrm{Str}_\infty(\mathbf{T})$ is somewhat exotic[2], Theorem 3.15 has some consequences which refer to another, more natural, topology on the set $\mathrm{Str}_\infty(\mathbf{T})$. Specifically, for every $\mathbf{S}, \mathbf{R} \in \mathrm{Str}_\infty(\mathbf{T})$ with $\mathbf{S} \neq \mathbf{R}$ let

$$d_{\mathbf{T}}(\mathbf{S}, \mathbf{R}) = 2^{-n} \tag{3.8}$$

where n is the least natural number with $\mathbf{S} \upharpoonright n \neq \mathbf{R} \upharpoonright n$. Also let $d_{\mathbf{T}}(\mathbf{S}, \mathbf{S}) = 0$ for every $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$. It is easy to see that $d_{\mathbf{T}}$ is a metric on $\mathrm{Str}_\infty(\mathbf{T})$ and the metric space $(\mathrm{Str}_\infty(\mathbf{T}), d_{\mathbf{T}})$ is separable and complete. Moreover, the family

$$\mathcal{M} = \{ \mathrm{Str}_\infty(\mathbf{A}, \mathbf{T}) : \mathbf{A} \text{ is a finite vector strong subtree of } \mathbf{T} \} \tag{3.9}$$

is a basis for this metric topology on $\mathrm{Str}_\infty(\mathbf{T})$. In particular, by (3.7) and (3.9), we see that the metric topology on $\mathrm{Str}_\infty(\mathbf{T})$ is coarser than the Ellentuck topology, and as a consequence we obtain the following corollary.

COROLLARY 3.16. *Let $\mathbf{T}$ be a rooted, pruned and finitely branching vector tree. Then every C-set of $(\mathrm{Str}_\infty(\mathbf{T}), d_{\mathbf{T}})$ is completely Ramsey.*

[2]For instance, the Ellentuck topology on $\mathrm{Str}_\infty(\mathbf{T})$ is neither second countable nor metrizable.

Proof. First recall that the family of C-sets of $(\mathrm{Str}_\infty(\mathbf{T}), d_{\mathbf{T}})$ is the smallest σ-algebra on $\mathrm{Str}_\infty(\mathbf{T})$ containing all metrically open sets and closed under the Souslin operation (see Appendix C). Since every metrically open set is open in the Ellentuck topology, by Proposition C.2 and Theorem C.3, we see that every C-set of $(\mathrm{Str}_\infty(\mathbf{T}), d_{\mathbf{T}})$ has the Baire property in the Ellentuck topology. By Theorem 3.15, the result follows. $\square$

The rest of this section is devoted to the proof of Theorem 3.15. The argument is somewhat lengthy and so we will comment on it for the benefit of the reader. After some initial reductions, Theorem 3.15 boils down to showing that every open set in the Ellentuck topology is completely Ramsey. Let $\mathcal{O}$ be such a set and fix $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ and $\mathbf{A} \in \mathrm{Str}_{<\infty}(\mathbf{S})$. Since $\mathcal{O}$ is open, we may find a basic open set $\mathrm{Str}_\infty(\mathbf{B}, \mathbf{R})$ which is contained in $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{S})$ and such that either $\mathrm{Str}_\infty(\mathbf{B}, \mathbf{R}) \subseteq \mathcal{O}$ or $\mathrm{Str}_\infty(\mathbf{B}, \mathbf{R}) \cap \mathcal{O} = \emptyset$. Note that this fact barely misses to prove that the set $\mathcal{O}$ is completely Ramsey, and observe that what we actually need to ensure is that the aforementioned basic open set $\mathrm{Str}_\infty(\mathbf{B}, \mathbf{R})$ can be chosen so that $\mathbf{B} = \mathbf{A}$. This selection is the combinatorial core of the proof and is achieved by implementing, as pigeonhole principle, the strong subtree version of the Halpern–Läuchli theorem in a powerful method discovered by Galvin and Prikry [**GR**] and further developed by Ellentuck [**E**]. The following definitions are the main conceptual tools.

Definition 3.17. *Let $\mathbf{T}$ be a rooted, pruned and finitely branching vector tree. Also let $\mathcal{F} \subseteq \mathrm{Str}_\infty(\mathbf{T})$, $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ and $\mathbf{A} \in \mathrm{Str}_{<\infty}(\mathbf{S})$. We say that $\mathbf{S}$* accepts *$\mathbf{A}$* into *$\mathcal{F}$ if $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{S}) \subseteq \mathcal{F}$. We say that $\mathbf{S}$* rejects *$\mathbf{A}$* from *$\mathcal{F}$ if there is no $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A}, \mathbf{S})$ accepting $\mathbf{A}$ into $\mathcal{F}$. Finally, we say that $\mathbf{S}$* decides *$\mathbf{A}$* relative to *$\mathcal{F}$ if either $\mathbf{S}$ accepts $\mathbf{A}$ into $\mathcal{F}$ or $\mathbf{S}$ rejects $\mathbf{A}$ from $\mathcal{F}$.*

If the family $\mathcal{F}$ is understood, then we will simply say that $\mathbf{S}$ accepts*,* rejects *and* decides *$\mathbf{A}$ respectively.*

We need some basic properties concerning the above notions which are gathered in the following fact.

Fact 3.18. *Let $\mathbf{T}$, $\mathcal{F}$, $\mathbf{S}$ and $\mathbf{A}$ be as in Definition 3.17. Also let $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{S})$ such that $\mathbf{A} \in \mathrm{Str}_{<\infty}(\mathbf{R})$. Then the following hold.*

(a) *If $\mathbf{S}$ accepts $\mathbf{A}$, then $\mathbf{R}$ accepts $\mathbf{A}$.*
(b) *If $\mathbf{S}$ rejects $\mathbf{A}$, then $\mathbf{R}$ rejects $\mathbf{A}$.*
(c) *If $\mathbf{S}$ decides $\mathbf{A}$, then $\mathbf{R}$ decides $\mathbf{A}$ in the same way that $\mathbf{S}$ does.*
(d) *There exists $\mathbf{Y} \in \mathrm{Str}_\infty(\mathrm{depth}_{\mathbf{S}}(\mathbf{A}), \mathbf{S})$ which decides $\mathbf{A}$.*

Proof. Parts (a), (b) and (c) are straightforward consequences of the relevant definitions. For part (d) notice that if $\mathbf{S}$ rejects $\mathbf{A}$, then we may set $\mathbf{Y} = \mathbf{S}$. Otherwise, there exists $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A}, \mathbf{S})$ such that $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \subseteq \mathcal{F}$. We select $\mathbf{Y} \in \mathrm{Str}_\infty(\mathrm{depth}_{\mathbf{S}}(\mathbf{A}), \mathbf{S})$ so that $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{Y}) = \mathrm{Str}_\infty(\mathbf{A}, \mathbf{R})$. Clearly, $\mathbf{Y}$ is as desired. The proof of Fact 3.18 is completed. $\square$

We proceed with the following lemma.

LEMMA 3.19. *Let* $\mathbf{T}$*,* $\mathcal{F}$*,* $\mathbf{S}$ *and* $\mathbf{A}$ *be as in Definition* 3.17*. Then there exists* $\mathbf{Y} \in \mathrm{Str}_\infty(\mathbf{A},\mathbf{S})$ *which decides every* $\mathbf{B} \in \mathrm{Str}_{<\infty}(\mathbf{A},\mathbf{Y})$*.*

PROOF. By passing to a vector strong subtree of $\mathbf{S}$ of infinite height, we may assume that $\mathbf{A}$ is an initial vector subtree of $\mathbf{S}$. Hence, setting $n_0 = h(\mathbf{A})$, we have $\mathbf{A} = \mathbf{S} \upharpoonright n_0$. Recursively, we select a sequence $(\mathbf{Y}_n)$ in $\mathrm{Str}_\infty(\mathbf{S})$ such that the following conditions are satisfied.

(C1) We have that $\mathbf{Y}_0 \in \mathrm{Str}_\infty(\mathbf{A},\mathbf{S})$ and $\mathbf{Y}_0$ decides $\mathbf{A}$.

(C2) For every $n \geqslant 1$ we have that $\mathbf{Y}_n \in \mathrm{Str}_\infty(n_0+n,\mathbf{Y}_{n-1})$. Moreover, $\mathbf{Y}_n$ decides every $\mathbf{B} \in \mathrm{Str}_{<\infty}(\mathbf{A},\mathbf{Y}_{n-1})$ with $\mathrm{depth}_{\mathbf{Y}_{n-1}}(\mathbf{B}) = n_0+n$.

The above construction can be easily carried out using part (d) of Fact 3.18. By condition (C1), there exists $\mathbf{Y} \in \mathrm{Str}_\infty(\mathbf{S})$ such that $\mathbf{Y} \in \mathrm{Str}_\infty(n_0+n,\mathbf{Y}_n)$ for every $n \in \mathbb{N}$. Using conditions (C1) and (C2) and Fact 3.18, we see that $\mathbf{Y}$ is as desired. The proof of Lemma 3.19 is completed. □

The heart of the proof of Theorem 3.15 lies in the following lemma.

LEMMA 3.20. *Let* $\mathbf{T}$*,* $\mathcal{F}$*,* $\mathbf{S}$ *and* $\mathbf{A}$ *be as in Definition* 3.17*. Assume that* $\mathbf{S}$ *rejects* $\mathbf{A}$*. Then there is* $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A},\mathbf{S})$ *such that* $\mathbf{R}$ *rejects every* $\mathbf{B} \in \mathrm{Str}_{<\infty}(\mathbf{A},\mathbf{R})$*.*

PROOF. By Lemma 3.19, there exists $\mathbf{Y} \in \mathrm{Str}_\infty(\mathbf{A},\mathbf{S})$ which decides every $\mathbf{B} \in \mathrm{Str}_{<\infty}(\mathbf{A},\mathbf{Y})$. Since $\mathbf{S}$ rejects $\mathbf{A}$, by Fact 3.18, we see that $\mathbf{Y}$ also rejects $\mathbf{A}$. We set $n_0 = h(\mathbf{A})$. Recursively, we will select a sequence $(\mathbf{R}_n)$ in $\mathrm{Str}_\infty(\mathbf{A},\mathbf{Y})$ such that the following conditions are satisfied for every $n \in \mathbb{N}$.

(C1) Setting
$$\mathcal{A}_n = \big\{\mathbf{A}' \in \mathrm{Str}_{<\infty}(\mathbf{A},\mathbf{R}_n) : \mathrm{depth}_{\mathbf{R}_n}(\mathbf{A}') = n_0+n\big\}, \tag{3.10}$$
we have that $\mathbf{R}_n$ rejects every $\mathbf{A}' \in \mathcal{A}_n$.

(C2) We have $\mathbf{R}_{n+1} \in \mathrm{Str}_\infty(n_0+n,\mathbf{R}_n)$.

(C3) Let $\mathcal{A}_n$ be as in (3.10). Then, setting
$$\mathcal{B}_{n+1} = \bigcup_{\mathbf{A}' \in \mathcal{A}_n} \big\{\mathbf{B} \in \mathrm{Str}_{<\infty}(\mathbf{A}',\mathbf{R}_{n+1}) : h(\mathbf{B}) = h(\mathbf{A}')+1\big\}, \tag{3.11}$$
we have that $\mathbf{R}_{n+1}$ rejects every $\mathbf{B} \in \mathcal{B}_{n+1}$.

Assuming that the above selection has been carried out, the proof of the lemma is completed as follows. First we observe that, by condition (C2), there exists a unique $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A},\mathbf{S})$ such that $\mathbf{R} \in \mathrm{Str}_\infty(n_0+n,\mathbf{R}_n)$ for every $n \in \mathbb{N}$. We claim that $\mathbf{R}$ is as desired. Indeed, let $\mathbf{A}' \in \mathrm{Str}_{<\infty}(\mathbf{A},\mathbf{R})$ be arbitrary. Then $\mathrm{depth}_{\mathbf{R}}(\mathbf{A}') = n_0+n$ for some $n \in \mathbb{N}$. Since $\mathbf{R} \in \mathrm{Str}_\infty(n_0+n,\mathbf{R}_n)$ we have $\mathbf{A}' \in \mathcal{A}_n$. Hence, by condition (C1) and Fact 3.18, we conclude that $\mathbf{R}$ rejects $\mathbf{A}'$.

It remains to carry out the recursive selection. First we set $\mathbf{R}_0 = Y$. Since $\mathcal{A}_0 = \{\mathbf{A}\}$ and $\mathbf{Y}$ rejects $\mathbf{A}$, we see that with this choice condition (C1) is satisfied. The other conditions are superfluous for "$n = 0$" and so the first step of the recursive selection is completed. Let $n \in \mathbb{N}$ and assume that $\mathbf{R}_0,\dots,\mathbf{R}_n$ have been selected so that the above conditions are satisfied. Recall that $\mathbf{Y}$ decides every $\mathbf{B} \in \mathrm{Str}_{<\infty}(\mathbf{A},\mathbf{Y})$. Since $\mathbf{R}_n \in \mathrm{Str}_\infty(\mathbf{A},\mathbf{Y})$, by Fact 3.18, we see that every

$\mathbf{B} \in \mathrm{Str}_{<\infty}(\mathbf{A}, \mathbf{R}_n)$ is either accepted or rejected by $\mathbf{R}_n$. Hence, by Corollary 3.14 and Fact 3.18, we may select $\mathbf{Z} \in \mathrm{Str}_\infty(n_0 + n, \mathbf{R}_n)$ such that for every $\mathbf{A}' \in \mathcal{A}_n$ we have that either: (i) $\mathbf{Z}$ accepts every $\mathbf{B} \in \mathrm{Str}_{<\infty}(\mathbf{A}', \mathbf{Z})$ with $h(\mathbf{B}) = h(\mathbf{A}') + 1$, or (ii) $\mathbf{Z}$ rejects every $\mathbf{B} \in \mathrm{Str}_{<\infty}(\mathbf{A}', \mathbf{Z})$ with $h(\mathbf{B}) = h(\mathbf{A}') + 1$. We claim that alternative (ii) is satisfied for every $\mathbf{A}' \in \mathcal{A}_n$. Indeed, let $\mathbf{A}' \in \mathcal{A}_n$ and assume that for $\mathbf{A}'$ the first alternative holds true. Then for every $\mathbf{X} \in \mathrm{Str}_\infty(\mathbf{A}', \mathbf{Z})$ we have that $\mathbf{Z}$ accepts $\mathbf{X} \upharpoonright (h(\mathbf{A}') + 1)$ which implies, in particular, that $\mathbf{X} \in \mathcal{F}$. Thus we see that $\mathrm{Str}_\infty(\mathbf{A}', \mathbf{Z}) \subseteq \mathcal{F}$, which is equivalent to saying that $\mathbf{Z}$ accepts $\mathbf{A}'$. Since $\mathbf{Z} \in \mathrm{Str}_\infty(\mathbf{R}_n)$, this contradicts our inductive assumption that $\mathbf{R}_n$ rejects $\mathbf{A}'$.

We set $\mathbf{R}_{n+1} = \mathbf{Z}$ and we claim that with this choice conditions (C1), (C2) and (C3) are satisfied. To this end notice, first, that $\mathbf{R}_{n+1} \in \mathrm{Str}_\infty(n_0 + n, \mathbf{R}_n)$. Moreover, by the discussion in the previous paragraph, we have that $\mathbf{R}_{n+1}$ rejects every member of $\mathcal{B}_{n+1}$. Therefore, we only need to check that condition (C1) is satisfied. Let $\mathbf{A}' \in \mathcal{A}_{n+1}$ be arbitrary. Looking at the initial vector subtree of $\mathbf{A}'$ of height $h(\mathbf{A}') - 1$, we see that there exists a unique $i \in [n+1]$ such that $\mathbf{A}'$ belongs to $\mathcal{B}_i$. If $i = n + 1$, then we are done. Otherwise, by our inductive assumptions, we have that $\mathbf{R}_i$ rejects $\mathbf{A}'$. Since $\mathbf{R}_{n+1} \in \mathrm{Str}_\infty(\mathbf{R}_i)$, by Fact 3.18, we conclude that $\mathbf{R}_{n+1}$ rejects $\mathbf{A}'$. The recursive selection is completed, and as we have already indicated, the proof of Lemma 3.20 is also completed. □

We are now ready to give the proof of Theorem 3.15.

Proof of Theorem 3.15. We emphasize that in what follows all topological notions refer to the Ellentuck topology. For every subset $\mathcal{X}$ of $\mathrm{Str}_\infty(\mathbf{T})$ by $\overline{\mathcal{X}}$ and $\mathrm{Int}(\mathcal{X})$ we denote the closure and the interior of $\mathcal{X}$ respectively.

The proof is based on a series of claims. We start with the following.

Claim 3.21. *If $\mathcal{X}$ is completely Ramsey, then $\mathcal{X} \setminus \mathrm{Int}(\mathcal{X})$ is nowhere dense.*

Proof of Claim 3.21. Assume not. It is then possible to find $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ and $\mathbf{A} \in \mathrm{Str}_{<\infty}(\mathbf{S})$ such that $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{S}) \subseteq \overline{\mathcal{X} \setminus \mathrm{Int}(\mathcal{X})}$. The set $\mathcal{X}$ is completely Ramsey, and so there exists $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A}, \mathbf{S})$ such that either $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \subseteq \mathcal{X}$ or $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \cap \mathcal{X} = \emptyset$. If the second alternative holds true, then $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \cap \overline{\mathcal{X}} = \emptyset$ which implies that $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \cap \overline{\mathcal{X} \setminus \mathrm{Int}(\mathcal{X})} = \emptyset$. Note that this is impossible, since $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \subseteq \mathrm{Str}_\infty(\mathbf{A}, \mathbf{S}) \subseteq \overline{\mathcal{X} \setminus \mathrm{Int}(\mathcal{X})}$. It follows that the first alternative must be satisfied, that is, $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \subseteq \mathcal{X}$. But then $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \subseteq \mathrm{Int}(\mathcal{X})$ which also implies that $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \cap \overline{\mathcal{X} \setminus \mathrm{Int}(\mathcal{X})} = \emptyset$. Having arrived to a contradiction, the proof of Claim 3.21 is completed. □

We proceed with the following crucial claim.

Claim 3.22. *Every open set is completely Ramsey.*

Proof of Claim 3.22. Fix an open subset $\mathcal{O}$ of $\mathrm{Str}_\infty(\mathbf{T})$. Let $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ and $\mathbf{A} \in \mathrm{Str}_{<\infty}(\mathbf{S})$ be arbitrary. We need to find $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A}, \mathbf{S})$ such that either $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \subseteq \mathcal{O}$ or $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \cap \mathcal{O} = \emptyset$.

We set $n_0 = h(\mathbf{A})$. By Fact 3.18, we may assume that $\mathbf{A} = \mathbf{S} \upharpoonright n_0$ and that $\mathbf{S}$ decides $\mathbf{A}$ relative to $\mathcal{O}$. If $\mathbf{S}$ accepts $\mathbf{A}$ into $\mathcal{O}$, then we are done. Otherwise, by Lemma 3.20, there exists $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A},\mathbf{S})$ such that $\mathbf{R}$ rejects every $\mathbf{B} \in \mathrm{Str}_{<\infty}(\mathbf{A},\mathbf{R})$ from $\mathcal{O}$. We will show that $\mathrm{Str}_\infty(\mathbf{A},\mathbf{R}) \cap \mathcal{O} = \emptyset$. Indeed, let $\mathbf{X} \in \mathrm{Str}_\infty(\mathbf{A},\mathbf{R})$ be arbitrary. Also let $n \in \mathbb{N}$ and set $\mathbf{B}_n = \mathbf{X} \upharpoonright (n_0 + n)$. Notice that $\mathbf{R}$ rejects $\mathbf{B}_n$ from $\mathcal{O}$ since $\mathbf{B}_n \in \mathrm{Str}_{<\infty}(\mathbf{A},\mathbf{R})$. This implies, in particular, that $\mathrm{Str}_\infty(n_0+n,\mathbf{X}) = \mathrm{Str}_\infty(\mathbf{B}_n,\mathbf{X}) \nsubseteq \mathcal{O}$ and so there is $\mathbf{Y}_n \in \mathrm{Str}_\infty(n_0+n,\mathbf{X})$ such that $\mathbf{Y}_n \notin \mathcal{O}$. In this way we select a sequence $(\mathbf{Y}_n)$ in the complement of $\mathcal{O}$ such that $\mathbf{Y}_n \in \mathrm{Str}_\infty(n_0+n,\mathbf{X})$ for every $n \in \mathbb{N}$. The family $\{\mathrm{Str}_\infty(n+n_0,\mathbf{X}) : n \in \mathbb{N}\}$ is a neighborhood basis for $\mathbf{X}$. Hence, the sequence $(\mathbf{Y}_n)$ converges to $\mathbf{X}$. Since the complement of $\mathcal{O}$ is closed, we conclude that $\mathbf{X} \notin \mathcal{O}$ which implies, of course, that $\mathrm{Str}_\infty(\mathbf{A},\mathbf{R}) \cap \mathcal{O} = \emptyset$. The proof of Claim 3.22 is completed. □

It is convenient to introduce the following terminology. We say that a subset $\mathcal{N}$ of $\mathrm{Str}_\infty(\mathbf{T})$ is *Ramsey null* if for every $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ and every $\mathbf{A} \in \mathrm{Str}_{<\infty}(\mathbf{S})$ there exists $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A},\mathbf{S})$ such that $\mathrm{Str}_\infty(\mathbf{A},\mathbf{R}) \cap \mathcal{N} = \emptyset$.

Also recall that a family $\mathcal{I}$ of subsets of a set X is called a *σ-ideal* provided that: (i) for every $A \in \mathcal{I}$ and every $B \subseteq A$ we have $B \in \mathcal{I}$, and (ii) for every sequence (A_n) in $\mathcal{I}$ we have $\bigcup_n A_n \in \mathcal{I}$.

Claim 3.23. *The family of Ramsey null subsets of* $\mathrm{Str}_\infty(\mathbf{T})$ *is a σ-ideal.*

Proof of Claim 3.23. It is easy to see that if $\mathcal{N}$ is Ramsey null, then so is every subset of $\mathcal{N}$. Therefore, it is enough to prove that the family of Ramsey null subsets of $\mathrm{Str}_\infty(\mathbf{T})$ is closed under countable unions. To this end, let $(\mathcal{N}_n)$ be a sequence of Ramsey null sets and denote by $\mathcal{N}$ their union. Also let $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ and $\mathbf{A} \in \mathrm{Str}_{<\infty}(\mathbf{S})$ be arbitrary. We set $n_0 = h(\mathbf{A})$. Notice that we may assume that $\mathbf{S} \upharpoonright n_0 = \mathbf{A}$. We need to find $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A},\mathbf{S})$ such that $\mathrm{Str}_\infty(\mathbf{A},\mathbf{R}) \cap \mathcal{N} = \emptyset$. Recursively and using our assumption that every set $\mathcal{N}_n$ is Ramsey null, we select a sequence $(\mathbf{R}_n)$ in $\mathrm{Str}_\infty(\mathbf{A},\mathbf{S})$ with $\mathbf{R}_0 = \mathbf{S}$ and satisfying the following conditions for every $n \in \mathbb{N}$.

(C1) We have $\mathbf{R}_{n+1} \in \mathrm{Str}_\infty(n_0+n,\mathbf{R}_n)$.
(C2) For every $\mathbf{B} \in \mathrm{Str}_{<\infty}(\mathbf{A},\mathbf{R}_n)$ with $\mathrm{depth}_{\mathbf{R}_n}(\mathbf{B}) = n_0 + n$ we have that $\mathrm{Str}_\infty(\mathbf{B},\mathbf{R}_{n+1}) \cap \mathcal{N}_n = \emptyset$.

By (C1), there exists $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A},\mathbf{S})$ such that $\mathbf{R} \in \mathrm{Str}_\infty(n_0+n,\mathbf{R}_{n+1})$ for every $n \in \mathbb{N}$. We will show that $\mathrm{Str}_\infty(\mathbf{A},\mathbf{R}) \cap \mathcal{N} = \emptyset$. Indeed, let $\mathbf{X} \in \mathrm{Str}_\infty(\mathbf{A},\mathbf{R})$. Also let $n \in \mathbb{N}$ be arbitrary and set $\mathbf{C} = \mathbf{X} \upharpoonright (n_0+n)$. By our construction, we have $\mathrm{Str}_\infty(\mathbf{C},\mathbf{R}_{n+1}) \cap \mathcal{N}_n = \emptyset$. This yields, in particular, that $\mathbf{X} \notin \mathcal{N}_n$. Since n was arbitrary, we see that $\mathbf{X} \notin \mathcal{N}$ which implies, of course, that $\mathrm{Str}_\infty(\mathbf{A},\mathbf{R}_\infty) \cap \mathcal{N} = \emptyset$. The proof of Claim 3.23 is completed. □

The following claim is the final step of the argument.

Claim 3.24. *A subset of* $\mathrm{Str}_\infty(\mathbf{T})$ *is meager if and only if it is Ramsey null.*

Proof of Claim 3.24. Let $\mathcal{M}$ be a meager subset of $\mathrm{Str}_\infty(\mathbf{T})$. We will show that $\mathcal{M}$ is Ramsey null. By Claim 3.23, we may assume that $\mathcal{M}$ is nowhere dense. Let $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ and $\mathbf{A} \in \mathrm{Str}_{<\infty}(\mathbf{S})$ be arbitrary. By Claim 3.22, the set $\overline{\mathcal{M}}$ is completely Ramsey. Hence, there exists $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A}, \mathbf{S})$ such that either $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \subseteq \overline{\mathcal{M}}$ or $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \cap \overline{\mathcal{M}} = \emptyset$. Since $\mathcal{M}$ is nowhere dense, we have $\mathrm{Int}\big(\overline{\mathcal{M}}\big) = \emptyset$. This implies that $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \cap \overline{\mathcal{M}} = \emptyset$ and so $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \cap \mathcal{M} = \emptyset$.

Conversely, let $\mathcal{N}$ be a Ramsey null subset of $\mathrm{Str}_\infty(\mathbf{T})$. We will show that $\mathcal{N}$ is nowhere dense in $\mathrm{Str}_\infty(\mathbf{T})$. Indeed, if not, then there exist $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ and $\mathbf{A} \in \mathrm{Str}_{<\infty}(\mathbf{S})$ such that $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{S}) \subseteq \overline{\mathcal{N}}$. Since $\mathcal{N}$ is Ramsey null, we may find $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A}, \mathbf{S})$ such that $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \cap \mathcal{N} = \emptyset$. This, of course, implies that $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \cap \overline{\mathcal{N}} = \emptyset$, a contradiction. The proof of Claim 3.24 is completed. $\square$

We are now ready to complete the proof of Theorem 3.15. First we observe that, by Claim 3.21, every completely Ramsey set has the Baire property. Conversely assume that $\mathcal{X}$ has the Baire property and write $\mathcal{X} = \mathcal{O} \triangle \mathcal{M}$ where $\mathcal{O}$ is open and $\mathcal{X}$ is meager. Let $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ and $\mathbf{A} \in \mathrm{Str}_{<\infty}(\mathbf{S})$ be arbitrary. By Claim 3.24, there exists $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{A}, \mathbf{S})$ such that $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{R}) \cap \mathcal{M} = \emptyset$. Next, by Claim 3.22, we may select $\mathbf{W} \in \mathrm{Str}_\infty(\mathbf{A}, \mathbf{R})$ such that either $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{W}) \subseteq \mathcal{O}$ or $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{W}) \cap \mathcal{O} = \emptyset$. The first case implies that $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{W}) \subseteq \mathcal{X}$ while the second case yields that $\mathrm{Str}_\infty(\mathbf{A}, \mathbf{W}) \cap \mathcal{X} = \emptyset$. Therefore, the set $\mathcal{X}$ is completely Ramsey, and so, the entire proof of Theorem 3.15 is completed. $\square$

3.2.3. Applications. Let $\mathbf{T}$ be a rooted, pruned and finitely branching vector tree. Also let k be a positive integer. With every finite coloring c of $\binom{\mathbb{N}}{k}$ we associate a finite coloring C of $\mathrm{Str}_k(\mathbf{T})$ defined by the rule

$$C(\mathbf{A}) = c\big(L_{\mathbf{T}}(\mathbf{A})\big).$$

Moreover, notice that for every $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ the map

$$\mathrm{Str}_k(\mathbf{S}) \ni \mathbf{A} \mapsto L_{\mathbf{T}}(\mathbf{A}) \in \binom{L_{\mathbf{T}}(\mathbf{S})}{k}$$

is a surjection. Taking into account these remarks, we see that Theorem 3.12 implies the infinite version of Ramsey's classical theorem [**Ra**]. Using essentially the same arguments, we also see that Theorem 3.15 implies Ellentuck's theorem [**E**] on definable partitions of infinite subsets of $\mathbb{N}$.

The next result is an extension of Theorem 3.12 in the spirit of the work of Nash-Williams [**NW**] and Galvin [**Ga1**].

Corollary 3.25. *Let $\mathbf{T}$ be a rooted, pruned and finitely branching vector tree, and $\mathcal{F}$ a family of vector strong subtrees of $\mathbf{T}$ of finite height. Then there exists $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ such that either:* (i) $\mathrm{Str}_{<\infty}(\mathbf{S}) \cap \mathcal{F} = \emptyset$, *or* (ii) *for every $\mathbf{R} \in \mathrm{Str}_\infty(\mathbf{S})$ there exists $n \in \mathbb{N}$ such that $\mathbf{R} \upharpoonright n \in \mathcal{F}$.*

Proof. We set $\mathcal{O} = \bigcup_{\mathbf{A} \in \mathcal{F}} \mathrm{Str}_\infty(\mathbf{A}, \mathbf{T})$ and we observe that $\mathcal{O}$ is open both in the Ellentuck and in the metric topology of $\mathrm{Str}_\infty(\mathbf{T})$. By Theorem 3.15, there exists $\mathbf{S} \in \mathrm{Str}_\infty(\mathbf{T})$ such that either $\mathrm{Str}_\infty(\mathbf{S}) \cap \mathcal{O} = \emptyset$ or $\mathrm{Str}_\infty(\mathbf{S}) \subseteq \mathcal{O}$. Notice that every vector strong subtree of $\mathbf{S}$ of finite height is the initial vector subtree of some vector strong subtree of $\mathbf{S}$ of infinite height. Therefore, the first alternative is equivalent to saying that $\mathrm{Str}_{<\infty}(\mathbf{S}) \cap \mathcal{F} = \emptyset$. The proof of Corollary 3.25 is thus completed. $\square$

Our last application deals with colorings of infinite chains of dyadic trees. We will use as a model the tree $[2]^{<\mathbb{N}}$, henceforth denoted for simplicity by D. Given a subset S of D, we denote by $\mathrm{Chains}_\infty(S)$ the set of all infinite chains of D which are contained in S. Notice that the set $\mathrm{Chains}_\infty(S)$ can be identified with the set of all sequences (s_n) in S such that $s_n \sqsubset s_{n+1}$ for every $n \in \mathbb{N}$. In particular, if D is endowed with the discrete topology and $D^{\mathbb{N}}$ with the product topology, then $\mathrm{Chains}_\infty(D)$ is a closed (hence, Polish) subspace $D^{\mathbb{N}}$. All topological notions below refer to the relative topology on $\mathrm{Chains}_\infty(D)$.

Now let $\mathcal{C}$ be a definable subset of $\mathrm{Chain}_\infty(D)$. We address the question whether there exists a "nice" subtree S of D such that either $\mathrm{Chains}_\infty(S) \subseteq \mathcal{C}$ or $\mathrm{Chains}_\infty(S) \cap \mathcal{C} = \emptyset$. It is initially unsettling to observe that this problem has a negative answer in the category of strong subtrees of D. Indeed, color an infinite chain (s_n) of D red if $s_0{}^\frown 1 \sqsubseteq s_1$; otherwise, color it blue. Notice that this is a clopen partition of $\mathrm{Chains}_\infty(D)$. However, every strong subtree of D of infinite height contains chains of both colors.

It turns out that the right category for studying this problem is that of *regular dyadic subtrees* of D (see [**Ka2**]). Recall that a subtree R of D is said to be regular dyadic provided that: (i) every $t \in R$ has exactly two immediate successors in R, and (ii) for every $n \in \mathbb{N}$ there exists $m \in \mathbb{N}$ such that the n-level $R(n)$ of R is contained in $T(m)$.

We have the following theorem, essentially due to Stern [**St**] (see also [**Paw**]).

THEOREM 3.26. *Let $\mathcal{C} \subseteq \mathrm{Chains}_\infty(D)$ be a C-set. Then there exists a regular dyadic subtree R of D such that either $\mathrm{Chains}_\infty(R) \subseteq \mathcal{C}$ or $\mathrm{Chains}_\infty(R) \cap \mathcal{C} = \emptyset$.*

PROOF. In spite of the examples mentioned above, we will reduce the result to Corollary 3.16. To this end, first we will "extend" $\mathcal{C}$ to a color of $\mathrm{Str}_\infty(D)$. Specifically, for every $S \in \mathrm{Str}_\infty(D)$ let (s_n) be the *leftmost branch* of S. (Recall that the leftmost branch of S is the unique sequence (s_n) such that for every $n \in \mathbb{N}$ the node s_n is the lexicographically least element of the n-level $S(n)$ of S.) Notice that the leftmost branch of S is an infinite chain of D. Also observe that the map $\Phi\colon \mathrm{Str}_\infty(D) \to \mathrm{Chains}_\infty(D)$ which assigns to each $S \in \mathrm{Str}_\infty(D)$ its leftmost branch, is continuous when $\mathrm{Str}_\infty(D)$ is equipped with the metric topology. By Proposition C.5, we see that $\Phi^{-1}(\mathcal{C})$ is a C-set with respect to the metric topology of $\mathrm{Str}_\infty(D)$. Therefore, by Corollary 3.16, there exists a strong subtree S of D of infinite height such that either $\mathrm{Str}_\infty(S) \subseteq \Phi^{-1}(\mathcal{C})$ or $\mathrm{Str}_\infty(S) \cap \Phi^{-1}(\mathcal{C}) = \emptyset$. Notice that this monochromatic subtree is not the desired one since the image of $\mathrm{Str}_\infty(S)$ under the map Φ is not onto $\mathrm{Chains}_\infty(S)$.

However, this is not a serious problem and can be easily bypassed if we appropriately "trim" the tree S. Specifically, let

$$R_0 = \{\emptyset\} \cup \bigcup_{n=1}^{\infty} \big\{(a_0, \dots, a_{2n-1}) \in [2]^{2n} : a_i = 1 \text{ if } i \text{ is even}\big\}. \tag{3.12}$$

Also let $\mathrm{I}_S\colon D \to S$ be the canonical isomorphism associated with the homogeneous tree S (see Definition 1.19) and define $R = \mathrm{I}_S(R_0)$. Observe that R is a regular dyadic subtree of D. Moreover, note that every infinite chain of R is the leftmost branch of a strong subtree of S. It follows, in particular, that $\mathrm{Chains}_\infty(R) \subseteq \mathcal{C}$ if $\mathrm{Str}_\infty(S) \subseteq \Phi^{-1}(\mathcal{C})$, while $\mathrm{Chains}_\infty(R) \cap \mathcal{C} = \emptyset$ if $\mathrm{Str}_\infty(S) \cap \Phi^{-1}(\mathcal{C}) = \emptyset$. The proof of Theorem 3.26 is completed. □

3.3. Homogeneous trees

In this section we study colorings of strong subtrees of a homogeneous tree of finite, but sufficiently large, height. Our main objective is to obtain quantitative information for the corresponding "Milliken numbers" for this important special class of trees.

To this end we will not rely on the strong subtree version of the Halpern–Läuchli theorem as we did in Section 3.2. Instead, we will use the following finite version of Theorem 3.1 which is, essentially, a consequence of the Hales–Jewett theorem. The relation between the Hales–Jewett theorem and the Halpern–Läuchli theorem for homogeneous trees is well understood and can be traced back to the work of Carlson and Simpson [**CS**].

PROPOSITION 3.27. *For every integer $d \geqslant 1$, every $b_1, \dots, b_d \in \mathbb{N}$ with $b_i \geqslant 2$ for all $i \in [d]$ and every pair ℓ, r of positive integers there exists a positive integer N with the following property. If $\mathbf{T} = (T_1, \dots, T_d)$ is a vector homogeneous tree with $b_{T_i} = b_i$ for all $i \in [d]$ and $h(\mathbf{T}) \geqslant N$, then for every r-coloring of $\otimes\mathbf{T}$ there exists $\mathbf{S} \in \mathrm{Str}_\ell(\mathbf{T})$ such that the level product $\otimes\mathbf{S}$ of $\mathbf{S}$ is monochromatic. The least positive integer with this property will be denoted by $\mathrm{HL}(b_1, \dots, b_d \,|\, \ell, r)$.*

Moreover, there exists a primitive recursive function $\phi\colon \mathbb{N}^3 \to \mathbb{N}$ belonging to the class $\mathcal{E}^5$ such that

$$\mathrm{HL}(b_1, \dots, b_d \,|\, \ell, r) \leqslant \phi\Big(\prod_{i=1}^{d} b_i, \ell, r\Big) \tag{3.13}$$

for every $d \geqslant 1$, every $b_1, \dots, b_d \geqslant 2$ and every $\ell, r \geqslant 1$.

PROOF. We have already pointed out that the result is a consequence of the Hales–Jewett theorem, but we will give a streamlined proof using Proposition 2.25 instead. Specifically, we will show that

$$\mathrm{HL}(b_1, \dots, b_d \,|\, \ell, r) \leqslant \mathrm{GR}\Big(\prod_{i=1}^{d} b_i, \ell, 1, r\Big). \tag{3.14}$$

Clearly, by Proposition 2.25, this is enough to complete the proof.

To see that the estimate in (3.14) is satisfied, fix the "dimension" d and the parameters $b_1, \dots, b_d, \ell, r$. Set $N = \mathrm{GR}\big(\prod_{i=1}^{d} b_i, \ell, 1, r\big)$ and let $\mathbf{T} = (T_1, \dots, T_d)$ be a vector homogeneous tree with $b_{T_i} = b_i$ for all $i \in [d]$ and $h(\mathbf{T}) \geqslant N$. Notice that we may assume that $T_i = [b_i]^{<N}$ for every $i \in [d]$.

The main observation of the proof is that we can "code" the level product of $\mathbf{T}$ with words over the alphabet $\mathbb{A} = [b_1] \times \dots \times [b_d]$. Specifically, for every $i \in [d]$ let

$\pi_i\colon \mathbb{A} \to [b_i]$ be the natural projection. We extend π_i to a map $\bar{\pi}_i\colon \mathbb{A}^{<N} \to [b_i]^{<N}$ by the rule $\bar{\pi}_i(\emptyset) = \emptyset$ and $\bar{\pi}_i\big((a_0,\dots,a_{n-1})\big) = \big(\pi_i(a_0),\dots,\pi_i(a_{n-1})\big)$ for every $n \in [N-1]$ and every $a_0,\dots,a_{n-1} \in \mathbb{A}$. Finally, we define $\mathrm{I}\colon \mathbb{A}^{<N} \to \otimes\mathbf{T}$ by

$$\mathrm{I}(w) = \big(\bar{\pi}_1(w),\dots,\bar{\pi}_d(w)\big)$$

for every $w \in \mathbb{A}^{<N}$. It is easy to see that the map I is a bijection. Moreover, for every natural number $n < N$ we have $\mathrm{I}(\mathbb{A}^n) = [b_1]^n \times \cdots \times [b_d]^n$.

Now let $c\colon \otimes\mathbf{T} \to [r]$ be a coloring. We will associate with c an r-coloring C of $\mathrm{Subsp}_1(\mathbb{A}^N)$. First we define a map $\Phi\colon \mathrm{Subsp}_1(\mathbb{A}^N) \to \mathbb{A}^{<N}$ as follows. Let L be a combinatorial line of $\mathbb{A}^N$ and let X be its wildcard set. We select $w \in L$ and we set $\Phi(L) = w \upharpoonright \min(X) \in \mathbb{A}^{<N}$. Notice that $\Phi(L)$ is independent of the choice of w. Next, we define $C\colon \mathrm{Subsp}_1(\mathbb{A}^N) \to [r]$ by $C = c \circ \mathrm{I} \circ \Phi$. By the choice of N, there exists $W \in \mathrm{Subsp}_\ell(\mathbb{A}^N)$ such that the set $\mathrm{Subsp}_1(W)$ is monochromatic with respect to C. Let $X_0,\dots,X_{\ell-1}$ be the wildcard sets of W and set

$$S = \bigcup_{i=0}^{\ell-1} \{w \upharpoonright \min(X_i) : w \in W\}.$$

Also for every $i \in [d]$ let $S_i = \bar{\pi}_i(S)$ and set $\mathbf{S} = (S_1,\dots,S_d)$. Notice that $\Phi\big(\mathrm{Subsp}_1(W)\big) = S$, $\mathrm{I}(S) = \otimes\mathbf{S}$ and $\mathbf{S} \in \mathrm{Str}_\ell(\mathbf{T})$. Moreover, by the definition of the coloring C and the choice of W, the level product of $\mathbf{S}$ is monochromatic with respect to c. This shows that the estimate in (3.14) is satisfied, and the proof of Proposition 3.27 is completed. □

We are now ready to state the main result of this section.

Theorem 3.28. *For every integer $d \geqslant 1$, every $b_1,\dots,b_d \in \mathbb{N}$ with $b_i \geqslant 2$ for all $i \in [d]$ and every triple ℓ, k, r of positive integers with $\ell \geqslant k$ there exists a positive integer N with the following property. If $\mathbf{T} = (T_1,\dots,T_d)$ is a vector homogeneous tree with $b_{T_i} = b_i$ for all $i \in [d]$ and $h(\mathbf{T}) \geqslant N$, then for every r-coloring of $\mathrm{Str}_k(\mathbf{T})$ there exists $\mathbf{S} \in \mathrm{Str}_\ell(\mathbf{T})$ such that the set $\mathrm{Str}_k(\mathbf{S})$ is monochromatic. The least positive integer with this property will be denoted by $\mathrm{Mil}(b_1,\dots,b_d \,|\, \ell,k,r)$.*

Moreover, for every integer $d \geqslant 1$ the numbers $\mathrm{Mil}(b_1,\dots,b_d \,|\, \ell,k,r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^7$.

The proof of Theorem 3.28 proceeds by induction on k and follows the general scheme we discussed in Section 2.1. The main tool is Lemma 3.30 below, which will enable us to "simplify" a given finite coloring of $\mathrm{Str}_k(\mathbf{T})$ by passing to a vector strong subtree of $\mathbf{T}$ of sufficiently large height.

We start with the following lemma.

Lemma 3.29. *Let $d \geqslant 1$ and let $b_1,\dots,b_d,k,r,n,M$ be positive integers with $b_i \geqslant 2$ for all $i \in [d]$ and $n \geqslant k$. Also let $\mathbf{T} = (T_1,\dots,T_d)$ be a vector homogeneous tree such that $b_{T_i} = b_i$ for all $i \in [d]$ and $c\colon \mathrm{Str}_{k+1}(\mathbf{T}) \to [r]$ a coloring. We set*

$$q = q(b_1,\dots,b_d,k,n) = |\{\mathbf{A} \in \mathrm{Str}_k(\mathbf{T}) : \mathrm{depth}_{\mathbf{T}}(\mathbf{A}) = n\}|. \tag{3.15}$$

Assume that

$$h(\mathbf{T}) \geqslant n + \phi\Big(\prod_{i=1}^{d} b_i^{b_i^n}, M, r^q\Big) \tag{3.16}$$

where $\phi\colon \mathbb{N}^3 \to \mathbb{N}$ *is as in Proposition* 3.27. *Then there exists* $\mathbf{W} \in \mathrm{Str}_{n+M}(n,\mathbf{T})$ *such that the set* $\mathrm{Str}_{k+1}(\mathbf{A},\mathbf{W})$ *is monochromatic for every* $\mathbf{A} \in \mathrm{Str}_k(\mathbf{W})$ *with* $\mathrm{depth}_{\mathbf{W}}(\mathbf{A}) = n$.

Proof. It is similar to the proof of Lemma 3.13. Unfortunately the notation is more cumbersome since we need to work with vector trees. However, the general strategy is identical.

We proceed to the details. Let $i \in [d]$ be arbitrary and fix an enumeration $\{t_1^i, \dots, t_{b_i^n}^i\}$ of the n-level $T_i(n)$ of T_i. For every $j \in [b_i^n]$ let $S_j^i = \mathrm{Succ}_{T_i}(t_j^i)$ and set $\mathbf{S}_i = (S_1^i, \dots, S_{b_i^n}^i)$. Observe that $b_{S_j^i} = b_i$ for every $j \in [b_i^n]$. Also let

$$\mathbf{S} = (\mathbf{S}_1, \dots, \mathbf{S}_d) = (S_1^1, \dots, S_{b_i^n}^1, \dots, S_1^d, \dots, S_{b_d^n}^d)$$

and notice that

$$\begin{aligned} h(\mathbf{S}) = h(\mathbf{T}) - n \quad & \overset{(3.16)}{\geqslant} \quad \phi\Big(\prod_{i=1}^{d} b_i^{b_i^n}, M, r^q\Big) \\ & \overset{(3.13)}{\geqslant} \quad \mathrm{HL}\big(\underbrace{b_1, \dots, b_1}_{b_1^n-\text{times}}, \dots, \underbrace{b_d, \dots, b_d}_{b_d^n-\text{times}} \,|\, M, r^q\big). \end{aligned} \tag{3.17}$$

Next, for every $\mathbf{A} \in \mathrm{Str}_k(\mathbf{T})$ with $\mathrm{depth}_{\mathbf{T}}(\mathbf{A}) = n$ and every $\mathbf{s} \in \otimes\mathbf{S}$ we define $\mathbf{E}(\mathbf{A},\mathbf{s}) \in \mathrm{Str}_{k+1}(\mathbf{A},\mathbf{T})$ as follows. Write $\mathbf{A} = (A_1, \dots, A_d)$ and $\mathbf{s} = (\mathbf{s}_1, \dots, \mathbf{s}_d)$ where $\mathbf{s}_i = (s_1^i, \dots, s_{b_i^n}^i) \in \otimes\mathbf{S}_i$ for every $i \in [d]$. For every $i \in [d]$ and every $t \in A_i(k-1)$ let

$$I(t) = \big\{ j \in [b_i^n] : t_j^i \in \mathrm{ImmSucc}_{T_i}(t) \big\}$$

and set

$$E(A_i, \mathbf{s}_i) = A_i \cup \big\{ s_j^i : j \in I(t) \text{ for some } t \in A_i(k-1) \big\}.$$

Finally, we define $\mathbf{E}(\mathbf{A},\mathbf{s}) = \big(E(A_1,\mathbf{s}_1), \dots, E(A_d,\mathbf{s}_d)\big)$.

Now set $\mathcal{F} = \{\mathbf{A} \in \mathrm{Str}_k(\mathbf{T}) : \mathrm{depth}_{\mathbf{T}}(\mathbf{A}) = n\}$. Observe that $|\mathcal{F}| = q$ by the choice of q in (3.15). Also let $C\colon \otimes\mathbf{S} \to [r]^{\mathcal{F}}$ be defined by the rule

$$C(\mathbf{s}) = \big\langle c\big(\mathbf{E}(\mathbf{A},\mathbf{s})\big) : \mathbf{A} \in \mathcal{F} \big\rangle.$$

By (3.17), there exists a vector strong subtree $\mathbf{R}$ of $\mathbf{S}$ with $h(\mathbf{R}) = M$ such that $\otimes\mathbf{R}$ is monochromatic with respect to C. Notice that $\mathbf{R}$ is of the form

$$\mathbf{R} = (\mathbf{R}_1, \dots, \mathbf{R}_d) = (R_1^1, \dots, R_{b_i^n}^1, \dots, R_1^d, \dots, R_{b_d^n}^d)$$

where R_j^i is a strong subtree of S_j^i for every $i \in [d]$ and every $j \in [b_i^n]$.

For every $i \in [d]$ let

$$W_i = (T_i \upharpoonright n) \cup \bigcup_{j=1}^{b_i^n} R_j^i.$$

We set $\mathbf{W} = (W_1, \dots, W_d)$ and we claim that $\mathbf{W}$ is as desired. Indeed, first observe that $\mathbf{W} \in \mathrm{Str}_{n+M}(n,\mathbf{T})$. Next, let $\mathbf{A} \in \mathrm{Str}_k(\mathbf{T})$ with $\mathrm{depth}_{\mathbf{T}}(\mathbf{A}) = n$ be arbitrary.

Note that $\mathbf{A} \in \mathcal{F}$ and $\mathrm{Str}_{k+1}(\mathbf{A}, \mathbf{W}) = \{\mathbf{E}(\mathbf{A}, \mathbf{s}) : \mathbf{s} \in \otimes\mathbf{R}\}$. By the definition of C and the choice of $\mathbf{R}$, we conclude that $\mathrm{Str}_{k+1}(\mathbf{A}, \mathbf{W})$ is monochromatic with respect to the coloring c. The proof of Lemma 3.29 is completed. $\square$

We need to introduce some numerical invariants. For every positive integer d we define a function $g_d \colon \mathbb{N}^{d+4} \to \mathbb{N}$ recursively by the rule

$$\begin{cases} g_d(\mathbf{b}, \ell, k, r, 0) = 1, \\ g_d(\mathbf{b}, \ell, k, r, j+1) = \phi\Big(\prod_{i=1}^d b_i^{b_i^{|\ell-j-1|}}, g_d(\mathbf{b}, \ell, k, r, j), r^{2^{\sum_{i=1}^d b_i^\ell}}\Big) + 1 \end{cases} \tag{3.18}$$

where $\mathbf{b} = (b_1, \dots, b_d) \in \mathbb{N}^d$ with $b_i \geqslant 2$ for all $i \in [d]$, ℓ, k, r are positive integers and $\phi \colon \mathbb{N}^3 \to \mathbb{N}$ is as in Proposition 3.27. If $b_i \leqslant 1$ for some $i \in [d]$ or if some of the parameters ℓ, k, r happens to be zero, then we set $g_d(b_1, \dots, b_d, \ell, k, r, j) = 0$. Since ϕ belongs to the class $\mathcal{E}^5$, we see that the function g_d is upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^6$. Note that this bound is uniform with respect to d.

As we have already pointed out, the following lemma is the main step towards the proof of Theorem 3.28.

LEMMA 3.30. *Let $d \geqslant 1$ and let $b_1, \dots, b_d, \ell, k, r$ be positive integers with $b_i \geqslant 2$ for all $i \in [d]$ and $\ell \geqslant k+1$. Also let $\mathbf{T} = (T_1, \dots, T_d)$ be a vector homogeneous tree such that $b_{T_i} = b_i$ for all $i \in [d]$ and*

$$h(\mathbf{T}) = k - 1 + g_d(b_1, \dots, b_d, \ell, k, r, \ell - k). \tag{3.19}$$

Finally, let $c \colon \mathrm{Str}_{k+1}(\mathbf{T}) \to [r]$ be a coloring. Then there exists $\mathbf{S} \in \mathrm{Str}_\ell(\mathbf{T})$ such that $c(\mathbf{B}) = c(\mathbf{C})$ for every $\mathbf{B}, \mathbf{C} \in \mathrm{Str}_{k+1}(\mathbf{S})$ with $\mathbf{B} \upharpoonright k = \mathbf{C} \upharpoonright k$.

PROOF. For every $n \in \{k-1, \dots, \ell-1\}$ let $M_n = g_d(b_1, \dots, b_d, \ell, k, r, \ell-1-n)$. Notice that $1 = M_{\ell-1} \leqslant M_n \leqslant M_{k-1} = h(\mathbf{T}) - (k-1)$ and if $n \geqslant k$, then

$$M_{n-1} = \phi\Big(\prod_{i=1}^d b_i^{b_i^n}, M_n, r^{2^{\sum_{i=1}^d b_i^\ell}}\Big) + 1. \tag{3.20}$$

Recursively, we will select a sequence $(\mathbf{S}_n)_{n=k-1}^{\ell-1}$ of vector strong subtrees of $\mathbf{T}$ with $\mathbf{S}_{k-1} = \mathbf{T}$ such that the following conditions are satisfied.

(C1) We have $h(\mathbf{S}_n) = n + M_n$.
(C2) If $n \geqslant k$, then $\mathbf{S}_n \in \mathrm{Str}_{n+M_n}(n, \mathbf{S}_{n-1})$.
(C3) If $n \geqslant k$, then for every $\mathbf{A} \in \mathrm{Str}_k(\mathbf{S}_n)$ with $\mathrm{depth}_{\mathbf{S}_n}(\mathbf{A}) = n$ the set $\mathrm{Str}_{k+1}(\mathbf{A}, \mathbf{S}_n)$ is monochromatic.

Let $n \in \{k, \dots, \ell-1\}$ and assume that the vector trees $\mathbf{S}_{k-1}, \dots, \mathbf{S}_{n-1}$ have been selected. Set $q = |\{\mathbf{A} \in \mathrm{Str}_k(\mathbf{S}_{n-1}) : \mathrm{depth}_{\mathbf{S}_{n-1}}(\mathbf{A}) = n\}|$ and notice that

$$\sum_{i=1}^d b_i^\ell \geqslant \sum_{i=1}^d b_i^n \geqslant \log_2 q. \tag{3.21}$$

By Corollary A.4, we may assume that for every $b, m, r, r' \in \mathbb{N}$ with $r \geqslant r'$ we have $\phi(b, m, r) \geqslant \phi(b, m, r')$. Hence,

$$h(\mathbf{S}_{n-1}) \overset{\text{(C1)}}{=} (n-1) + M_{n-1} \overset{(3.20)}{=} n + \phi\Big(\prod_{i=1}^{d} b_i^{b_i^n}, M_n, r^{2^{\sum_{i=1}^{d} b_i^{\ell}}}\Big) \overset{(3.21)}{\geqslant} n + \phi\Big(\prod_{i=1}^{d} b_i^{b_i^n}, M_n, r^q\Big). \tag{3.22}$$

By Lemma 3.29, there exists a vector strong subtree $\mathbf{S}_n$ of $\mathbf{S}_{n-1}$ satisfying conditions (C1), (C2) and (C3). The recursive selection is completed.

We set $\mathbf{S} = \mathbf{S}_{\ell-1}$ and we claim that $\mathbf{S}$ is as desired. To this end notice, first, that $h(\mathbf{S}) = \ell - 1 + M_{\ell-1} = \ell$. Let $\mathbf{B}, \mathbf{C} \in \mathrm{Str}_{k+1}(\mathbf{S})$ with $\mathbf{B} \upharpoonright k = \mathbf{C} \upharpoonright k$ be arbitrary. Set $\mathbf{A} = \mathbf{B} \upharpoonright k = \mathbf{C} \upharpoonright k$ and $n = \mathrm{depth}_{\mathbf{S}}(\mathbf{A})$. Observe that $n \in \{k, \dots, \ell-1\}$. By condition (C2), we see that $\mathbf{S} \in \mathrm{Str}_{\ell}(n, \mathbf{S}_n)$. This implies, in particular, that $\mathbf{A} \in \mathrm{Str}_k(\mathbf{S}_n)$, $\mathrm{depth}_{\mathbf{S}_n}(\mathbf{A}) = n$ and $\mathbf{B}, \mathbf{C} \in \mathrm{Str}_{k+1}(\mathbf{A}, \mathbf{S}_n)$. By condition (C3), we conclude that $c(\mathbf{B}) = c(\mathbf{C})$ and the proof of Lemma 3.30 is completed. □

We are ready to give the proof of Theorem 3.28.

Proof of Theorem 3.28. Fix the positive integer d. Observe that

$$\mathrm{Mil}\big(b_1, \dots, b_d \,|\, \ell, 1, r\big) = \mathrm{HL}(b_1, \dots, b_d \,|\, \ell, r). \tag{3.23}$$

On the hand, by Lemma 3.30, we see that

$$\mathrm{Mil}(\mathbf{b} \,|\, \ell, k+1, r) \leqslant k - 1 + g_d\big(\mathbf{b}, \mathrm{Mil}(\mathbf{b} \,|\, \ell, k, r), k, r, \mathrm{Mil}(\mathbf{b} \,|\, \ell, k, r) - k\big) \tag{3.24}$$

for every $\mathbf{b} = (b_1, \dots, b_d) \in \mathbb{N}^d$ with $b_i \geqslant 2$ for all $i \in [d]$ and every triple ℓ, k, r of positive integers with $\ell \geqslant k + 1$. Now recall that the function g_d is upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^6$. Hence, by (3.23), (3.24) and Proposition 3.27, we conclude that the numbers $\mathrm{Mil}\big(b_1, \dots, b_d \,|\, \ell, k, r\big)$ are upper bounded by a primitive recursive function in the class $\mathcal{E}^7$. The proof of Theorem 3.28 is completed. □

3.4. Notes and remarks

3.4.1. Theorem 3.2 was discovered in 1966 as a result needed for the construction of a model of set theory in which the Boolean prime ideal theorem is true but not the full axiom of choice (see [**HLe**]). The original proof was based on tools from logic and somewhat later a second proof was found by Harrington (unpublished) using set theoretic techniques. Purely combinatorial proofs were given much later by Argyros, Felouzis and Kanellopoulos [**AFK**], and by Todorcevic [**To**].

Theorem 3.1 was first formulated in the late 1960s by Laver who also obtained in [**L**] an extension of this result that concerns partitions of products of infinitely many trees. Yet another proof of Theorem 3.1 was given by Milliken [**M2**] using methods developed by Halpern and Läuchli in [**HL**].

3.4.2. As we have already mentioned, all the results in Subsections 3.2.1 and 3.2.2 are due to Milliken. Milliken also addressed the natural problem of getting quantitative refinements of Theorem 3.12 (see, in particular, [**M2**, Section 5]). In this direction, the primitive recursive bounds obtained by Theorem 3.28 are new and encompass all cases considered by Milliken.

We also note that Theorem 3.15 is naturally placed in the general context of *topological Ramsey spaces*. This theory, initiated by Carlson [**C**] and further developed by Todorcevic [**To**], is an extension of the works of Galvin and Prikry [**GP**], and Ellentuck [**E**].

3.4.3. Besides the work of Milliken [**M2, M3**] and Stern [**St**], there is a large number of results in the literature dealing with Ramsey properties of trees and perfect sets of reals. Examples include the work of Galvin [**Ga2**] and Blass [**B1**] on colorings of finite subsets of the reals, the work of Louveau, Shelah and Veličković [**LSV**] on colorings of "rapidly increasing" sequences of reals, and the work of Kanellopoulos [**Ka2**] on colorings of "rapidly increasing" dyadic trees. These results are based on the strong subtree version of the Halpern–Läuchli theorem and can be derived from Milliken's tree theorem arguing as in the proof of Theorem 3.26.

CHAPTER 4

Variable words

4.1. Carlson's theorem

The central theme of this chapter is the study of Ramsey properties of sequences of variable words. Most of the results that we present can be roughly classified as infinite-dimensional extensions of the Hales–Jewett theorem and its consequences, though there are finite-dimensional phenomena in this context not covered by the analysis in Chapter 2 (see, in particular, Section 4.3). The present section is entirely devoted to the proof of the following theorem due to Carlson [**C**]. General facts about extracted variable words can be found in Section 1.4.

THEOREM 4.1. *Let A be a finite alphabet with $|A| \geqslant 2$ and $\mathbf{w} = (w_n)$ a sequence of variable words over A. Then for every finite coloring of the set* $\mathrm{EV}[\mathbf{w}]$ *of all extracted variable words of* $\mathbf{w}$ *there exists an extracted subsequence* $\mathbf{v} = (v_n)$ *of* $\mathbf{w}$ *such that the set* $\mathrm{EV}[\mathbf{v}]$ *is monochromatic.*

Carlson's theorem is, arguably, one of the finest results in Ramsey theory. It unifies and extends several strong results, including the Carlson–Simpson theorem, Hindman's theorem and many more. We present in detail a number of its consequences in Section 4.2.

We also note that no combinatorial proof of Carlson's theorem has been found so far[1] and all known proofs are based on the use of ultrafilters and/or methods from topological dynamics. Proofs of this sort were first discovered by Galvin and Glazer. This line of research was subsequently further developed by several authors and is now an active part of Ramsey theory. We review this theory, and in particular those tools needed for the proof of Carlson's theorem, in Appendix D.

PROOF OF THEOREM 4.1. We follow the proof from [**HS**]. We fix a letter x not belonging to A which we view as a variable, and we set $S = (A \cup \{x\})^{<\mathbb{N}}$. Notice that the set S equipped with the operation of concatenation is a semigroup. Therefore, by Proposition D.5, the space βS of all ultrafilters on S equipped with the binary operation $^\frown$ defined by

$$R \in \mathcal{V}^\frown\mathcal{W} \Leftrightarrow (\mathcal{V}v)(\mathcal{W}w)\ [v^\frown w \in R]$$

is a compact semigroup. We recall that a basic open set of βS is of the form $(R)_{\beta S} = \{\mathcal{V} \in \beta S : R \in \mathcal{V}\}$ for some $R \subseteq S$.

[1] We remark, however, that most of the consequences of Carlson's theorem can be proved by purely combinatorial means.

We fix a sequence $\mathbf{w} = (w_n)$ of variable words over A. For every $m \in \mathbb{N}$ let

$$C_m = \mathrm{E}[(w_n)_{n=m}^{\infty}], \quad V_m = \mathrm{EV}[(w_n)_{n=m}^{\infty}] \ \text{ and } \ S_m = C_m \cup V_m \tag{4.1}$$

and define

$$\gamma C = \bigcap_{m=0}^{\infty} (C_m)_{\beta S}, \quad \gamma V = \bigcap_{m=0}^{\infty} (V_m)_{\beta S} \ \text{ and } \ \gamma S = \bigcap_{m=0}^{\infty} (S_m)_{\beta S}. \tag{4.2}$$

We have the following claim.

Claim 4.2. *The following hold.*

(a) *The spaces γS and γC are compact subsemigroups of βS. Moreover, γV is a two-sided ideal of γS.*

(b) *For every $a \in A$ the semigroup homomorphism $S \ni v \mapsto v(a) \in S$ (with the convention that $v(a) = v$ if $v \in A^{<\mathbb{N}}$) is extended to a continuous homomorphism $T_a \colon \gamma S \to \gamma C$ which is the identity on γC.*

Proof of Claim 4.2. (a) As in Appendix D, for every $R \subseteq S$ by $\mathrm{C}\ell_{\beta S}(R)$ we denote the closure of the set $e_S(R) = \{e_S(r) : r \in R\}$ in βS. By (D.2), we have $\mathrm{C}\ell_{\beta S}(R) = (R)_{\beta S}$ for every $R \subseteq S$. Hence, the family $\{(S_m)_{\beta S} : m \in \mathbb{N}\}$ is a decreasing sequence of nonempty closed subsets of βS which implies that the set $\gamma S = \bigcap_{m=0}^{\infty} (S_m)_{\beta S}$ is a nonempty compact subset of βS. Arguing similarly we see that γC and γV are both nonempty compact subsets of βS.

We proceed to show that γS is a subsemigroup of βS. Let $\mathcal{V}, \mathcal{W} \in \gamma S$ be arbitrary. Notice that

$$\mathcal{V}^\frown\mathcal{W} \in \gamma S \Leftrightarrow (\forall m \in \mathbb{N}) \ [S_m \in \mathcal{V}^\frown\mathcal{W}]. \tag{4.3}$$

Therefore, it is enough to prove that $S_m \in \mathcal{V}^\frown\mathcal{W}$ for every $m \in \mathbb{N}$. We first observe that

$$(\forall m \in \mathbb{N})(\forall v \in S_m)(\exists \ell \in \mathbb{N}) \ [S_\ell \subseteq \{w \in S : v^\frown w \in S_m\}]. \tag{4.4}$$

Indeed, let $m \in \mathbb{N}$ and $v \in S_m$. Then $v \in \mathrm{E}[(w_n)_{n=m}^{\ell-1}] \cup \mathrm{EV}[(w_n)_{n=m}^{\ell-1}]$ for some $\ell \geqslant m+1$. Clearly $v^\frown S_\ell \subseteq S_m$ and so $S_\ell \subseteq \{w \in S : v^\frown w \in S_m\}$. Since $S_\ell \in \mathcal{W}$, we see that $\{w \in S : v^\frown w \in S_m\} \in \mathcal{W}$. Hence,

$$S_m \subseteq \{v \in S : \{w \in S : v^\frown w \in S_m\} \in \mathcal{W}\}.$$

This implies that $\{v \in S : \{w \in S : v^\frown w \in S_m\} \in \mathcal{W}\} \in \mathcal{V}$ which is equivalent to saying that $S_m \in \mathcal{V}^\frown\mathcal{W}$.

With identical arguments we see that γC is a compact subsemigroup of βS. Therefore, the proof of this part of the claim will be completed once we show that γV is a two-sided ideal of γS. So let $\mathcal{V} \in \gamma V$ and $\mathcal{W} \in \gamma S$. Arguing as in the proof of (4.4), we see that

$$(\forall m \in \mathbb{N})(\forall v \in V_m)(\exists \ell \in \mathbb{N}) \ [S_\ell \subseteq \{w \in S : v^\frown w \in V_m\}] \tag{4.5}$$

which is easily seen to imply that $V_m \in \mathcal{V}^\frown\mathcal{W}$ for every $m \in \mathbb{N}$. This shows, of course, that $\mathcal{V}^\frown\mathcal{W} \in \gamma V$. Conversely notice that

$$(\forall m \in \mathbb{N})(\forall w \in S_m)(\exists \ell \in \mathbb{N}) \ [V_\ell \subseteq \{v \in S : w^\frown v \in V_m\}]. \tag{4.6}$$

This yields that $V_m \in \mathcal{W}^\frown\mathcal{V}$ for every $m \in \mathbb{N}$ and so $\mathcal{W}^\frown\mathcal{V} \in \gamma V$. Thus γV is a two-sided ideal of γS.

(b) Fix $a \in A$. By Proposition D.9, there exists a unique continuous semigroup homomorphism $T_a \colon \beta S \to \beta S$ such that $T_a(v) = v(a)$ for every $v \in S$. Let $\mathcal{W} \in \gamma S$ and recall that $T_a(\mathcal{W}) = \{X \subseteq S : T_a^{-1}(X) \in \mathcal{W}\}$. Also let $m \in \mathbb{N}$ be arbitrary. It is easy to see that $S_m \subseteq T_a^{-1}(C_m)$. Since $S_m \in \mathcal{W}$, we have $T_a^{-1}(C_m) \in \mathcal{W}$. Hence, $C_m \in T_a(\mathcal{W})$ for every $m \in \mathbb{N}$ which implies that $T_a(\mathcal{W}) \in \gamma C$.

Now let $\mathcal{W} \in \gamma C$. We will show that $T_a(\mathcal{W}) = \mathcal{W}$. By the maximality of ultrafilters, it is enough to prove that $X \in T_a(\mathcal{W})$ for every $X \in \mathcal{W}$. To this end, let $X \in \mathcal{W}$. Set $Y = X \cap C_0$ and notice that $Y \in \mathcal{W}$. Moreover, $Y \subseteq T_a^{-1}(Y)$ and so $T_a^{-1}(Y) \in \mathcal{W}$ which is equivalent to saying that $Y \in T_a(\mathcal{W})$. Since $Y \subseteq X$ we conclude that $X \in T_a(\mathcal{W})$. The proof of Claim 4.2 is completed. $\square$

The following claim is the main step of the proof.

Claim 4.3. *There exists an idempotent $\mathcal{V} \in \gamma V$ such that for every $a \in A$ we have $T_a(\mathcal{V})^\frown\mathcal{V} = \mathcal{V}^\frown T_a(\mathcal{V}) = \mathcal{V}$.*

Proof of Claim 4.3. The space γC is a compact semigroup on its own. Hence, by Lemma D.11 and Proposition D.12, there exists a minimal idempotent $\mathcal{W}$ of γC. Note that $\mathcal{W}$ is also an idempotent of γS. Next recall that, by Claim 4.2, γV is a two-sided ideal of γS. Therefore, by Corollary D.15, there exists an idempotent $\mathcal{V} \in \gamma V$ with $\mathcal{V} \preccurlyeq \mathcal{W}$. This implies, in particular, that $\mathcal{W}^\frown\mathcal{V} = \mathcal{V}^\frown\mathcal{W} = \mathcal{V}$. Fix $a \in A$. Since $T_a(\mathcal{W}) = \mathcal{W}$ we have

$$T_a(\mathcal{V}) = T_a(\mathcal{W}^\frown\mathcal{V}) = T_a(\mathcal{W})^\frown T_a(\mathcal{V}) = \mathcal{W}^\frown T_a(\mathcal{V}) \tag{4.7}$$

and

$$T_a(\mathcal{V}) = T_a(\mathcal{V}^\frown\mathcal{W}) = T_a(\mathcal{V})^\frown T_a(\mathcal{W}) = T_a(\mathcal{V})^\frown\mathcal{W}. \tag{4.8}$$

Moreover,

$$T_a(\mathcal{V}) = T_a(\mathcal{V}^\frown\mathcal{V}) = T_a(\mathcal{V})^\frown T_a(\mathcal{V}). \tag{4.9}$$

It follows that $T_a(\mathcal{V})$ is an idempotent of γC and $T_a(\mathcal{V}) \preccurlyeq \mathcal{W}$. The ultrafilter $\mathcal{W}$ is a minimal idempotent of γC and so $T_a(\mathcal{V}) = \mathcal{W}$. Hence, $T_a(\mathcal{V})^\frown\mathcal{V} = \mathcal{W}^\frown\mathcal{V} = \mathcal{V}$ and $\mathcal{V}^\frown T_a(\mathcal{V}) = \mathcal{V}^\frown\mathcal{W} = \mathcal{V}$. The proof of Claim 4.3 is completed. $\square$

We proceed with the following claim.

Claim 4.4. *Let $\mathcal{V}$ be as in Claim 4.3. Also let $X \in \mathcal{V}$. Then there exists an extracted subsequence $\mathbf{v}$ of $\mathbf{w}$ such that $\mathrm{EV}[\mathbf{v}] \subseteq X$.*

Proof of Claim 4.4. Since $\mathcal{V} \in \gamma V$ and $\mathrm{EV}[\mathbf{w}] = \mathrm{EV}[(w_n)_{n=0}^\infty] = V_0 \in \mathcal{V}$, we may assume that $\mathcal{V}$ consists of subsets of $\mathrm{EV}[\mathbf{w}]$. Let $Y \in \mathcal{V}$ be arbitrary. By Claim 4.3, for every $a \in A$ we have $T_a(\mathcal{V})^\frown\mathcal{V} = \mathcal{V}^\frown T_a(\mathcal{V}) = \mathcal{V}$ which implies that

$$(\mathcal{V}v)(\mathcal{V}u)(\forall a \in A)\ [v(a)^\frown u \in Y \wedge v^\frown u(a) \in Y].$$

On the other hand, we have $\mathcal{V}^\frown\mathcal{V} = \mathcal{V}$ and so

$$(\mathcal{V}v)(\mathcal{V}u)\ [v \in Y \wedge u \in Y \wedge v^\frown u \in Y].$$

Since $\mathrm{EV}[(v,u)] = \{v(a)^\frown u : a \in A\} \cup \{v^\frown u(a) : a \in A\} \cup \{v, u, v^\frown u\}$ for every $v, u \in \mathrm{EV}[\mathbf{w}]$, we conclude that

$$(\mathcal{V}v)(\mathcal{V}u)\ \big[\mathrm{EV}[(v,u)] \subseteq Y\big]. \tag{4.10}$$

In what follows for every $v, u \in \mathrm{EV}[\mathbf{w}]$ we write $v < u$ if there exists a positive integer m such that $v \in \mathrm{EV}[(w_n)_{n=0}^{m-1}]$ and $u \in V_m = \mathrm{EV}[(w_n)_{n=m}^{\infty}]$.

Now let $X \in \mathcal{V}$. Recursively, we will select a decreasing sequence (X_n) of subsets of X and a sequence (v_n) in $\mathrm{EV}[\mathbf{w}]$ such that for every $n \in \mathbb{N}$ we have

(C1) $v_n \in X_n$, $X_n \in \mathcal{V}$, $v_n < v_{n+1}$ and $X_{n+1} = \{u \in V_0 : \mathrm{EV}[(v_n, u)] \subseteq X_n\}$.

First we set $X_0 = X$. By (4.10) applied for "$Y = X_0$", there exists $v_0 \in V_0$ such that, setting $X_1 = \{u \in V_0 : \mathrm{EV}[(v_0, u)] \subseteq X_0\}$, we have $X_1 \in \mathcal{V}$. Next, let n be a positive integer and assume that the sets $X_0, \dots, X_n$ and the variable words $v_0, \dots, v_{n-1}$ have been selected. Let $m \geqslant 1$ be such that $v_{n-1} \in \mathrm{EV}[(w_n)_{n=0}^{m-1}]$. Notice that $\mathrm{EV}[(w_n)_{n=m}^{\infty}] = V_m \in \mathcal{V}$. Therefore, by (4.10) applied for "$Y = X_n$", we may select $v_n \in V_m$ such that $\{u \in V_0 : \mathrm{EV}[(v_n, u)] \subseteq X_n\} \in \mathcal{V}$. Finally, let $X_{n+1} = \{u \in V_0 : \mathrm{EV}[(v_n, u)] \subseteq X_n\}$ and observe that with these choices the recursive selection is completed.

We set $\mathbf{v} = (v_n)$. Since $v_n \in V_0$ and $v_n < v_{n+1}$ for every $n \in \mathbb{N}$, we see that $\mathbf{v}$ is an extracted subsequence of (w_n). We will show that $\mathrm{EV}[\mathbf{v}] \subseteq X$. To this end, for every $m, i \in \mathbb{N}$ let $\mathrm{EV}(\mathbf{v}, m, i)$ be the set of all variable words of the form $v_{i_0}(a_0)^\frown \dots ^\frown v_{i_m}(a_m)$ where $i_0 < \dots < i_m$ is a finite strictly increasing sequence in $\mathbb{N}$ with $i_0 = i$ and $(a_0, \dots, a_m)$ is a variable word over A. By induction on m, we will show that $\mathrm{EV}(\mathbf{v}, m, i) \subseteq X_i$ for every $i \in \mathbb{N}$. Notice that $\mathrm{EV}(\mathbf{v}, 0, i) = \{v_i\}$ for every $i \in \mathbb{N}$, and so the case "$m = 0$" follows immediately by the properties of the above recursive selection. Let $m \in \mathbb{N}$ and assume that $\mathrm{EV}(\mathbf{v}, m, i) \subseteq X_i$ for every $i \in \mathbb{N}$. Fix a strictly increasing sequence $i_0 < \dots < i_{m+1}$ in $\mathbb{N}$ and a variable word $(a_0, \dots, a_{m+1})$ over A. We have to prove that $v_{i_0}(a_0)^\frown \dots ^\frown v_{i_{m+1}}(a_{m+1}) \in X_{i_0}$. We consider the following cases.

Case 1: *there exists $j \in [m+1]$ such that $a_j = x$.* In this case, setting $v = v_{i_1}(a_1)^\frown \dots ^\frown v_{i_{m+1}}(a_{m+1})$, we have $v \in \mathrm{EV}(\mathbf{v}, m, i_1) \subseteq X_{i_1}$. Hence,

$$v \in X_{i_1} \subseteq X_{i_0+1} = \{u \in V_0 : \mathrm{EV}[(v_{i_0}, u)] \subseteq X_{i_0}\}.$$

Also observe that $v_{i_0}(a)^\frown v \in \mathrm{EV}[(v_{i_0}, v)]$ for every $a \in A \cup \{x\}$. Therefore, we conclude that $v_{i_0}(a_0)^\frown \dots ^\frown v_{i_{m+1}}(a_{m+1}) \in X_{i_0}$.

Case 2: *we have $a_j \in A$ for every $j \in [m+1]$.* Recall that $(a_0, \dots, a_{m+1})$ is a variable word over A. Therefore, by our assumptions, we obtain that $a_0 = x$. We set $v' = v_{i_1}(x)^\frown \dots ^\frown v_{i_{m+1}}(a_{m+1})$. Arguing precisely as in the previous case we see that $v' \in \{u \in V_0 : \mathrm{EV}[(v_{i_0}, u)] \subseteq X_{i_0}\}$. Hence,

$$v_{i_0}(a_0)^\frown v_{i_1}(a_1)^\frown \dots ^\frown v_{i_{m+1}}(a_{m+1}) = v_{i_0}{}^\frown v'(a_1) \in \mathrm{EV}[(v_{i_0}, v')] \subseteq X_{i_0}$$

as desired.

The above cases are exhaustive, and so this completes the proof of the general inductive step. Finally, recall that the sequence (X_n) is decreasing and $X_0 = X$.

It follows, in particular, that $\mathrm{EV}(\mathbf{v},m,i)\subseteq X_i\subseteq X$ for every $m,i\in\mathbb{N}$. Therefore, $\mathrm{EV}[\mathbf{v}]=\bigcup_{(m,i)\in\mathbb{N}^2}\mathrm{EV}(\mathbf{v},m,i)\subseteq X$ and the proof of Claim 4.4 is completed. □

We are now ready to complete the proof of the theorem. Let $\mathcal{V}\in\gamma V$ be as in Claim 4.3 and fix a finite coloring $c\colon \mathrm{EV}[\mathbf{w}]\to[r]$. There exists $p\in[r]$ such that the set $c^{-1}(\{p\})$ belongs to the ultrafilter $\mathcal{V}$. By Claim 4.4, there exists an extracted sequence $\mathbf{v}=(v_n)$ of $\mathbf{w}$ such that $\mathrm{EV}[\mathbf{v}]\subseteq c^{-1}(\{p\})$. The proof of Theorem 4.1 is thus completed. □

4.2. Applications

In this section we present several applications of Carlson's theorem. We start with the following theorem due to Carlson and Simpson [**CS**].

THEOREM 4.5. *Let A be a finite alphabet with $|A|\geqslant 2$. Then for every finite coloring of the set of all words over A there exist a word w over A and a sequence (u_n) of left variable words over A such that the set*

$$\{w\}\cup\big\{w^\frown u_0(a_0)^\frown\dots^\frown u_n(a_n): n\in\mathbb{N} \text{ and } a_0,\dots,a_n\in A\big\}$$

is monochromatic.

The Carlson–Simpson theorem is not only an infinite-dimensional extension of the Hales–Jewett theorem but it also refines the Hales–Jewett theorem by providing information on the structure of the wildcard set of the monochromatic variable word. This additional information (namely, that the sequence (u_n) consists of left variable words) has further combinatorial consequences. For instance, Theorem 4.5 is easily seen to imply the strong subtree version of the Halpern–Läuchli theorem for vector homogeneous trees.

The idea to extend the scope of applications of the Hales–Jewett theorem by providing information on the structure of the wildcard set of the monochromatic variable word, is quite fruitful and some recent trends in Ramsey theory are pointing in this direction. A well-known example is the *polynomial Hales–Jewett theorem* due to Bergelson and Leibman [**BL**].

PROOF OF THEOREM 4.5. Let $c\colon A^{<\mathbb{N}}\to[r]$ be a finite coloring. Notice that every variable word v over A is written, uniquely, as $v^{*\frown}v^{**}$ where v^* is a word over A and v^{**} is a left variable word over A. (If v is a left variable word, then v^* is the empty word and $v^{**}=v$.) Using this decomposition we see that the coloring c corresponds to an r-coloring C of the set of all variable words over A which is defined by the rule $C(v)=c(v^*)$. By Theorem 4.1, there exist $p\in[r]$ and a sequence $\mathbf{v}=(v_n)$ of variable words over A such that $\mathrm{EV}[\mathbf{v}]\subseteq C^{-1}(\{p\})$.

Fix $\alpha\in A$ and set $w=v_0(\alpha)^\frown v_1^*$. Also let $u_n=v_{n+1}^{**}{}^\frown v_{n+2}^*$ for every $n\in\mathbb{N}$. We will show that the word w and the sequence (u_n) are as desired. Notice, first, that u_n is left variable word for every $n\in\mathbb{N}$. Next observe that $w=(v_0(\alpha)^\frown v_1)^*$ and $v_0(\alpha)^\frown v_1\in\mathrm{EV}[\mathbf{v}]$. This implies, of course, that $c(w)=p$. Finally, let $n\in\mathbb{N}$

and $a_0, \dots, a_n \in A$ be arbitrary. Observe that

$$\begin{aligned} w^\frown u_0(a_0)^\frown \dots ^\frown u_n(a_n) &= (v_0(\alpha)^\frown v_1^*)^\frown (v_1^{**}(a_0)^\frown v_2^*)^\frown \dots ^\frown (v_{n+1}^{**}(a_n)^\frown v_{n+2}^*) \\ &= v_0(\alpha)^\frown v_1(a_0)^\frown \dots ^\frown v_{n+1}(a_n)^\frown v_{n+2}^* \\ &= (v_0(\alpha)^\frown v_1(a_0)^\frown \dots ^\frown v_{n+1}(a_n)^\frown v_{n+2})^*. \end{aligned}$$

Since $v_0(\alpha)^\frown v_1(a_0)^\frown \dots ^\frown v_{n+1}(a_n)^\frown v_{n+2} \in \mathrm{EV}[\mathbf{v}]$ we conclude that

$$c\big(w^\frown u_0(a_0)^\frown \dots ^\frown u_n(a_n)\big) = C\big(v_0(\alpha)^\frown v_1(a_0)^\frown \dots ^\frown v_{n+1}(a_n)^\frown v_{n+2}\big) = p.$$

The proof of Theorem 4.5 is completed. □

The following result is a version of Theorem 4.1 for extracted words. It was obtained independently by Carlson [**C**], and by Furstenberg and Katznelson [**FK3**].

THEOREM 4.6. *Let A be a finite alphabet with $|A| \geqslant 2$ and $\mathbf{w} = (w_n)$ a sequence of variable words over A. Then for every finite coloring of the set $\mathrm{E}[\mathbf{w}]$ of all extracted words of $\mathbf{w}$ there exists an extracted subsequence $\mathbf{v}$ of $\mathbf{w}$ such that the set $\mathrm{E}[\mathbf{v}]$ is monochromatic.*

PROOF. We fix a finite coloring $c \colon \mathrm{E}[\mathbf{w}] \to [r]$. We also fix an element $\alpha \in A$ and we define a coloring $C \colon \mathrm{EV}[\mathbf{w}] \to [r]$ by the rule $C(v) = c\big(v(\alpha)\big)$. By Theorem 4.1, there exist $p \in [r]$ and an extracted subsequence $\mathbf{u} = (u_n)$ of $\mathbf{w}$ such that $\mathrm{EV}[\mathbf{u}] \subseteq C^{-1}(\{p\})$. We set $v_n = u_{2n}{}^\frown u_{2n+1}(\alpha)$ for every $n \in \mathbb{N}$ and we observe that $\mathbf{v} = (v_n)$ is an extracted subsequence of $\mathbf{w}$. We claim that $c(w) = p$ for every $w \in \mathrm{E}[\mathbf{v}]$. Indeed, let w be an extracted word of $\mathbf{v}$. There exist $n \in \mathbb{N}$, a finite strictly increasing sequence $i_0 < \dots < i_n$ in $\mathbb{N}$ and $a_0, \dots, a_n \in A$ such that

$$w = v_{i_0}(a_0)^\frown \dots ^\frown v_{i_n}(a_n) = u_{2i_0}(a_0)^\frown u_{2i_0+1}(\alpha)^\frown \dots ^\frown u_{2i_n}(a_n)^\frown u_{2i_n+1}(\alpha).$$

Set $v = u_{2i_0}(a_0)^\frown u_{2i_0+1}(x)^\frown \dots ^\frown u_{2i_n}(a_n)^\frown u_{2i_n+1}(x)$ and observe that $v \in \mathrm{EV}[\mathbf{u}]$ and $v(\alpha) = w$. Hence, $c(w) = c\big(v(\alpha)\big) = C(v) = p$ and the proof of Theorem 4.6 is completed. □

The next application is a higher-dimensional extension of Carlson's theorem.

THEOREM 4.7. *Let A be a finite alphabet with $|A| \geqslant 2$. Also let m be a positive integer. Then for every sequence $\mathbf{w} = (w_n)$ of variable words over A and every finite coloring of the set $\mathrm{EV}_m[\mathbf{w}]$ there exists an extracted subsequence $\mathbf{v}$ of $\mathbf{w}$ such that the set $\mathrm{EV}_m[\mathbf{v}]$ is monochromatic.*

For the proof of Theorem 4.7 we need to introduce some pieces of notation. Specifically, let A be a finite alphabet with $|A| \geqslant 2$ and $\mathbf{w} = (w_n)$ a sequence of variable words over A. For every pair ℓ, m of positive integers with $\ell \leqslant m$ and every $(u_i)_{i=0}^{\ell-1} \in \mathrm{EV}_\ell[\mathbf{w}]$ we set

$$\mathrm{EV}_m[(u_i)_{i=0}^{\ell-1}, \mathbf{w}] = \big\{\mathbf{v} \in \mathrm{EV}_m[\mathbf{w}] : \mathbf{v} \upharpoonright \ell = (u_i)_{i=0}^{\ell-1}\big\}. \tag{4.11}$$

Moreover, for every positive integer n let

$$\mathrm{EV}_\infty[n, \mathbf{w}] = \{\mathbf{v} \in \mathrm{EV}_\infty[\mathbf{w}] : \mathbf{v} \upharpoonright n = \mathbf{w} \upharpoonright n\}. \tag{4.12}$$

We are ready to give the proof of Theorem 4.7.

Proof of Theorem 4.7. By induction on m. The case "$m = 1$" is the content of Theorem 4.1, and so let $m \geqslant 1$ and assume that the result has been proved up to m. Fix a sequence $\mathbf{w} = (w_n)$ of variable words over A and let $c\colon \mathrm{EV}_{m+1}[\mathbf{w}] \to [r]$ be a finite coloring. Recursively, we will select a sequence $(\mathbf{w}_n)$ of extracted subsequences of $\mathbf{w}$ with $\mathbf{w}_0 = \mathbf{w}$ and satisfying the following conditions for every $n \in \mathbb{N}$.

(C1) We have $\mathbf{w}_{n+1} \in \mathrm{EV}_\infty[m+n, \mathbf{w}_n]$.

(C2) For every $(v_i)_{i=0}^{m-1} \in \mathrm{EV}_m[\mathbf{w}_{n+1} \upharpoonright (m+n)]$ the set $\mathrm{EV}_{m+1}[(v_i)_{i=0}^{m-1}, \mathbf{w}_{n+1}]$ is monochromatic.

Assuming that the above selection has been carried out, the proof is completed as follows. By condition (C1) and the fact that $\mathbf{w}_0 = \mathbf{w}$, there exists a unique $\mathbf{u} \in \mathrm{EV}_\infty[\mathbf{w}]$ such that $\mathbf{u} \in \mathrm{EV}_\infty[m+n, \mathbf{w}_n]$ for every $n \in \mathbb{N}$. By condition (C2), we see that $c\big((s_i)_{i=0}^m\big) = c\big((t_i)_{i=0}^m\big)$ for every pair $(s_i)_{i=0}^m$ and $(t_i)_{i=0}^m$ in $\mathrm{EV}_{m+1}[\mathbf{u}]$ with $s_i = t_i$ for every $i \in \{0, \dots, m-1\}$. It follows, in particular, that the coloring of $\mathrm{EV}_{m+1}[\mathbf{u}]$ is reduced to a coloring of $\mathrm{EV}_m[\mathbf{u}]$. Hence, by our inductive assumptions, there exists $\mathbf{v} \in \mathrm{EV}_\infty[\mathbf{u}]$ such that $\mathrm{EV}_{m+1}[\mathbf{v}]$ is monochromatic with respect to c.

It remains to carry out the recursive selection. First we set $\mathbf{w}_0 = \mathbf{w}$. Let $n \in \mathbb{N}$ and assume that the sequences $\mathbf{w}_0, \dots, \mathbf{w}_n$ have been selected so that (C1) and (C2) are satisfied. Write $\mathbf{w}_n = (w_i^{(n)})$ and notice that the set $F \coloneqq \mathrm{EV}_m[(w_i^{(n)})_{i=0}^{m+n-1}]$ is finite. Define $C\colon \mathrm{EV}[(w_i^{(n)})_{i=m+n}^\infty] \to [r]^F$ by the rule

$$C(v) = \big\langle c\big((v_0, \dots, v_{m-1}, v)\big) : (v_0, \dots, v_{m-1}) \in F \big\rangle.$$

By Theorem 4.1, there exists an extracted subsequence $\mathbf{s} = (s_i)$ of $(w_i^{(n)})_{i=m+n}^\infty$ such that the set $\mathrm{EV}[\mathbf{s}]$ is monochromatic with respect to C. We set $w_i^{(n+1)} = w_i^{(n)}$ if $i \in \{0, \dots, m+n-1\}$ and $w_i^{(n+1)} = s_{i-m-n}$ if $i \geqslant m+n$. It is then clear that the sequence $\mathbf{w}_{n+1} = (w_i^{(n+1)})$ is as desired. This completes the recursive selection, and as we have already indicated, the proof of Theorem 4.7 is also completed. □

Recall that a nonempty finite sequence $\mathcal{F} = (F_0, \dots, F_{n-1})$ of nonempty finite subsets of $\mathbb{N}$ is said to be block if $\max(F_i) < \min(F_j)$ for every $i, j \in \{0, \dots, n-1\}$ with $i < j$. Respectively, we say that an infinite sequence $\mathcal{X} = (X_n)$ of nonempty finite subsets of $\mathbb{N}$ is *block* if $\max(X_i) < \min(X_j)$ for every $i, j \in \mathbb{N}$ with $i < j$. For every infinite block sequence $\mathcal{X} = (X_n)$ and every positive integer m we denote by $\mathrm{Block}_m(\mathcal{X})$ the set of all block sequences $\mathcal{F} = (F_0, \dots, F_{m-1})$ of length m such that for every $i \in \{0, \dots, m-1\}$ there exists $G \subseteq \mathbb{N}$ with $F_i = \bigcup_{n \in G} X_n$.

The following result is due, independently, to Milliken [**M1**] and Taylor [**Tay1**]. The case "$m = 1$" is Hindman's theorem [**H**].

Theorem 4.8. *For every positive integer m and every finite coloring of the set of all block sequences of length m there exists an infinite block sequence $\mathcal{X} = (X_n)$ such that the set* $\mathrm{Block}_m(\mathcal{X})$ *is monochromatic.*

Proof. Fix an integer $m \geqslant 1$ and a finite coloring c of the set of all block sequences of length m. Also fix a finite alphabet A with $|A| \geqslant 2$ and a sequence $\mathbf{w}$ of variable words over A. For every $\ell \in [m]$ and every $\mathbf{v} = (v_i)_{i=0}^{\ell-1} \in \mathrm{EV}_\ell[\mathbf{w}]$ let $F(\mathbf{v})$ be the sequence of wildcard sets of the ℓ-dimensional combinatorial space

$[(v_i)_{i=0}^{\ell-1}]$ and notice the $F(\mathbf{v})$ is a block sequence of length ℓ. Thus, we can define a finite coloring C of $\mathrm{EV}_m[\mathbf{w}]$ by the rule $C(\mathbf{v}) = c\big(F(\mathbf{v})\big)$. By Theorem 4.7, there exists an extracted subsequence $\mathbf{u} = (u_n)$ of $\mathbf{w}$ such that the set $\mathrm{EV}_m[\mathbf{u}]$ is monochromatic with respect to the coloring C. Fix $\alpha \in A$. We set $X_0 = F(u_0)$ and $X_n = F\big(u_0(\alpha)^\frown \dots ^\frown u_{n-1}(\alpha)^\frown u_n\big)$ for every $n \geqslant 1$. It is easy to see that the infinite block sequence (X_n) is as desired. The proof of Theorem 4.8 is completed. $\square$

We close this section with the following analogue of Theorem 4.7 for reduced variable words (see Subsection 1.4.1 for the relevant definitions).

THEOREM 4.9. *Let A be a finite alphabet with $|A| \geqslant 2$. Also let m be a positive integer. Then for every sequence $\mathbf{w} = (w_n)$ of variable words over A and every finite coloring of the set $\mathrm{V}_m[\mathbf{w}]$ of all reduced subsequences of $\mathbf{w}$ of length m there exists a reduced subsequence $\mathbf{v}$ of $\mathbf{w}$ such that the set $\mathrm{V}_m[\mathbf{v}]$ is monochromatic.*

Although Theorem 4.9 is quite similar to Theorem 4.7, the reader should have in mind that these two statements refer to different types of structures. Also note that Theorem 4.9 implies Theorem 2.15 via a standard argument but, of course, this reduction is ineffective. We proceed to the proof of Theorem 4.9.

PROOF OF THEOREM 4.9. First we select a reduced subsequence $\mathbf{u} = (u_n)$ of $\mathbf{w}$ such that for every $t \in \mathrm{EV}[\mathbf{u}]$ there exist a unique $n \in \mathbb{N}$, a unique finite strictly increasing sequence $i_0 < \dots < i_n$ in $\mathbb{N}$ and a unique a variable word $(a_0, \dots, a_n)$ over A such that $t = u_{i_0}(a_0)^\frown \dots ^\frown u_{i_n}(a_n)$. (This property is guaranteed, for example, if the reduced subsequence (u_n) of (w_n) satisfies $|u_{n+1}| > \sum_{i=0}^{n} |u_n|$ for every $n \in \mathbb{N}$.) For every $t \in \mathrm{EV}[\mathbf{u}]$ we shall denote by $\mathrm{supp}_{\mathbf{u}}(t)$ the unique set $\{i_0, \dots, i_n\}$ of indices which correspond to t.

Fix $\alpha \in A$. For every $t = u_{i_0}(a_0)^\frown \dots ^\frown u_{i_n}(a_n) \in \mathrm{EV}[\mathbf{u}]$ and every $m_1, m_2 \in \mathbb{N}$ with $m_1 \leqslant i_0 \leqslant i_n \leqslant m_2$ let $Q_{\mathbf{u}}|_{m_1}^{m_2}(t)$ be the reduced variable word of $(u_i)_{i=m_1}^{m_2}$ which is defined by the rule

$$Q_{\mathbf{u}}|_{m_1}^{m_2}(t) = u_{m_1}(b_{m_1})^\frown \dots ^\frown u_{m_2}(b_{m_2}) \tag{4.13}$$

where $b_{i_0} = a_0, \dots, b_{i_n} = a_n$ and $b_i = \alpha$ if $i \in \{m_1, \dots, m_2\} \setminus \mathrm{supp}_{\mathbf{u}}(t)$. We view the word $Q_{\mathbf{u}}|_{m_1}^{m_2}(t)$ as a "reduced extension" of t. Using these "reduced extensions" we define a map $\mathbf{Q}_{\mathbf{u}} \colon \mathrm{EV}_m[\mathbf{u}] \to \mathrm{V}_m[\mathbf{u}]$ as follows. Let $\mathbf{t} = (t_0, \dots, t_{m-1}) \in \mathrm{EV}_m[\mathbf{u}]$. Set $p_0 = 0$ and $p_{i+1} = \max\big(\mathrm{supp}_{\mathbf{u}}(t_i)\big) + 1$ for every $i \in \{0, \dots, m-1\}$, and let

$$\mathbf{Q}_{\mathbf{u}}\big(\mathbf{t}\big) = \big(Q_{\mathbf{u}}|_{p_0}^{p_1-1}(t_0), \dots, Q_{\mathbf{u}}|_{p_{m-1}}^{p_m-1}(t_{m-1})\big). \tag{4.14}$$

Now fix a finite coloring $c \colon \mathrm{V}_m[\mathbf{u}] \to [r]$. We define $C \colon \mathrm{EV}_m[\mathbf{u}] \to [r]$ by the rule $C(\mathbf{t}) = c\big(\mathbf{Q}_{\mathbf{u}}(\mathbf{t})\big)$ for every $\mathbf{t} \in \mathrm{EV}_m[\mathbf{u}]$. By Theorem 4.7, there exist $p \in [r]$ and an extracted subsequence $\mathbf{y} = (y_n)$ of $\mathbf{u}$ such that $\mathrm{EV}_m[\mathbf{y}] \subseteq C^{-1}(\{p\})$. Let $m_0 = 0$ and $m_{n+1} = \max\big(\mathrm{supp}_{\mathbf{u}}(y_n)\big) + 1$ for every $n \in \mathbb{N}$. We set

$$v_n = Q_{\mathbf{u}}|_{m_n}^{m_{n+1}-1}(y_n) \tag{4.15}$$

for every $n \in \mathbb{N}$ and we observe that $\mathbf{v} = (v_n)$ is a reduced subsequence of $\mathbf{u}$. Since $\mathbf{u}$ is a reduced subsequence of $\mathbf{w}$, we see that $\mathbf{v} \in \mathrm{V}_\infty[\mathbf{w}]$. We will show that $c(\mathbf{s}) = p$ for every $\mathbf{s} \in \mathrm{V}_m[\mathbf{v}]$. Indeed, let $\mathbf{s} = (s_0, \dots, s_{m-1}) \in \mathrm{V}_m[\mathbf{v}]$ be arbitrary. There exist a finite strictly increasing sequence $0 = \ell_0 < \cdots < \ell_m$ in $\mathbb{N}$ and $a_0, \dots, a_{\ell_m - 1} \in A \cup \{x\}$ such that for every $i \in \{0, \dots, m-1\}$ we have

$$s_i = v_{\ell_i}(a_{\ell_i})^\frown \dots ^\frown v_{\ell_{i+1}-1}(a_{\ell_{i+1}-1}) \in \mathrm{V}[(v_j)_{j=\ell_i}^{\ell_{i+1}-1}].$$

For every $i \in \{0, \dots, m-1\}$ set

$$t_i = y_{\ell_i}(a_{\ell_i})^\frown \dots ^\frown y_{\ell_{i+1}-1}(a_{\ell_{i+1}-1})$$

and notice that $\mathbf{t} = (t_0, \dots, t_{m-1}) \in \mathrm{V}_m[\mathbf{y}] \subseteq \mathrm{EV}_m[\mathbf{y}]$ and $\mathbf{Q_u}(\mathbf{t}) = \mathbf{s}$. It follows that $c(\mathbf{s}) = c\big(\mathbf{Q_u}(\mathbf{t})\big) = C(\mathbf{t}) = p$ and the proof of Theorem 4.9 is completed. $\square$

4.3. Finite versions

We now turn our attention to the study of the finite analogues of the main results presented in this chapter so far. Our primary objective is to obtain quantitative information for the numerical invariants associated with these finite analogues. The main tools are the Hales–Jewett theorem and its consequences.

4.3.1. The finite version of the Carlson–Simpson theorem. The following result is the finite version of Theorem 4.5.

PROPOSITION 4.10. *For every triple k, d, r of positive integers with $k \geqslant 2$ there exists a positive integer N with the following property. If A is an alphabet with $|A| = k$, then for every Carlson–Simpson space T of $A^{<\mathbb{N}}$ of dimension at least N and every r-coloring of T there exists a d-dimensional Carlson–Simpson subspace of T which is monochromatic. The least positive integer with this property will be denoted by $\mathrm{cs}(k, d, r)$.*

Moreover, the numbers $\mathrm{cs}(k, d, r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^5$.

PROOF. It is similar to the proof of Theorem 4.5. Precisely, we will show that

$$\mathrm{cs}(k, d, r) \leqslant \mathrm{GR}(k, d+1, 1, r). \tag{4.16}$$

By Proposition 2.25, this will complete the proof.

To this end, fix a triple k, d, r of positive integers with $k \geqslant 2$, and let A be an alphabet with $|A| = k$. Also let

$$n \geqslant \mathrm{GR}(k, d+1, 1, r) \tag{4.17}$$

and T an n-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ generated by the system $\langle t, (v_i)_{i=0}^{n-1} \rangle$. Finally, let c be an r-coloring of T. Set $\mathbf{v} = (t^\frown v_0, v_1, \dots, v_{n-1})$ and recall that every $v \in \mathrm{V}_1[\mathbf{v}]$ is written as $v^{*\frown} v^{**}$ where v^* is a word over A and v^{**} is a left variable word over A; moreover, observe that $v^* \in T$. We define an r-coloring C of $\mathrm{V}_1[\mathbf{v}]$ by $C(v) = c(v^*)$ and we notice that, by (4.17), there exists $(u_i)_{i=0}^d \in \mathrm{V}_{d+1}[\mathbf{v}]$ such that the set $\mathrm{V}_1[(u_i)_{i=0}^d]$ is monochromatic with respect to C. Set $s = u_0^*$ and $w_i = u_i^{**\frown} u_{i+1}^*$ for every $i \in \{0, \dots, d-1\}$, and let S be the Carlson–Simpson space of $A^{<\mathbb{N}}$ generated by the system $\langle s, (w_i)_{i=0}^{d-1} \rangle$. Clearly, S is a d-dimensional Carlson–Simpson subspace of T which is monochromatic with respect to c. The proof of Proposition 4.10 is completed. $\square$

We have already pointed out that the Carlson–Simpson theorem implies both the Hales–Jewett theorem and the strong subtree version of the Halpern–Läuchli theorem for vector homogeneous trees. The following corollary provides finer quantitative information.

COROLLARY 4.11. *The following hold.*

(a) *For every triple k, d, r of positive integers with $k \geqslant 2$ we have*

$$\mathrm{MHJ}(k, d, r) \leqslant \mathrm{cs}(k, d, r). \tag{4.18}$$

(b) *For every integer $d \geqslant 1$, every $b_1, \dots, b_d \in \mathbb{N}$ with $b_i \geqslant 2$ for all $i \in [d]$ and every pair ℓ, r of positive integers with $\ell \geqslant 2$ we have*

$$\mathrm{HL}(b_1, \dots, b_d \,|\, \ell, r) \leqslant \mathrm{cs}\Big(\prod_{i=1}^{d} b_i, \ell - 1, r\Big) + 1. \tag{4.19}$$

PROOF. First we argue for part (a). Let k, d, r be positive integers with $k \geqslant 2$ and set $N = \mathrm{cs}(k, d, r)$. Also let A be an alphabet with $|A| = k$ and $c \colon A^N \to [r]$ a coloring. Fix an element $\alpha \in A$ and for every $i \in \mathbb{N}$ let α^i be as in (2.1). We define a coloring $c' \colon A^{<N+1} \to [r]$ by the rule $c'(w) = c(w^\frown \alpha^{N-|w|})$ for every $w \in A^{<N+1}$. By the choice of N, there exists a d-dimensional Carlson–Simpson system $\langle s, (w_i)_{i=0}^{d-1} \rangle$ over A such that the Carlson–Simpson space generated by this system is monochromatic with respect to c'. Setting $j = N - |s| - \sum_{i=0}^{d-1} |w_i|$ and $W = \{ s^\frown w_0(a_0)^\frown \dots {}^\frown w_{d-1}(a_{d-1})^\frown \alpha^j : a_0, \dots, a_{d-1} \in A \}$, we see that W is a monochromatic, with respect to c, d-dimensional combinatorial subspace of A^N.

The proof of part (b) is similar to the proof of Proposition 3.27. Fix the "dimension" d and the parameters $b_1, \dots, b_d, \ell, r$ and set $N = \mathrm{cs}\big(\prod_{i=1}^d b_i, \ell - 1, r\big)$. Let $\mathbf{T} = (T_1, \dots, T_d)$ be a vector homogeneous tree with $b_{T_i} = b_i$ for all $i \in [d]$ and $h(\mathbf{T}) \geqslant N + 1$. Clearly, we may assume that $T_i = [b_i]^{<N+1}$ for every $i \in [d]$. Also let $c \colon \otimes \mathbf{T} \to [r]$ be a coloring. We set $\mathbb{A} = [b_1] \times \dots \times [b_d]$ and for every $i \in [d]$ we define $\bar{\pi}_i \colon \mathbb{A}^{<N+1} \to [b_i]^{<N+1}$ precisely as in the proof of Proposition 3.27. Next, define $\mathrm{I} \colon \mathbb{A}^{<N+1} \to \otimes \mathbf{T}$ by the rule $\mathrm{I}(w) = \big(\bar{\pi}_1(w), \dots, \bar{\pi}_d(w)\big)$ and recall that the map I is a bijection. Hence, the coloring c induces a coloring $c' \colon \mathbb{A}^{<N+1} \to [r]$ defined by $c' = c \circ \mathrm{I}$. Let S be a Carlson–Simpson subspace of $\mathbb{A}^{<N+1}$ of dimension $\ell - 1$ which is monochromatic with respect to c'. For every $i \in [d]$ set $S_i = \bar{\pi}_i(S)$ and observe that S_i is a Carlson–Simpson subspace of $[b_i]^{<N+1}$ having the same level set as S. It follows, in particular, that $\mathbf{S} = (S_1, \dots, S_d)$ is a vector strong subtree of $\mathbf{T}$. Since $\mathrm{I}(S) = \otimes \mathbf{S}$, we conclude that the level product $\otimes \mathbf{S}$ of $\mathbf{S}$ is monochromatic with respect to c and the proof of Corollary 4.11 is completed. □

4.3.2. Extracted words of finite sequences of variable words. This subsection is devoted to the proof of the following finite version of Theorem 4.6.

THEOREM 4.12. *For every triple k, d, r of positive integers with $k \geqslant 2$ there exists a positive integer N with the following property. If $n \geqslant N$ and A is an alphabet with $|A| = k$, then for every finite sequence $(w_i)_{i=0}^{n-1}$ of variable words over A and every r-coloring of the set $\mathrm{E}[(w_i)_{i=0}^{n-1}]$ of all extracted words of $(w_i)_{i=0}^{n-1}$ there* *exists an extracted subsequence* $(v_i)_{i=0}^{d-1}$ *of* $(w_i)_{i=0}^{n-1}$ *such that the set* $\mathrm{E}[(v_i)_{i=0}^{d-1}]$ *is monochromatic. The least integer with this property will be denoted by* $\mathrm{c}(k,d,r)$*.*

Moreover, the numbers $\mathrm{c}(k,d,r)$ *are upper bounded by a primitive recursive function belonging to the class* $\mathcal{E}^5$*.*

It is convenient to introduce the following notation. Let F be a nonempty finite subset of $\mathbb{N}$ and A a finite alphabet with $|A| \geqslant 2$. For every $n \in F$ let w_n be either a word or a variable word over A. By $\prod_{n\in F} w_n$ we shall denote the concatenation of $\{w_n : n \in F\}$ in increasing order of indices. That is, if $i_0 < \cdots < i_m$ is the increasing enumeration of F, then

$$\prod_{n\in F} w_n = w_{i_0}{}^\frown \dots {}^\frown w_{i_m}. \tag{4.20}$$

We are ready to give the proof of Theorem 4.12.

Proof of Theorem 4.12. It is similar to the second proof of the multidimensional Hales–Jewett theorem in Section 2.2.

Fix a triple k,d,r of positive integers with $k \geqslant 2$ and set $m = \mathrm{H}(d,r)$ where $\mathrm{H}(d,r)$ is as in Proposition 2.19. We will show that

$$\mathrm{c}(k,d,r) \leqslant m \cdot \mathrm{HJ}\big(k^m, r^{2^m-1}\big). \tag{4.21}$$

Indeed, let $n \geqslant m \cdot \mathrm{HJ}\big(k^m, r^{2^m-1}\big)$ and fix an alphabet A with $|A| = k$. Also let $(w_i)_{i=0}^{n-1}$ be a finite sequence of variable words over A and let $c\colon \mathrm{E}[(w_i)_{i=0}^{n-1}] \to [r]$ be a coloring. For every $i \in \{0,\dots,m-1\}$ fix $x_i \notin A$ and set $\mathbf{x} = (x_0,\dots,x_{m-1})$. Observe that for every positive integer p the set of all variable words over A^m of length p is naturally identified with the set $(A^m \cup \{\mathbf{x}\})^p \setminus (A^m)^p$. Next, set

$$\ell = \mathrm{HJ}\big(k^m, r^{2^m-1}\big) \tag{4.22}$$

and for every $i \in \{0,\dots,m-1\}$ define $\pi_i\colon A^m \cup \{\mathbf{x}\} \to A \cup \{x\}$ by

$$\pi_i\big((a_0,\dots,a_{m-1})\big) = \begin{cases} a_i & \text{if } a_i \in A, \\ x & \text{if } a_i = x_i \end{cases} \tag{4.23}$$

and $\mathrm{T}_i\colon (A^m \cup \{\mathbf{x}\})^\ell \to [(w_{i\cdot\ell+j})_{j=0}^{\ell-1}] \cup \mathrm{V}[(w_{i\cdot\ell+j})_{j=0}^{\ell-1}]$ by

$$\mathrm{T}_i\big((b_0,\dots,b_{\ell-1})\big) = \prod_{j=0}^{\ell-1} w_{i\cdot\ell+j}\big(\pi_i(b_j)\big). \tag{4.24}$$

Notice that for every $i \in \{0,\dots,m-1\}$ we have

$$\mathrm{T}_i\big((A^m)^\ell\big) = [(w_{i\cdot\ell+j})_{j=0}^{\ell-1}] \ \text{ and } \ \mathrm{T}_i\big((A^m \cup \{\mathbf{x}\})^\ell \setminus (A^m)^\ell\big) \subseteq \mathrm{V}[(w_{i\cdot\ell+j})_{j=0}^{\ell-1}].$$

Also observe that for every $v \in (A^m \cup \{\mathbf{x}\})^\ell \setminus (A^m)^\ell$ (that is, v is a variable word over A^m of length ℓ) and every $(a_0,\dots,a_{m-1}) \in A^m$, setting $u_i = \mathrm{T}_i(v)$, we have

$$\mathrm{T}_i\Big(v\big((a_0,\dots,a_{m-1})\big)\Big) = u_i(a_i). \tag{4.25}$$

We have the following claim.

CLAIM 4.13. *There exists a finite sequence* $(u_i)_{i=0}^{m-1}$ *of variable words over* A *with* $u_i \in \mathrm{V}[(w_{i\cdot\ell+j})_{j=0}^{\ell-1}]$ *for every* $i \in \{0,\dots,m-1\}$ *and satisfying the following property. For every nonempty subset* F *of* $\{0,\dots,m-1\}$ *there exists* $p_F \in [r]$ *such that* $c\big(\prod_{i\in F} u_i(a_i)\big) = p_F$ *for every* $(a_i)_{i\in F} \in A^F$.

PROOF OF CLAIM 4.13. Let $\mathcal{F} = \big\{F \subseteq \{0,\dots,m-1\} : F \neq \emptyset\big\}$ and define $C\colon (A^m)^\ell \to [r]^{\mathcal{F}}$ by the rule

$$C\big((b_0,\dots,b_{\ell-1})\big) = \Big\langle c\Big(\prod_{i\in F} \mathrm{T}_i\big((b_0,\dots,b_{\ell-1})\big)\Big) : F \in \mathcal{F}\Big\rangle. \tag{4.26}$$

By the choice of ℓ in (4.22), there exists $v \in (A^m \cup \{\mathbf{x}\})^\ell \setminus (A^m)^\ell$ such that the combinatorial line $L \coloneqq \big\{v\big((a_0,\dots,a_{m-1})\big) : (a_0,\dots,a_{m-1}) \in A^m\big\}$ of $(A^m)^\ell$ is monochromatic with respect to C. It follows, in particular, that for every $F \in \mathcal{F}$ there exists $p_F \in [r]$ such that for every $(a_0,\dots,a_{m-1}) \in A^m$ we have

$$c\Big(\prod_{i\in F} \mathrm{T}_i\Big(v\big((a_0,\dots,a_{m-1})\big)\Big)\Big) = p_F. \tag{4.27}$$

For every $i \in \{0,\dots,m-1\}$ we set $u_i = \mathrm{T}_i(v)$. Notice that $u_i \in \mathrm{V}[(w_{i\cdot\ell+j})_{j=0}^{\ell-1}]$. Moreover, by (4.25) and (4.27), we see that for every $F \in \mathcal{F}$ and every $(a_i)_{i\in F} \in A^F$ we have $c\big(\prod_{i\in F} u_i(a_i)\big) = p_F$. The proof of Claim 4.13 is completed. □

The following claim is the second step of the proof. It is a consequence of Proposition 2.19.

CLAIM 4.14. *Let* $(u_i)_{i=0}^{m-1}$ *be as in Claim* 4.13. *Then there exists an extracted subsequence* $(v_i)_{i=0}^{d-1}$ *of* $(u_i)_{i=0}^{m-1}$ *such that the set* $\mathrm{E}[(v_i)_{i=0}^{d-1}]$ *is monochromatic.*

PROOF OF CLAIM 4.14. For every nonempty $F \subseteq \{0,\dots,m-1\}$ let $p_F \in [r]$ be as in Claim 4.13. The map $F \mapsto p_F$ is, of course, an r-coloring of the set of all nonempty subsets of $\{0,\dots,m-1\}$. Since $m = \mathrm{H}(d,r)$, by Proposition 2.19, there exist $p_0 \in [r]$ and a block sequence $(F_0,\dots,F_{d-1})$ of nonempty subsets of $\{0,\dots,m-1\}$ such that $\bigcup_{i\in G} F_i \mapsto p_0$ for every nonempty $G \subseteq \{0,\dots,d-1\}$. We set $v_i = \prod_{j\in F_i} u_i$ for every $i \in \{0,\dots,d-1\}$. It is clear that the sequence $(v_i)_{i=0}^{d-1}$ is as desired. The proof of Claim 4.14 is completed. □

By Claims 4.13 and 4.14, we conclude that the estimate in (4.21) is satisfied. Finally, the fact that the numbers $\mathrm{c}(k,d,r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^5$ is an immediate consequence of Theorem 2.1, Proposition 2.19 and (4.21). The proof of Theorem 4.12 is completed. □

4.3.3. The finite version of Carlson's theorem. The last result of this section is the following finite version of Theorem 4.1.

THEOREM 4.15. *For every triple* k,d,r *of positive integers with* $k \geqslant 2$ *there exists a positive integer* N *with the following property. If* $n \geqslant N$ *and* A *is an alphabet with* $|A| = k$, *then for every finite sequence* $(w_i)_{i=0}^{n-1}$ *of variable words over* A *and every* r*-coloring of the set* $\mathrm{EV}[(w_i)_{i=0}^{n-1}]$ *of all extracted variable words of* $(w_i)_{i=0}^{n-1}$ *there exists an extracted subsequence* $(v_i)_{i=0}^{d-1}$ *of* $(w_i)_{i=0}^{n-1}$ *such that the set*

$\mathrm{EV}[(v_i)_{i=0}^{d-1}]$ *is monochromatic. The least positive integer with this property will be denoted by* $\mathrm{C}(k,d,r)$.

Moreover, the numbers $\mathrm{C}(k,d,r)$ *are upper bounded by a primitive recursive function belonging to the class* $\mathcal{E}^6$.

The proof of Theorem 4.15 is conceptually close to the proof of Theorem 2.15. However, the argument is slightly more involved since we work with extracted variable words. The following lemma is the analogue of Lemma 2.23. It is the first main step of the proof of Theorem 4.15.

LEMMA 4.16. *Let* k, r, L *be positive integers with* $k \geqslant 2$ *and define a finite sequence* $(N_i)_{i=0}^{L}$ *in* $\mathbb{N}$ *recursively by the rule*

$$\begin{cases} N_0 = 0, \\ N_{i+1} = N_i + \mathrm{HJ}\big(k, r^{(k+1)^{2(N_i+L-i-1)}}\big). \end{cases} \tag{4.28}$$

Let A *be a finite alphabet with* $|A| = k$, $\mathbf{w} = (w_i)_{i=0}^{N_L-1}$ *a finite sequence of variable words over* A *and* $c\colon \mathrm{EV}[\mathbf{w}] \to [r]$ *a coloring. Then there exists a reduced subsequence* $\mathbf{u} = (u_i)_{i=0}^{L-1}$ *of* $\mathbf{w}$ *with the following property. For every nonempty subset* F *of* $\{0,\dots,L-1\}$ *and every* $(a_i)_{i\in F}, (b_i)_{i\in F} \in (A\cup\{x\})^F \setminus A^F$ *with* $\{i\in F : a_i = x\} = \{i \in F : b_i = x\}$ *we have* $c\big(\prod_{i\in F} u_i(a_i)\big) = c\big(\prod_{i\in F} u_i(b_i)\big)$.

PROOF. By backwards induction, we will select a sequence $(u_i)_{i=0}^{L-1}$ of variable words over A such that for every $i \in \{0,\dots,L-1\}$ the following are satisfied.

- (C1) We have that u_i is a reduced variable word of $(w_j)_{j=N_i}^{N_{i+1}-1}$.
- (C2) Let $F = F_1 \cup F_2$ where $F_1 \subseteq \{0,\dots,N_i-1\}$ and $F_2 \subseteq \{i+1,\dots,L-1\}$ with the convention that $F_1 = \emptyset$ if $i=0$ while $F_2=\emptyset$ if $i = L-1$. Assume that the set F is nonempty and let $(a_j)_{j\in F} \in (A\cup\{x\})^F \setminus A^F$. Then for every $a, b \in A$ we have

$$c\Big(\prod_{j\in F_1} w_j(a_j)^\frown u_i(a)^\frown \prod_{j\in F_2} u_j(a_j)\Big) = c\Big(\prod_{j\in F_1} w_j(a_j)^\frown u_i(b)^\frown \prod_{j\in F_2} u_j(a_j)\Big).$$

Assuming that the above selection has been carried out, the proof of the lemma is completed as follows. We set $\mathbf{u} = (u_i)_{i=0}^{L-1}$ and we observe that, by condition (C1), $\mathbf{u}$ is a reduced subsequence of $\mathbf{w}$. Moreover, using condition (C2), we see that $\mathbf{u}$ satisfies the requirements of the lemma.

It remains to carry out the above selection. The first step is identical to the general one, and so let $i \in \{0,\dots,L-2\}$ and assume that the variable words $u_{i+1},\dots,u_{L-1}$ have been selected so that (C1) and (C2) are satisfied. We set

$$n_i = N_{i+1} - N_i \overset{(4.28)}{=} \mathrm{HJ}\big(k, r^{(k+1)^{2(N_i+L-i-1)}}\big). \tag{4.29}$$

Denote by $\mathcal{F}$ the family of all pairs (F_1,F_2) such that: (i) $F_1 \subseteq \{0,\dots,N_i-1\}$ with $F_1 = \emptyset$ if $i=0$, (ii) $F_2 \subseteq \{i+1,\dots,L-1\}$, and (iii) $F_1 \cup F_2 \neq \emptyset$. Moreover, for every nonempty $F \subseteq \mathbb{N}$ let $B(F) = (A\cup\{x\})^F \setminus A^F$. Finally, set $W_i = [(w_j)_{j=N_i}^{N_{i+1}-1}]$ and let $\mathrm{I}_{W_i}\colon A^{n_i} \to W_i$ be the canonical isomorphism associated with the combinatorial space W_i (see Definition 1.2). We define a coloring C of A^{n_i} by the rule

$$C(z) = \Big\langle c\Big(\prod_{j\in F_1} w_j(a_j)^\frown \mathrm{I}_{W_i}(z)^\frown \prod_{j\in F_2} u_j(a_j)\Big) : (F_1,F_2)\in\mathcal{F} \text{ and } (a_j)\in B(F_1\cup F_2)\Big\rangle.$$

Notice that $|B(F_1\cup F_2)| \leqslant (k+1)^{N_i+L-i-1}$ for every $(F_1,F_2)\in\mathcal{F}$. Moreover, we have $|\mathcal{F}| = 2^{N_i}\cdot 2^{L-i-1}-1 \leqslant (k+1)^{N_i+L-i-1}$. Hence, by (4.29), there exists a variable word u over A of length n_i such that the combinatorial line $\{u(a) : a\in A\}$ of A^{n_i} is monochromatic with respect to C. Let u_i be the unique variable word in $\mathrm{V}[(w_j)_{j=N_i}^{N_{i+1}-1}]$ such that $u_i(a) = \mathrm{I}_{W_i}\big(u(a)\big)$ for every $a\in A$, and observe that (C1) and (C2) are satisfied for u_i. The proof of Lemma 4.16 is completed. □

As in Subsection 2.3.2, for every block sequence $\mathcal{X} = (X_0,\dots,X_{n-1})$ of nonempty finite subsets of $\mathbb{N}$ and every $\ell\in[n]$ by $\mathrm{Block}_\ell(\mathcal{X})$ we denote the set of all block subsequences of $\mathcal{X}$ of length ℓ. The following fact is a variant of Fact 2.20.

Fact 4.17. *Let m,r be positive integers and set*

$$L = \mathrm{MT}(2m,m,r^m). \tag{4.30}$$

Also let $\mathcal{X} = (X_0,\dots,X_{L-1})$ be a block sequence and $c\colon \bigcup_{\ell=1}^m \mathrm{Block}_\ell(\mathcal{X})\to[r]$. Then there exists $\mathcal{Z}\in\mathrm{Block}_m(\mathcal{X})$ such that for every $\ell\in[m]$ the set $\mathrm{Block}_\ell(\mathcal{Z})$ is monochromatic.

Proof. We define $C\colon \mathrm{Block}_m(\mathcal{X})\to[r]^m$ by $C(\mathcal{H}) = \langle c(\mathcal{H}\upharpoonright \ell) : \ell\in[m]\rangle$. By the choice of L in (4.30) and Theorem 2.21, there exists $\mathcal{Y}\in\mathrm{Block}_{2m}(\mathcal{X})$ such that the set $\mathrm{Block}_m(\mathcal{Y})$ is monochromatic with respect to C. It is then clear that the block sequence $\mathcal{Z} = \mathcal{Y}\upharpoonright m$ is as desired. The proof of Fact 4.17 is completed. □

We proceed with the following lemma.

Lemma 4.18. *Let k,m,r be positive integers with $k\geqslant 2$. Set*

$$L = \mathrm{MT}(6m,3m,r^{3m})$$

and let $(N_i)_{i=0}^L$ be as in (4.28). Also let A be a finite alphabet with $|A| = k$, $\mathbf{w} = (w_i)_{i=0}^{N_L-1}$ a finite sequence of variable words over A and $c\colon \mathrm{EV}[\mathbf{w}]\to[r]$. Then there exist an extracted subsequence $\mathbf{v} = (v_i)_{i=0}^{m-1}$ of $\mathbf{w}$ and a finite sequence $(r_n)_{n=1}^m$ in $[r]$ with the following property. For every nonempty $F\subseteq\{0,\dots,m-1\}$ and every $(a_i)_{i\in F}\in (A\cup\{x\})^F\setminus A^F$ we have $c\big(\prod_{i\in F} v_i(a_i)\big) = r_n$ if $|\{i\in F : a_i = x\}| = n$.

Proof. Let $(u_i)_{i=0}^{L-1}$ be the reduced subsequence of $\mathbf{w}$ obtained by Lemma 4.16. For every $n\in\{0,\dots,L-1\}$ we set $X_n = \{n\}$ and we observe that the finite sequence $\mathcal{X} = (X_n)_{n=0}^{L-1}$ is block. Moreover, for every $\ell\in\{2,\dots,3m\}$ and every $\mathcal{H} = (H_0,\dots,H_{\ell-1})\in\mathrm{Block}_\ell(\mathcal{X})$ denote by $\Lambda(\mathcal{H})$ the set of all finite sequences $(a_i)_{i\in\cup\mathcal{H}}$ in $A\cup\{x\}$ such that

(a) $a_i\in A$ if and only if $i\in\bigcup\{H_j : j \text{ is even and } 0\leqslant j\leqslant \ell-1\}$ and
(b) $a_i = x$ if and only if $i\in\bigcup\{H_j : j \text{ is odd and } 0\leqslant j\leqslant \ell-1\}$.

Notice that, by Lemma 4.16, for every $\mathcal{H} \in \bigcup_{\ell=2}^{3m} \mathrm{Block}_\ell(\mathcal{X})$ there exists $p_{\mathcal{H}} \in [r]$ such that for every $(a_i)_{i \in \cup\mathcal{H}} \in \Lambda(\mathcal{H})$ we have

$$c\Big(\prod_{i \in \cup\mathcal{H}} u_i(a_i) \Big) = p_{\mathcal{H}}. \tag{4.31}$$

Thus, we may define a coloring $C \colon \bigcup_{\ell=1}^{3m} \mathrm{Block}_\ell(\mathcal{X}) \to [r]$ by the rule $C(\mathcal{H}) = p_{\mathcal{H}}$ if $\mathcal{H} \in \bigcup_{\ell=2}^{3m} \mathrm{Block}_\ell(\mathcal{X})$ and $C(\mathcal{H}) = r$ if $\mathcal{H}$ is a block sequence of length one. By Fact 4.17, there exists a block sequence $\mathcal{Z} = (Z_0, \dots, Z_{3m-1})$ of subsets of $\{0, \dots, L-1\}$ such that for every $\ell \in [3m]$ the set $\mathrm{Block}_\ell(\mathcal{Z})$ is monochromatic with respect to the coloring C. In particular, for every $\ell \in \{2, \dots, 3m\}$ there exists $p_\ell \in [r]$ such that $p_{\mathcal{H}} = p_\ell$ for every $\mathcal{H} \in \mathrm{Block}_\ell(\mathcal{Z})$.

We fix a letter $\alpha \in A$. For every $i \in \{0, \dots, m-1\}$ we set

$$v_i = \Big(\prod_{j \in Z_{3i}} u_j(\alpha) \Big)^\frown \Big(\prod_{j \in Z_{3i+1}} u_j \Big)^\frown \Big(\prod_{j \in Z_{3i+2}} u_j(\alpha) \Big). \tag{4.32}$$

We have the following claim.

Claim 4.19. *Let $n \in [m]$, $F \subseteq \{0, \dots, m-1\}$ and $(a_i)_{i \in F} \in (A \cup \{x\})^F \setminus A^F$ such that $|\{i \in F : a_i = x\}| = n$. Then we have $c\big(\prod_{i \in F} v_i(a_i)\big) = p_{2n+1}$.*

Proof of Claim 4.19. First we observe that $2n+1 \leqslant 3m$ and so the number p_{2n+1} is well defined. For every $i \in \{0, \dots, m-1\}$ we set

$$T_i = Z_{3i} \cup Z_{3i+1} \cup Z_{3i+2}.$$

Also let $i_1 < \dots < i_n$ be the increasing enumeration of the set $\{i \in F : a_i = x\}$. We define $\mathcal{H} = (H_j)_{j=0}^{2n} \in \mathrm{Block}_{2n+1}(\mathcal{Z})$ as follows.

(a) If $j = 0$, then $H_0 = \bigcup \big\{ T_i : i \in F \text{ and } i < i_1 \big\} \cup Z_{3i_1}$.
(b) If $j = 2n$, then $H_{2n} = Z_{3i_n+2} \cup \bigcup \big\{ T_i : i \in F \text{ and } i_n < i \big\}$.
(c) If $j \in [n]$, then $H_{2j-1} = Z_{3i_j+1}$.
(d) If $n \geqslant 2$ and $j \in [n-1]$, then

$$H_{2j} = Z_{3i_j+2} \cup \bigcup \big\{ T_i : i \in F \text{ and } i_j < i < i_{j+1} \big\} \cup Z_{3i_{j+1}}.$$

We also define $(b_q)_{q \in \cup\mathcal{H}} \in (A \cup \{x\})^{\cup\mathcal{H}} \setminus A^{\cup\mathcal{H}}$ by the rule

$$b_q = \begin{cases} x & \text{if } q \in \bigcup\{Z_{3i_j+1} : 1 \leqslant j \leqslant n\} = \bigcup\{H_{2j-1} : 1 \leqslant j \leqslant 2n-1\}, \\ a_i & \text{if } q \in Z_{3i+1} \text{ for some } i \in F \setminus \{i_1, \dots, i_n\}, \\ \alpha & \text{if } q \in \bigcup\{Z_{3i} \cup Z_{3i+2} : i \in F\}. \end{cases} \tag{4.33}$$

Notice that

$$\cup\mathcal{H} = \bigcup_{i \in F} T_i = \bigcup_{i \in F} (Z_{3i} \cup Z_{3i+1} \cup Z_{3i+2})$$

and $b_q = x$ if and only if $q \in \bigcup\{H_j : j \text{ is odd}\}$. Therefore, $(b_q)_{q \in \cup\mathcal{H}} \in \Lambda(\mathcal{H})$. Moreover, it is easy to see that $\prod_{i \in F} v_i(a_i) = \prod_{q \in \cup\mathcal{H}} u_q(b_q)$ and so

$$c\Big(\prod_{i \in F} v_i(a_i) \Big) = c\Big(\prod_{q \in \cup\mathcal{H}} u_q(b_q) \Big) \overset{(4.31)}{=} p_{\mathcal{H}} = p_{2n+1}.$$

The proof of Claim 4.19 is completed. $\square$

For every $n\in[m]$ we set $r_n=p_{2n+1}$. By Claim 4.19, we see that the sequences $\mathbf{v}=(v_i)_{i=0}^{m-1}$ and $(r_n)_{n=1}^{m}$ are as desired. The proof of Lemma 4.18 is completed. $\square$

The following lemma is the last step of the proof of Theorem 4.15.

Lemma 4.20. *Let k,d,r be positive integers with $k\geqslant 2$. Set*

$$L=\mathrm{MT}\big(6\mathrm{H}(d,r),3\mathrm{H}(d,r),r^{3\mathrm{H}(d,r)}\big) \tag{4.34}$$

and let $(N_i)_{i=0}^{L}$ be as in (4.28). Also let A be a finite alphabet with $|A|=k$, $\mathbf{w}=(w_i)_{i=0}^{N_L-1}$ a finite sequence of variable words over A and $c\colon \mathrm{EV}[\mathbf{w}]\to[r]$. Then there exists an extracted subsequence $\mathbf{s}=(s_i)_{i=0}^{d-1}$ of $\mathbf{w}$ such that the set $\mathrm{EV}[\mathbf{s}]$ is monochromatic.

Proof. We set $m=\mathrm{H}(d,r)$. By (4.34) and Lemma 4.18, there exist an extracted subsequence $(v_i)_{i=0}^{m-1}$ of $\mathbf{w}$ and a finite sequence $(r_n)_{n=1}^{m}$ in $[r]$ such that $c\big(\prod_{i\in F}v_i(a_i)\big)=r_n$ for every nonempty $F\subseteq\{0,\dots,m-1\}$ and every $(a_i)_{i\in F}\in(A\cup\{x\})^F\setminus A^F$ with $|\{i\in F:a_i=x\}|=n$. By the choice of m and Proposition 2.19, there exist $p\in[r]$ and a block sequence $\mathcal{F}=(F_0,\dots,F_{d-1})$ of subsets of $\{0,\dots,m-1\}$ such that $r_{|Y|}=p$ for every $Y\in\mathrm{NU}(\mathcal{F})$. We set

$$s_i=\prod_{j\in F_i}v_j$$

for every $i\in\{0,\dots,d-1\}$ and we claim that $\mathbf{s}=(s_i)_{i=0}^{d-1}$ is as desired. Indeed, first observe that $\mathbf{s}$ is an extracted subsequence of $\mathbf{w}$. Let $H\subseteq\{0,\dots,d-1\}$ and $(a_i)_{i\in H}\in(A\cup\{x\})^H\setminus A^H$ be arbitrary. Set $Q=\bigcup_{i\in H}F_i\in\mathrm{NU}(\mathcal{F})$. Moreover, for every $q\in Q$ let $i(q)$ be the unique element of H such that $q\in F_{i(q)}$ and set $b_q=a_{i(q)}$. Notice that

$$\prod_{i\in H}s_i(a_i)=\prod_{i\in H}\Big(\prod_{j\in F_i}v_j(a_i)\Big)=\prod_{q\in Q}v_q(b_q).$$

Also observe that, setting $H'=\{i\in H:a_i=x\}$ and $Y=\bigcup_{i\in H'}F_i$, we have $\{q\in Q:b_q=x\}=Y$. Since $Y\in\mathrm{NU}(\mathcal{F})$ we conclude that

$$c\Big(\prod_{i\in H}s_i(a_i)\Big)=c\Big(\prod_{q\in Q}v_q(b_q)\Big)=r_{|Y|}=p$$

and the proof of Lemma 4.20 is completed. $\square$

We are now ready to complete the proof of Theorem 4.15.

Proof of Theorem 4.15. Fix a triple k,d,r of positive integers with $k\geqslant 2$ and let L be as in (4.34). By Lemma 4.20, we see that

$$\mathrm{C}(k,d,r)\leqslant N_L \tag{4.35}$$

where the number N_L is as in (4.28). Now recall that, by Theorems 2.1 and 2.21 and Proposition 2.19, the Hales–Jewett numbers $\mathrm{HJ}(k,r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^5$, the Milliken–Taylor numbers $\mathrm{MT}(d,m,r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^6$ and, finally, the numbers $\mathrm{H}(d,r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^4$. Hence, by (4.28), (4.34) and (4.35), we conclude that the numbers $\mathrm{C}(k,d,r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^6$. The proof of Theorem 4.15 is completed. $\square$

4.4. Carlson–Simpson spaces

In this section we will present some basic Ramsey properties of Carlson–Simpson spaces. These properties are naturally placed in the general context of this chapter, though their importance will be highlighted in Chapters 5 and 9. The first result in this direction is the following theorem.

Theorem 4.21. *For every quadruple k,d,m,r of positive integers with $k \geqslant 2$ and $d \geqslant m$ there exists a positive integer N with the following property. For every alphabet A with $|A|=k$, every Carlson–Simpson space T of $A^{<\mathbb{N}}$ of dimension at least N and every r-coloring of $\mathrm{SubCS}_m(T)$ there exists $S \in \mathrm{SubCS}_d(T)$ such that the set $\mathrm{SubCS}_m(S)$ is monochromatic. The least positive integer with this property will be denoted by $\mathrm{CS}(k,d,m,r)$.*

Moreover, the numbers $\mathrm{CS}(k,d,m,r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^6$.

As we discussed in Subsection 1.5.1, there is a natural correspondence between Carlson–Simpson spaces and Carlson–Simpson systems. Using this correspondence, we see that Theorem 4.21 is equivalently formulated as follows.

Theorem 4.21′. *Let k,d,m,r be positive integers with $k \geqslant 2$ and $d \geqslant m$. Also let n be an integer with $n \geqslant \mathrm{CS}(k,d,m,r)$. If A is an alphabet with $|A|=k$, then for every Carlson–Simpson system $\mathbf{w}=\langle w,(w_i)_{i=0}^{n-1}\rangle$ over A and every r-coloring of $\mathrm{Subsys}_m(\mathbf{w})$ there exists $\mathbf{u} \in \mathrm{Subsys}_d(\mathbf{w})$ such that $\mathrm{Subsys}_m(\mathbf{u})$ is monochromatic.*

We proceed to the proof of Theorem 4.21.

Proof of Theorem 4.21. We will show that

$$\mathrm{CS}(k,d,m,r) \leqslant \mathrm{GR}(k,d+1,m+1,r)-1 \tag{4.36}$$

for every choice of admissible parameters. By Theorem 2.15, this is enough to complete the proof.

First we need to do some preparatory work. Let A be a finite alphabet with $|A| \geqslant 2$ and define $\Phi\colon \bigcup_{n=1}^{\infty}\bigcup_{\ell=1}^{n}\mathrm{Subsp}_{\ell+1}(A^{n+1}) \to \bigcup_{\ell=1}^{\infty}\mathrm{SubCS}_\ell(A^{<\mathbb{N}})$ as follows. Let $n \geqslant \ell \geqslant 1$ and $V \in \mathrm{Subsp}_{\ell+1}(A^{n+1})$. Let $(X_0,\dots,X_\ell)$ be the sequence of the wildcard sets of V and set $\Phi(V)=\bigcup_{i=0}^{\ell}\{v \upharpoonright \min(X_i) : v \in V\}$. Notice that $\Phi(V)$ is a Carlson–Simpson space of $A^{<\mathbb{N}}$ of dimension ℓ. In particular, the map Φ is well-defined. Also observe that for every $i \in \{0,\dots,\ell\}$ the i-level of $\Phi(V)$ is the set $\{v \upharpoonright \min(X_i) : v \in V\}$. We will need the following elementary fact.

Fact 4.22. *Let $n \geqslant \ell \geqslant 1$. Then $\Phi(A^{n+1})=A^{<n+1}$ and*

$$\Phi\big(\mathrm{Subsp}_{\ell+1}\big(A^{n+1}\big)\big)=\mathrm{SubCS}_\ell(A^{<n+1}). \tag{4.37}$$

More generally, for every $(n+1)$-dimensional combinatorial space W of $A^{<\mathbb{N}}$ we have

$$\Phi\big(\mathrm{Subsp}_{\ell+1}(W)\big) = \mathrm{SubCS}_{\ell}\big(\Phi(W)\big). \tag{4.38}$$

We are ready to show the estimate in (4.36). Fix the parameters k, d, m, r and an integer $n \geqslant \mathrm{GR}(k, d+1, m+1, r)-1$. Let A be an alphabet with $|A| = k$. Also let T be a Carlson–Simpson space of $A^{<\mathbb{N}}$ with $\dim(T) = n$ and $c\colon \mathrm{SubCS}_m(T) \to [r]$ a coloring. By Fact 4.22, there exists a $(n+1)$-dimensional combinatorial space W of $A^{<\mathbb{N}}$ such that $\Phi(W) = T$. Notice that $n \geqslant m$ and so, by (4.38), we have $\Phi\big(\mathrm{Subsp}_{m+1}(W)\big) = \mathrm{SubCS}_m(T)$. Therefore, by restricting Φ on the set of all $(m+1)$-dimensional combinatorial subspaces of W, we see that the map $c \circ \Phi$ is an r-coloring of $\mathrm{Subsp}_{m+1}(W)$. Since $\dim(W) = n+1 \geqslant \mathrm{GR}(k, d+1, m+1, r)$, by Theorem 2.15, there exists a $(d+1)$-dimensional combinatorial subspace V of W such that the set $\mathrm{Subsp}_{m+1}(V)$ is monochromatic with respect to $c \circ \Phi$. We set $S = \Phi(V)$. By Fact 4.22, we see that $S \in \mathrm{SubCS}_d(T)$ and, moreover, $\Phi\big(\mathrm{Subsp}_{m+1}(V)\big) = \mathrm{SubCS}_m(S)$. It follows, in particular, that the set $\mathrm{SubCS}_m(S)$ is monochromatic with respect to c. The proof of Theorem 4.21 is completed. □

We close this section with the following infinite version of Theorem 4.21.

THEOREM 4.23. *Let A be a finite alphabet with $|A| \geqslant 2$. Also let m be a positive integer. Then for every infinite-dimensional Carlson–Simpson space T of $A^{<\mathbb{N}}$ and every finite coloring of the set $\mathrm{SubCS}_m(T)$ there exists an infinite-dimensional Carlson–Simpson subspace S of T such that the set $\mathrm{SubCS}_m(S)$ is monochromatic.*

PROOF. We fix an infinite-dimensional Carlson–Simpson space T of $A^{<\mathbb{N}}$ and a finite coloring $c\colon \mathrm{SubCS}_m(T) \to [r]$. Let $\langle t, (t_n)\rangle$ be the Carlson–Simpson system generating T. We set $w_0 = t^\frown t_0$ and $w_n = t_n$ for every $n \geqslant 1$. Clearly $\mathbf{w} = (w_n)$ is a sequence of variable words over A.

As in the proof of Theorem 4.5, we write (uniquely) every variable word v over A as $v^{*\frown} v^{**}$ where v^* is a word over A and v^{**} is a left variable word over A. Using this decomposition we define two maps $\Psi\colon \mathrm{V}_\infty[\mathbf{w}] \to \mathrm{SubCS}_\infty(T)$ and $\psi\colon \mathrm{V}_{m+1}[\mathbf{w}] \to \mathrm{SubCS}_m(T)$ as follows. If $(z_n) \in \mathrm{V}_\infty[\mathbf{w}]$, then let $\Psi\big((z_n)\big)$ be the infinite-dimensional Carlson–Simpson subspace of T which is generated by the Carlson–Simpson system $\langle z_0^*, (z_n^{**\frown} z_{n+1}^*)\rangle$. Respectively, if $(u_n)_{n=0}^m \in \mathrm{V}_{m+1}[\mathbf{w}]$, then let $\psi\big((u_n)_{n=0}^m\big)$ be the m-dimensional Carlson–Simpson subspace of T which is generated by the Carlson–Simpson system $\langle u_0^*, (u_n^{**\frown} u_{n+1}^*)_{n=0}^{m-1}\rangle$. Notice that

$$\Psi\big(\mathrm{V}_\infty[\mathbf{z}]\big) = \mathrm{SubCS}_\infty\big(\Psi(\mathbf{z})\big) \ \text{ and } \ \psi\big(\mathrm{V}_{m+1}[\mathbf{z}]\big) = \mathrm{SubCS}_m\big(\Psi(\mathbf{z})\big) \tag{4.39}$$

for every $\mathbf{z} \in \mathrm{V}_\infty[\mathbf{w}]$. Moreover, $\Psi(\mathbf{w}) = T$ and $\psi\big(\mathrm{V}_{m+1}[\mathbf{w}]\big) = \mathrm{SubCS}_m(T)$ and so the map $C = c \circ \psi$ is an r-coloring of $\mathrm{V}_{m+1}[\mathbf{w}]$. By Theorem 4.9, there exist $p \in [r]$ and $\mathbf{v} \in \mathrm{V}_\infty[\mathbf{w}]$ such that $\mathrm{V}_{m+1}[\mathbf{v}] \subseteq C^{-1}(\{p\})$. Setting $S = \Psi(\mathbf{v})$, we see that $S \in \mathrm{SubCS}_\infty(T)$ and $\mathrm{SubCS}_m(S) \subseteq c^{-1}(\{p\})$. The proof of Theorem 4.23 is thus completed. □

4.5. Notes and remarks

4.5.1. Theorem 4.1 is the content of Lemma 9.6 in [**C**]. Carlson worked in a more general context and studied Ramsey properties of sequences of multivariable words. As such, his arguments are somewhat different. The proof we presented is more streamlined and is taken from [**HS**, Theorem 18.23]. Closely related proofs appear in [**FK3**] and [**BBH**].

After the seminal work of Carlson, colorings of variable words have been studied by several authors; see, e.g., [**BBH, HM2, FK3**]. Another result in this direction was obtained by Gowers in [**Go1**] (see, also, [**Ka1**]). Gowers' work was motivated by a problem concerning the geometry of the Banach space c_0.

4.5.2. The Carlson–Simpson theorem was the first infinite-dimensional extension of the Hales–Jewett and influenced, strongly, all subsequent related advances in Ramsey theory. It appears as Theorem 6.3 in [**CS**]. The original proof was combinatorial in nature and was based on a method invented by Baumgartner in his proof [**Bau**] of Hindman's theorem [**H**].

We also note that there are several results in the literature related to the Carlson–Simpson theorem. Theorem 4.6 is of course in this direction, though closer to the spirit of the Carlson–Simpson theorem is the work of McCutcheon in [**McC2**]. Other variants appear in [**BBH**] and [**HM1**].

4.5.3. All applications of Theorem 4.1 presented in Section 4.2 were observed in [**C**]. Carlson also extended Theorems 4.7 and 4.9 for definable partitions of infinite sequences of variable words (see Theorems 2 and 12 in [**C**]). These results are in the spirit of Theorem 3.15 and are obtained by implementing Theorem 4.1 in the method developed by Galvin and Prikry [**GP**], and Ellentuck [**E**]. Detailed presentations can be found in [**HS, To**].

4.5.4. The primitive recursive bounds obtained by Theorems 4.12 and 4.15 are new. Using these theorems one can obtain, of course, quantitative analogues of all the results presented in Section 4.2. For instance, we have the following finite version of Theorem 4.7.

THEOREM 4.24. *For every quadruple k, d, m, r of positive integers with $k \geqslant 2$ and $d \geqslant m$ there exists a positive integer N with the following property. If A is an alphabet with $|A| = k$, then for every finite sequence $\mathbf{w}$ of variable words over A of length at least N and every r-coloring of the set $\mathrm{EV}_m[\mathbf{w}]$ there exists $\mathbf{v} \in \mathrm{EV}_d[\mathbf{w}]$ such that the set $\mathrm{EV}_m[\mathbf{v}]$ is monochromatic. The least positive integer with this property will be denoted by $\mathrm{C}(k, d, m, r)$.*

Moreover, the numbers $\mathrm{C}(k, d, m, r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^8$.

4.5.5. Theorem 4.21 was observed in [**DKT3**]. We also note that there is a version of Theorem 4.23 which is analogous to Theorem 3.15 and concerns definable partitions of infinite-dimensional Carlson–Simpson spaces. This result can be proved arguing as in the proof of Theorem 4.23 and using [**C**, Theorem 12] instead of Theorem 4.9.

CHAPTER 5

Finite sets of words

In this chapter we study colorings of arbitrary nonempty finite sets of words. Specifically, given a finite alphabet A with at least two letters, we will characterize the Ramsey classes of finite subsets of $A^{<\mathbb{N}}$. This is achieved by introducing the *type* of a nonempty finite subset F of $A^{<\mathbb{N}}$, an isomorphic invariant which encodes a canonical embedding of F in a substructure of $A^{<\mathbb{N}}$. Substructures of interest in this context are combinatorial spaces and Carlson–Simpson spaces of sufficiently large dimension.

5.1. Subsets of combinatorial spaces

5.1.1. Definitions. Let A be a finite alphabet with $|A| \geqslant 2$ and $<_A$ a linear order on A. For every $a \in A$ and every integer $p \geqslant 1$ let a^p be as in (2.1) and set

$$\Delta(A^p) = \{a^p : a \in A\} \subseteq A^p. \tag{5.1}$$

Also let F be a nonempty subset of A^n for some $n \in \mathbb{N}$ and set $p = |F|$. We are about to define the following objects related to the set F.

5.1.1.1. *The word representation* $\mathrm{R}(F)$ *of* F. If $F = \{w\}$ for some $w \in A^n$, then we set $\mathrm{R}(F) = w$. Assume that $p \geqslant 2$ and let $w_0 <_{\mathrm{lex}} \cdots <_{\mathrm{lex}} w_{p-1}$ be the lexicographical increasing enumeration of F. For every $i \in \{0, \dots, n-1\}$ and every $j \in \{0, \dots, p-1\}$ let $w_{i,j}$ be the i-th coordinate of w_j and set

$$\alpha_i = (w_{i,0}, \dots, w_{i,p-1}).$$

We define

$$\mathrm{R}(F) = (\alpha_0, \dots, \alpha_{n-1}). \tag{5.2}$$

Notice that $\mathrm{R}(F)$ is a word over the alphabet A^p of length n.

5.1.1.2. *The type* $\tau(F)$ *of* F. If $p = 1$, then we define $\tau(F)$ to be the empty word. If $p \geqslant 2$, then let $\mathrm{R}(F) = (\alpha_0, \dots, \alpha_{n-1})$ be the word representation of F and set

$$X(F) = \big\{i \in \{0, \dots, n-1\} : \alpha_i \in A^p \setminus \Delta(A^p)\big\}. \tag{5.3}$$

Observe that the set $X(F)$ is nonempty. Also note that there exists a unique block sequence $\mathcal{X}(F) = (X_0, \dots, X_{m-1})$ of nonempty subsets of $X(F)$ satisfying the following properties.

(P1) We have $X(F) = \cup \mathcal{X}(F)$.
(P2) For every $\ell \in \{0, \dots, m-1\}$ the located word $\mathrm{R}(F) \upharpoonright X_\ell$ is constant.
(P3) If $m \geqslant 2$, then for every $\ell \in \{0, \dots, m-2\}$, every $i \in X_\ell$ and every $i' \in X_{\ell+1}$ we have $\alpha_i \neq \alpha_{i'}$.

For every $\ell \in \{0,\dots,m-1\}$ let τ_ℓ be the unique letter of $A^p \setminus \Delta(A^p)$ such that $\alpha_i = \tau_\ell$ for every $i \in X_\ell$ and define

$$\tau(F) = (\tau_0,\dots,\tau_{m-1}). \tag{5.4}$$

That is, $\tau(F)$ is the word over the alphabet $A^p \setminus \Delta(A^p)$ of length at most n which is obtained by erasing first all letters of $\mathrm{R}(F)$ which belong to $\Delta(A^p)$, then shortening the runs of the same letters of $A^p \setminus \Delta(A^p)$ to singletons and, finally, pushing everything back together.

Example 5.1. Let $A=[3]$ equipped with its natural order and let F be the subset of $[3]^4$ consisting of the words $(1,1,2,2)$, $(1,1,2,1)$, $(2,2,2,1)$ and $(2,2,2,3)$. Write the set F in lexicographical increasing order as

$$\{(1,1,2,1) <_{\mathrm{lex}} (1,1,2,2) <_{\mathrm{lex}} (2,2,2,1) <_{\mathrm{lex}} (2,2,2,3)\}$$

and note that $\alpha_0=(1,1,2,2)$, $\alpha_1=(1,1,2,2)$, $\alpha_2=(2,2,2,2)$ and $\alpha_3=(1,2,1,3)$. Therefore,

$$\mathrm{R}(F) = \big((1,1,2,2),(1,1,2,2),(2,2,2,2),(1,2,1,3)\big)$$

and $\tau(F) = \big((1,1,2,2),(1,2,1,3)\big)$.

5.1.2. Basic properties. We first observe that the type is, essentially, independent of the choice of the linear order on the alphabet A. Indeed, let $<_A, <'_A$ be two linear orders on A. Also let $F \subseteq A^n$ and $G \subseteq A^l$ for some $n,l \in \mathbb{N}$ and assume that both are nonempty. Denote by $\tau(F),\tau(G)$ the types of F,G when computed using the linear order $<_A$ and by $\tau'(F),\tau'(G)$ the types of F,G when computed using the linear order $<'_A$. Then notice that $\tau(F)=\tau(G)$ if and only if $\tau'(F)=\tau'(G)$. In light of this remark, in what follows we will not refer explicitly to the linear order which is used to define the type.

We proceed with the following lemma which asserts that the type is preserved under canonical isomorphisms.

Lemma 5.1. *Let A be a finite alphabet with $|A| \geqslant 2$. Also let $d \in \mathbb{N}$ and F a nonempty subset of A^d. Finally, let V be a d-dimensional combinatorial space of $A^{<\mathbb{N}}$ and I_V the canonical isomorphism associated with V (see Definition* 1.2*). Then we have $\tau(F)=\tau\big(\mathrm{I}_V(F)\big)$.*

Proof. Let v be the d-variable word over A which generates V and notice that the set $G=\{v(a_0,\dots,a_{d-1}) : (a_0,\dots,a_{d-1}) \in F\}$ has the same type with F. Since $\mathrm{I}_V(F)=G$, the proof of Lemma 5.1 is completed. □

Let A be a finite alphabet with $|A| \geqslant 2$ and set

$$\mathcal{T} = \{\tau(F) : F \text{ is a nonempty subset of } A^n \text{ for some } n \in \mathbb{N}\}.$$

Let $\tau=(\tau_0,\dots,\tau_{n-1}) \in \mathcal{T}$ be nonempty and observe that there exists a unique positive integer $p(\tau)$ such that τ is a word over $A^{p(\tau)}$. Also notice that there is a canonical way to "decode" τ and produce a set of type τ. Specifically, for every $i \in \{0,\dots,n-1\}$ and every $j \in \{0,\dots,p(\tau)-1\}$ let $a_{i,j}$ be the j-th coordinate of τ_i and set

$$[\tau] = \big\{(a_{0,j},\dots,a_{n-1,j}) : 0 \leqslant j \leqslant p(\tau)-1\big\}. \tag{5.5}$$

Observe that $[\tau]$ is a subset of $A^{|\tau|}$, has cardinality $p(\tau)$ and is of type τ. More generally, let V be a combinatorial space of $A^{<\mathbb{N}}$ with $\dim(V) = |\tau|$. Let I_V be the canonical isomorphism associated with V and define

$$[V, \tau] = \mathrm{I}_V\big([\tau]\big). \tag{5.6}$$

By Lemma 5.1 and the previous remarks, we obtain the following fact.

Fact 5.2. *Let A be a finite alphabet with $|A| \geqslant 2$. Also let $\tau \in \mathcal{T}$ be nonempty and V a combinatorial space of $A^{<\mathbb{N}}$ with $\dim(V) = |\tau|$. We set $F = [V, \tau]$. Then we have $F \subseteq V$, $|F| = p(\tau)$ and $\tau(F) = \tau$.*

We will need a converse of Fact 5.2. More precisely, given a nonempty subset F of A^n for some $n \in \mathbb{N}$, we seek for a combinatorial space V of $A^{<\mathbb{N}}$ of dimension $|\tau(F)|$ such that $[V, \tau(F)] = F$. It turns out that, in this context, this problem has a very satisfactory answer.

Lemma 5.3. *Let A be a finite alphabet with $|A| \geqslant 2$. Also let $n \in \mathbb{N}$ and $F \subseteq A^n$ with $|F| \geqslant 2$, and set $m = |\tau(F)|$. Then there exists a unique m-dimensional combinatorial space W of $A^{<\mathbb{N}}$ such that:* (i) *$[W, \tau(F)] = F$, and* (ii) *$W \subseteq V$ for every combinatorial space V of $A^{<\mathbb{N}}$ with $F \subseteq V$.*

We will denote by $\mathrm{Env}(F)$ the combinatorial space obtained by Lemma 5.3 and we will call it the *envelope* of F. We proceed to the proof of Lemma 5.3.

Proof of Lemma 5.3. Let $\mathrm{R}(F) = (\alpha_0, \dots, \alpha_{n-1})$ be the word representation of F. Also let $X(F)$ be as in (5.3) and set

$$S(F) = \{0, \dots, n-1\} \setminus X(F). \tag{5.7}$$

Notice that if $S(F) \neq \emptyset$, then for every $i \in S(F)$ there exists a unique $a_i \in A$ such that $\alpha_i = a_i^p$. Finally, let $\mathcal{X}(F) = (X_0, \dots, X_{m-1})$ be the block sequence satisfying properties (P1)–(P3) in Subsection 5.1.1. We define W to be the combinatorial subspace of A^n with wildcard sets $X_0, \dots, X_{m-1}$ and constant part $(f_i)_{i \in S(F)}$ where $f_i = a_i$ for every $i \in S(F)$. We will show that W is as desired.

To this end notice, first, that $\dim(W) = |\tau(F)|$ and $[W, \tau(F)] = F$. Next let V be a combinatorial space of $A^{<\mathbb{N}}$ with $F \subseteq V$. (This implies, in particular, that $V \subseteq A^n$.) Let $Y_0, \dots, Y_{d-1}$ be the wildcard sets of V, Σ the set of its fixed coordinates and $(g_i)_{i \in \Sigma} \in A^\Sigma$ its constant part. Since $F \subseteq V$, there exists a block sequence $(H_0, \dots, H_{m-1})$ of nonempty finite subsets of $\{0, \dots, d-1\}$ such that

(a) for every $\ell \in \{0, \dots, m-1\}$ we have $X_\ell = \bigcup_{j \in H_\ell} Y_j$,
(b) for every $j \in \{0, \dots, d-1\}$ if $Y_j \cap S(F) \neq \emptyset$, then $Y_j \subseteq S(F)$ and the located word $f \upharpoonright Y_j$ is constant, and
(c) for every $i \in S(F) \cap \Sigma$ we have $f_i = g_i$.

Using (a), (b) and (c), we conclude that $W \subseteq V$. Finally, let U be an arbitrary m-dimensional combinatorial space of $A^{<\mathbb{N}}$ satisfying (i) and (ii). Observe that $F = [U, \tau(F)] \subseteq U$ and so, by property (ii) applied for "$V = U$", we obtain that $W \subseteq U$. With the same reasoning and by switching the roles of U and W, we see that $U \subseteq W$. Therefore, $W = U$ and the proof of Lemma 5.3 is completed. $\square$

We close this subsection with the following result which shows that finite sets of words of a given type are ubiquitous.

LEMMA 5.4. *Let A be a finite alphabet with $|A| \geqslant 2$. Also let $n \in \mathbb{N}$ and F a nonempty subset of A^n. Then for every combinatorial space V of $A^{<\mathbb{N}}$ of dimension at least $|\tau(F)|$ there exists a subset G of V with $\tau(G) = \tau(F)$.*

In particular, for every positive integer d the set

$$\mathcal{T}_d = \{\tau(F) : F \subseteq A^n \text{ for some } n \in \mathbb{N} \text{ and } |\tau(F)| \leqslant d\}$$

has cardinality at most $2^{|A|^d}$.

PROOF. We set $p = |F|$ and $m = |\tau(F)|$. We may assume, of course, that $p \geqslant 2$ and $m \geqslant 1$. Fix $U \in \mathrm{Subsp}_m(V)$ and set $G = [U, \tau(F)]$. By Fact 5.2, we see that the set G is as desired.

Now let d be a positive integer and notice that, by the previous discussion, for every $\tau \in \mathcal{T}_d$ we may select a subset G_τ of A^d with $\tau(G_\tau) = \tau$. Observe that if two subsets G, G' of A^d have different types, then they are distinct. It follows that the map $\mathcal{T}_d \ni \tau \mapsto G_\tau \in \mathcal{P}(A^d)$ is an injection, and so $|\mathcal{T}_d| \leqslant |\mathcal{P}(A^d)| = 2^{|A|^d}$. The proof of Lemma 5.4 is completed. □

5.1.3. The main result. It is easy to see that there is no analogue of Ramsey's classical theorem for colorings of subsets of combinatorial spaces of a fixed cardinality. Indeed, let A be a finite alphabet with $|A| \geqslant 2$ and $d, \ell \in \mathbb{N}$ with $|A|^d \geqslant \ell \geqslant 2$. Also let W be a combinatorial space of $A^{<\mathbb{N}}$ of dimension at least $d+1$ and define a coloring c of $\binom{W}{\ell}$ as follows. Let $F \in \binom{W}{\ell}$ be arbitrary and set $c(F) = \tau(F)$ if the type of F has length at most d; otherwise set $c(F) = 0$. Regardless of how large the dimension of W is, by Lemma 5.4 we see that for every $V \in \mathrm{Subsp}_{d+1}(W)$ the coloring c restricted on $\binom{V}{\ell}$ takes all possible colors.

It turns out, however, that colorings which depend on the type are the only obstacles to the Ramsey property. Specifically we have the following theorem.

THEOREM 5.5. *For every triple k, d, r of positive integers with $k \geqslant 2$ there exists a positive integer N with the following property. If $n \geqslant N$ and A is an alphabet with $|A| = k$, then for every n-dimensional combinatorial space W of $A^{<\mathbb{N}}$ and every r-coloring of $\mathcal{P}(W)$ there exists $V \in \mathrm{Subsp}_d(W)$ such that every pair of nonempty subsets of V with the same type is monochromatic. The least positive integer with this property will be denoted by $\mathrm{RamSp}(k, d, r)$.*

Moreover, the numbers $\mathrm{RamSp}(k, d, r)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^6$.

The proof of Theorem 5.5 is based on the following fact.

FACT 5.6. *Let k, d, r be positive integers with $k \geqslant 2$. Also let A be an alphabet with $|A| = k$ and W a combinatorial space of $A^{<\mathbb{N}}$ with*

$$\dim(W) \geqslant \mathrm{GR}\big(k, 2d, d, r^{d+1}\big). \tag{5.8}$$

Then for every coloring $c\colon W \cup \bigcup_{m=1}^{d} \mathrm{Subsp}_m(W) \to [r]$ there exists $V \in \mathrm{Subsp}_d(W)$ such that: (i) *$c(v_1) = c(v_2)$ for every $v_1, v_2 \in V$, and* (ii) *$c(X) = c(Y)$ for every $m \in [d]$ and every $X, Y \in \mathrm{Subsp}_m(V)$.*

Proof. Fix $a \in A$. For every $U \in \mathrm{Subsp}_d(W)$ and every $0 \leqslant m \leqslant d$ let

$$U(a,m) = \big\{ \mathrm{I}_U(t^\frown a^{d-m}) : t \in A^m \big\} \tag{5.9}$$

where I_U is the canonical isomorphism associated with U and a^{d-m} is as in (2.1). Observe that $U(a,0) \in W$ while $U(a,m) \in \mathrm{Subsp}_m(W)$ for every $m \in [d]$. We define a coloring $C\colon \mathrm{Subsp}_d(W) \to [r^{d+1}]$ by the rule

$$C(U) = \big\langle c\big(U(a,m)\big) : m \in \{0,\dots,d\} \big\rangle.$$

By Theorem 2.15 and (5.8), there is $Y \in \mathrm{Subsp}_{2d}(W)$ such that the set $\mathrm{Subsp}_d(Y)$ is monochromatic with respect to C. Notice that this is equivalent to saying that for every $m \in \{0,\dots,d\}$ there exists $r_m \in [r]$ such that $c\big(U(a,m)\big) = r_m$ for every $U \in \mathrm{Subsp}_d(Y)$. Write $Y = Y_1^\frown Y_2$ where Y_1 and Y_2 are both d-dimensional combinatorial spaces of $A^{<\mathbb{N}}$ and set

$$V = Y_1^\frown \mathrm{I}_{Y_2}(a^d) \in \mathrm{Subsp}_d(Y). \tag{5.10}$$

We claim that V is as desired. We will argue only for part (ii) since the verification of part (i) is similar. Fix $m \in [d]$ and $\widetilde{X} \in \mathrm{Subsp}_m(V)$. By the choice of V in (5.10), we see that $\widetilde{X}$ is of the form $X^\frown \mathrm{I}_{Y_2}(a^d)$ for some (unique) $X \in \mathrm{Subsp}_m(Y_1)$. Set $U = X^\frown\{\mathrm{I}_{Y_2}(s^\frown a^m) : s \in A^{d-m}\}$ and notice that $U \in \mathrm{Subsp}_d(Y)$ and

$$U(a,m) \overset{(5.9)}{=} \mathrm{I}_X(A^m)^\frown \mathrm{I}_{Y_2}(a^{d-m\frown} a^m) = X^\frown \mathrm{I}_{Y_2}(a^d) = \widetilde{X}.$$

Therefore, $c(\widetilde{X}) = c\big(U(a,m)\big) = r_m$ and the proof of Fact 5.6 is completed. $\square$

We proceed to the proof of Theorem 5.5.

Proof of Theorem 5.5. Fix a triple k,d,r of positive integers with $k \geqslant 2$ and set $\rho = r^{2^{k^d}}$. We will show that

$$\mathrm{RamSp}(k,d,r) \leqslant \mathrm{GR}\big(k,2d,d,\rho^{d+1}\big). \tag{5.11}$$

By Theorem 2.15, this is enough to complete the proof. To this end, fix an alphabet A with $|A| = k$ and a combinatorial space W of $A^{<\mathbb{N}}$ with

$$\dim(W) \geqslant \mathrm{GR}\big(k,2d,d,\rho^{d+1}\big). \tag{5.12}$$

Let $c\colon \mathcal{P}(W) \to [r]$ be a coloring and let $\mathcal{T}_d$ be as in Lemma 5.4. For every $\tau \in \mathcal{T}_d$ we will define an r-coloring C_τ of the set $W \cup \bigcup_{m=1}^d \mathrm{Subsp}_m(W)$ as follows. Assume, first, that τ is the empty word. Then for every $w \in W$ we set $C_\tau(w) = c(\{w\})$; if $X \in \mathrm{Subsp}_m(W)$ for some $m \in [d]$, then we set $C_\tau(X) = 1$. Next assume that τ is nonempty. Then for every $X \in \mathrm{Subsp}_{|\tau|}(W)$ we define $C_\tau(X) = c\big([X,\tau]\big)$ where $[X,\tau]$ is as in (5.6). If $X \in W$ or if $X \in \mathrm{Subsp}_m(W)$ for some $m \in [d]$ with $m \neq |\tau|$, then we set $C_\tau(X) = 1$.

Now define $C\colon W \cup \bigcup_{m=1}^d \mathrm{Subsp}_m(W) \to [r]^{\mathcal{T}_d}$ by the rule

$$C(X) = \langle C_\tau(X) : \tau \in \mathcal{T}_d \rangle. \tag{5.13}$$

By Lemma 5.4, the set $\mathcal{T}_d$ has cardinality at most 2^{k^d}. This implies, of course, that C is a ρ-coloring of $W \cup \bigcup_{m=1}^d \mathrm{Subsp}_m(W)$. Hence, by Fact 5.6 and (5.12), there exists $V \in \mathrm{Subsp}_d(W)$ such that: (i) C is constant on V, and (ii) C is constant on $\mathrm{Subsp}_m(V)$ for every $m \in [d]$. We claim that V is as desired. Indeed, fix a pair G, G' of nonempty subsets of V with $\tau(G) = \tau(G')$. Set $\tau = \tau(G)$ and notice that $\tau \in \mathcal{T}_d$. If τ is the empty word, then both G and G' are singletons. Since C is constant on V, this is easily seen to imply that $c(G) = c(G')$. So assume that $|\tau| \geqslant 1$. Set $m = |\tau|$ and observe that $m \in [d]$. Let $X = \mathrm{Env}(G)$, $Y = \mathrm{Env}(G')$ and notice that, by Lemma 5.3, we have $X, Y \in \mathrm{Subsp}_m(V)$. Using the fact that C is constant on $\mathrm{Subsp}_m(V)$ and invoking Lemma 5.3 once again, we conclude that

$$c(G) = c\big([X,\tau]\big) = C_\tau(X) = C_\tau(Y) = c\big([Y,\tau]\big) = c(G').$$

The proof of Theorem 5.5 is completed. $\square$

5.2. Subsets of Carlson–Simpson spaces

5.2.1. Definitions. Let A be a finite alphabet with $|A| \geqslant 2$. Fix a linear order $<_A$ on A and a letter χ not belonging to A. For every positive integer p set

$$A^p_\chi = \bigcup_{q=0}^{p-1} \{\chi^q{}^\frown w : w \in A^{p-q}\} \ \text{ and } \ \Delta(A^p_\chi) = \bigcup_{q=0}^{p-1} \{\chi^q{}^\frown a^{p-q} : a \in A\} \tag{5.14}$$

where χ^q and a^{p-q} are as in (2.1). Observe that $\Delta(A^p_\chi) \subseteq A^p_\chi$.

Now let F be a nonempty finite subset of $A^{<\mathbb{N}}$ and set $p = |F|$. We are about to extend the analysis presented in Subsection 5.1.1 and introduce the word representation $\mathrm{R}(F)$ and the type $\tau(F)$ associated with F.

To this end let $L(F) = \{n \in \mathbb{N} : F \cap A^n \neq \emptyset\}$ be the level set of F and set $L = |L(F)|$. Write the set $L(F)$ in increasing order as $n_0 < \cdots < n_{L-1}$ and set $p_l = |F \cap A^{n_l}|$ for every $l \in \{0, \dots, L-1\}$. Notice that $p = \sum_{l=0}^{L-1} p_l$. Also set $I_0 = \{n \in \mathbb{N} : n < n_0\}$ and $I_l = \{n \in \mathbb{N} : n_{l-1} \leqslant n < n_l\}$ if $L \geqslant 2$ and $l \in [L-1]$. Observe that the family $\{I_0, \dots, I_{L-1}\}$ is a partition of the set $\{0, \dots, n_{L-1} - 1\}$ into successive intervals. (However, note that the set I_0 could be empty.)

5.2.1.1. *The word representation* $\mathrm{R}(F)$ *of* F. It is a word over the alphabet A^p_χ of length n_{L-1}. For every $i \in \{0, \dots, n_{L-1} - 1\}$ the i-th coordinate α_i of $\mathrm{R}(F)$ is defined as follows. Let $l(i) \in \{0, \dots, L-1\}$ be the unique integer such that $i \in I_{l(i)}$. For every $l \in \{l(i), \dots, L-1\}$ let $\mathrm{R}(F \cap A^{n_l})$ be the word representation of the set $F \cap A^{n_l}$ computed using the linear order $<_A$. Recall that $\mathrm{R}(F \cap A^{n_l})$ is a word $(\alpha_{l,0}, \dots, \alpha_{l,n_l-1})$ of length n_l over the alphabet A^{p_l}. Also notice that $n_l \geqslant n_{l(i)} > i$. We set

$$q(i) = p - \sum_{l=l(i)}^{L-1} p_l \tag{5.15}$$

and we define

$$\alpha_i = \chi^{q(i)}{}^\frown\Big(\prod_{l=l(i)}^{L-1} \alpha_{l,i}\Big) = \chi^{q(i)}{}^\frown \alpha_{l(i),i}{}^\frown \dots {}^\frown \alpha_{L-1,i} \in A^p_\chi. \tag{5.16}$$

Observe that if $L = 1$ (equivalently, if $F \subseteq A^{n_0}$), then this definition leads to the word representation of F as described in Subsection 5.1.1.

5.2.1.2. *The type* $\tau(F)$ *of* F. If $L = 1$, then the type of F is as defined in Subsection 5.1.1. So assume that $L \geqslant 2$. Let $\mathrm{R}(F) = (\alpha_0, \dots, \alpha_{n_{L-1}-1})$ be the word representation of F. For every $l \in \{0, \dots, L-2\}$ we define $Y_l \subseteq I_{l+1}$ by

$$\begin{aligned} i \in Y_l \quad \Leftrightarrow \quad & \alpha_i = \alpha_{n_l} \text{ and for every } j \in \{n_l, \dots, i\} \\ & \text{we have that either } \alpha_j \in \Delta(A^p_\chi) \text{ or } \alpha_j = \alpha_{n_l}. \end{aligned} \tag{5.17}$$

Note that for every $l \in \{0, \dots, L-2\}$ the located word $\mathrm{R}(F) \upharpoonright Y_l$ is constant and $n_l = \min(Y_l)$. Moreover, the finite sequence $\mathcal{Y}(F) = (Y_0, \dots, Y_{L-2})$ is a block sequence of subsets of $\{n_0, \dots, n_{L-1} - 1\}$. Next we set

$$Y(F) = \cup\mathcal{Y}(F) \ \text{ and } \ X(F) = \big\{ i \in \{0, \dots, n_{L-1}-1\} \setminus Y(F) : \alpha_i \notin \Delta(A^p_\chi) \big\}. \tag{5.18}$$

If $X(F) \neq \emptyset$, then there exists a unique block sequence $\mathcal{X}(F) = (X_0, \dots, X_{M-1})$ of nonempty subsets of $X(F)$ with the following properties.

- (P1) We have $X(F) = \cup\mathcal{X}(F)$.
- (P2) For every $j \in \{0, \dots, M-1\}$ the located word $\mathrm{R}(F) \upharpoonright X_j$ is constant.
- (P3) If $M \geqslant 2$, then for every $j \in \{0, \dots, M-2\}$, every $i \in X_j$ and every $i' \in X_{j+1}$ we have $\alpha_i \neq \alpha_{i'}$.

Observe that for every $l \in \{0, \dots, L-2\}$ and every $j \in \{0, \dots, M-1\}$ we have that either $\max(Y_l) < \min(X_j)$ or $\max(X_j) < \min(Y_l)$. Therefore, there exists a unique block sequence $\mathcal{Z}(F) = (Z_0, \dots, Z_{m-1})$, where $m = (L-1) + M$, such that each of the coordinates of $\mathcal{Z}(F)$ is a coordinate of either $\mathcal{Y}(F)$ or $\mathcal{X}(F)$. In particular, for every $\ell \in \{0, \dots, m-1\}$ there exists $\tau_\ell \in A^p_\chi$ such that $\alpha_i = \tau_\ell$ for every $i \in Z_\ell$. We define the type of F by the rule

$$\tau(F) = (\tau_0, \dots, \tau_{m-1}). \tag{5.19}$$

Notice that $\tau(F)$ is a word over the alphabet A^p_χ with $L - 1 \leqslant |\tau(F)| \leqslant n_{L-1}$.

EXAMPLE 5.2. Let $A = [3]$ equipped with its natural order and F the subset of $[3]^{<\mathbb{N}}$ consisting of the words (1), $(2,2,2,2)$, $(2,2,1,2)$ and $(1,2,1,2,3)$. Notice that $L(F) = \{1,4,5\}$ and $F \cap [3]^1 = \{(1)\}$, $F \cap [3]^4 = \{(2,2,1,2) <_{\mathrm{lex}} (2,2,2,2)\}$ and $F \cap [3]^5 = \{(1,2,1,2,3)\}$. Therefore,

$$\mathrm{R}(F) = \big((1,2,2,1), (\chi,2,2,2), (\chi,1,2,1), (\chi,2,2,2), (\chi,\chi,\chi,3)\big)$$

and $\tau(F) = \big((1,2,2,1), (\chi,2,2,2), (\chi,1,2,1), (\chi,\chi,\chi,3)\big)$.

5.2.2. Basic properties. As in Subsection 5.1.2, we remark that the type is an intrinsic characteristic in the sense that the question whether two nonempty finite sets of words over an alphabet A have the same type, is independent of the particular choice of the linear order $<_A$ on A and the letter χ. Thus, we will not refer explicitly to these data when we talk of properties of types.

The following lemma is the analogue of Lemma 5.1 and shows that the type is preserved under canonical isomorphisms of Carlson–Simpson spaces.

LEMMA 5.7. *Let A be a finite alphabet with $|A| \geqslant 2$, $d \in \mathbb{N}$ and F a nonempty subset of $A^{<d+1}$. Also let T be a d-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ and let I_T be the canonical isomorphism associated with T* (*see Definition* 1.10). *Then we have $\tau(F) = \tau\big(\mathrm{I}_T(F)\big)$.*

Proof. We set $p = |F|$ and $L = |L(F)|$. We may assume that $p, L \geqslant 2$ since the other cases follow from Lemma 5.1. Set $G = \mathrm{I}_T(F)$ and observe that $|G| = p$ and $|L(G)| = L$. Notice that the d-level $T(d)$ of T is a d-dimensional combinatorial space of $A^{<\mathbb{N}}$. Let $H_0, \dots, H_{d-1}$ be the wildcard sets of $T(d)$, S the set of its constant coordinates and $(c_j)_{j\in S} \in A^S$ its constant part. Finally, let $\mathrm{R}(F) = (\alpha_0, \dots, \alpha_{n-1})$ and $\mathrm{R}(G) = (\beta_0, \dots, \beta_{\ell-1})$ be the word representations of F and G respectively. Then observe that

(a) for every $i \in \{0, \dots, d-1\}$ and every $j \in H_i$ we have $\beta_j = \alpha_i$, and
(b) for every $j \in S$ we have $\beta_j \in \Delta(A^p_{\chi})$.

Next, let $\mathcal{X}(F)$, $\mathcal{Y}(F)$ and $\mathcal{X}(G)$, $\mathcal{Y}(G)$ be the block sequences used for the definition of $\tau(F)$ and $\tau(G)$ respectively. Using (a) and (b), we see that the following properties are satisfied.

(P1) We have $|\mathcal{Y}(F)| = |\mathcal{Y}(G)| = L-1$ and $|\mathcal{X}(F)| = |\mathcal{X}(G)|$.
(P2) If $\mathcal{Y}(F) = (Y_0, \dots, Y_{L-2})$ and $\mathcal{Y}(G) = (Y'_0, \dots, Y'_{L-2})$, then for every $l \in \{0, \dots, L-2\}$ we have $\bigcup_{i\in Y_l} H_i \subseteq Y'_l$. Moreover, the constant located words $\mathrm{R}(F) \upharpoonright Y_l$ and $\mathrm{R}(G) \upharpoonright Y'_l$ take the same value.
(P3) If $\mathcal{X}(F) = (X_0, \dots, X_{M-1})$ and $\mathcal{X}(G) = (X'_0, \dots, X'_{M-1})$, then for every $j \in \{0, \dots, M-1\}$ we have $X'_j = \bigcup_{i\in X_j} H_i$. Moreover, the constant located words $\mathrm{R}(F) \upharpoonright X_j$ and $\mathrm{R}(G) \upharpoonright X'_j$ take the same value.
(P4) For every $l \in \{0, \dots, L-2\}$ and every $j \in \{0, \dots, M-1\}$ we have $\max(Y_l) < \min(X_j)$ if and only if $\max(Y'_l) < \min(X'_j)$; respectively, we have $\max(X_j) < \min(Y_l)$ if and only if $\max(X'_j) < \min(Y'_l)$.

By properties (P1)–(P4) and the definition of type, we conclude that $\tau(F) = \tau(G)$. The proof of Lemma 5.7 is completed. □

Now let A be a finite alphabet with $|A| \geqslant 2$ and set

$$\mathcal{T} = \{\tau(F) : F \text{ is a nonempty subset of } A^{<\mathbb{N}}\}.$$

Let $\tau = (\tau_0, \dots, \tau_{n-1}) \in \mathcal{T}$ be nonempty and notice that there exists a unique positive integer $p(\tau)$ such that τ is a word over $A^{p(\tau)}_{\chi}$. Our goal is to "decode" τ and produce a set of words of type τ. To this end, for every $i \in \{0, \dots, n-1\}$ and every $j \in \{0, \dots, p(\tau)-1\}$ let $a_{i,j}$ be the j-th coordinate of τ_i. Since $\tau \in \mathcal{T}$, we see that for every $j \in \{0, \dots, p(\tau)-1\}$ there exists $n_j \in \{0, \dots, n\}$ such that $a_{i,j} \in A$ if and only if $i < n_j$. Let w_j be the empty word if $n_j = 0$ while $w_j = (a_{0,j}, \dots, a_{n_j-1,j})$ if $n_j \geqslant 1$, and define

$$[\tau] = \{w_0, \dots, w_{p(\tau)-1}\}. \tag{5.20}$$

Note that $[\tau]$ is a subset of $A^{<|\tau|+1}$, has cardinality $p(\tau)$ and is of type τ. Also observe that this construction can be performed inside any Carlson–Simpson space. Indeed, let T be a Carlson–Simpson space of $A^{<\mathbb{N}}$ of dimension $|\tau|$, let I_T be the canonical isomorphism associated with T and define

$$[T, \tau] = \mathrm{I}_T\big([\tau]\big). \tag{5.21}$$

We have the following fact which follows from Lemma 5.7 taking into account the previous remarks.

FACT 5.8. *Let A be a finite alphabet with $|A| \geqslant 2$. Also let $\tau \in \mathcal{T}$ be nonempty and T a Carlson–Simpson space of $A^{<\mathbb{N}}$ with* $\dim(T) = |\tau|$. *We set $F = [T, \tau]$. Then we have $F \subseteq T$, $|F| = p(\tau)$ and $\tau(F) = \tau$.*

As in Section 5.1, we will need a converse of Fact 5.8. This is the content of the following lemma.

LEMMA 5.9. *Let A be a finite alphabet with $|A| \geqslant 2$. Also let F be a finite subset of $A^{<\mathbb{N}}$ with $|F| \geqslant 2$ and set $m = |\tau(F)|$. Finally, let S be a* (*finite or infinite dimensional*) *Carlson–Simpson space of $A^{<\mathbb{N}}$ with $F \subseteq S$. Then there exists $T \in \mathrm{SubCS}_m(S)$ such that $[T, \tau(F)] = F$.*

We point out that, in contrast with Lemma 5.3, the Carlson–Simpson space T obtained by Lemma 5.9 is not necessarily unique. For instance, let F be the subset of $[2]^{<5}$ consisting of the words (2), $(1,1,1,1)$ and $(1,1,1,2)$ and notice that $\tau(F) = \{(2,1,1), (\chi,1,1), (\chi,1,2)\}$. Also let T_1 and T_2 be the 3-dimensional Carlson–Simpson subspaces of $[2]^{<5}$ generated by the systems $\langle \emptyset, (x), (x,1), (x)\rangle$ and $\langle \emptyset, (x), (x,x), (x)\rangle$ respectively. Clearly T_1 and T_2 are incomparable under inclusion, yet observe that $[T_1, \tau(F)] = [T_2, \tau(F)] = F$.

PROOF OF LEMMA 5.9. Clearly we may assume that S is finite-dimensional. We set $d = \dim(S)$ and we claim that we may also assume that $S = A^{<d+1}$. Indeed, let I_S be the canonical isomorphism associated with S and set $G = \mathrm{I}_S^{-1}(F)$. By Lemma 5.7, we have $G \subseteq A^{<d+1}$ and $\tau(G) = \tau(F)$. Let R be a Carlson–Simpson subspace of $A^{<d+1}$ of dimension m such that $[R, \tau(G)] = G$. We set $T = \mathrm{I}_S(R)$ and we observe that $T \in \mathrm{SubCS}_m(S)$. Moreover, note that $\mathrm{I}_T = \mathrm{I}_S \circ \mathrm{I}_R$ and so

$$[T, \tau(F)] \overset{(5.21)}{=} \mathrm{I}_T\big([\tau(F)]\big) = \mathrm{I}_S\Big(\mathrm{I}_R\big([\tau(F)]\big)\Big) = \mathrm{I}_S\big([R, \tau(G)]\big) = \mathrm{I}_S(G) = F.$$

Hence, in what follows we may assume that $S = A^{<d+1}$.

Let $L(F)$ be the level set of F and set $L = |L(F)|$. Also set $p = |F|$. Assume that $L = 1$ or, equivalently, that $F \subseteq A^n$ for some $n \in [d]$. Consider the envelope $\mathrm{Env}(F)$ of F and notice that, by Lemma 5.3, we have $\mathrm{Env}(F) \in \mathrm{Subsp}_m(A^n)$ and $[\mathrm{Env}(F), \tau(F)] = F$. (Here, $[\mathrm{Env}(F), \tau(F)]$ is as in (5.6).) By Lemma 1.13, there exists a unique m-dimensional Carlson–Simpson subspace R of $A^{<n+1}$ whose m-level $R(m)$ is $\mathrm{Env}(F)$. Since $n \in [d]$, it follows that $R \in \mathrm{SubCS}_m(A^{<d+1})$ and $[R, \tau(F)] = [\mathrm{Env}(F), \tau(F)] = F$.

It remains to deal with the case $|L| \geqslant 2$. Write the set $L(F)$ in increasing order as $n_0 < \cdots < n_{L-1}$ and let $\mathrm{R}(F) = (\alpha_0, \dots, \alpha_{n_{L-1}-1})$ be the word representation of F. Also let $Y(F)$ and $X(F)$ be as in (5.18), and let $\mathcal{Z}(F) = (Z_0, \dots, Z_{m-1})$ be the block sequence of subsets of $\{0, \dots, n_{L-1}-1\}$ which is used to define the type $\tau(F)$ of F. We set

$$S(F) = \{0, \dots, n_{L-1}-1\} \setminus \big(Y(F) \cup X(F)\big). \tag{5.22}$$

Observe that if $S(F) \neq \emptyset$, then for every $i \in S(F)$ we have that $\alpha_i \in \Delta(A^p_\chi)$. Therefore, for every $i \in S(F)$ there exist $a_i \in A$ and $q \in \{0, \dots, p-1\}$, both unique, such that $\alpha_i = \chi^q{}^\frown a_i^{p-q}$. Let W be the m-dimensional combinatorial subspace of $A^{n_{L-1}}$ with wildcard sets $Z_0,\dots,Z_{m-1}$ and constant part $(f_i)_{i\in S(F)}$ where $f_i=a_i$ for every $i\in S(F)$. By Lemma 1.13 once again, there exists a unique m-dimensional Carlson–Simpson subspace R of $A^{<n_{L-1}+1}$ such that $R(m)=W$. Notice that $n_{L-1}\leqslant d$ and so $R\in \mathrm{SubCS}_m(A^{<d+1})$. Moreover, by the definition of $\tau(F)$ and the choice of W, we see that $[R,\tau(F)]=F$. The proof of Lemma 5.9 is completed. □

We close this section with the following analogue of Lemma 5.4.

LEMMA 5.10. *Let A be a finite alphabet with $|A|\geqslant 2$. Also let F be a nonempty subset of $A^{<\mathbb{N}}$. Then for every Carlson–Simpson space T of $A^{<\mathbb{N}}$ of dimension at least $|\tau(F)|$ there exists a subset G of T with $\tau(G)=\tau(F)$.*

In particular, for every positive integer d the set

$$\mathcal{T}_d=\{\tau(F): F\subseteq A^{<\mathbb{N}} \text{ and } |\tau(F)|\leqslant d\}$$

has cardinality at most $2^{|A|^{d+1}}$.

PROOF. We set $p=|F|$ and $m=|\tau(F)|$. We may assume that $p\geqslant 2$. Let T be a Carlson–Simpson space of $A^{<\mathbb{N}}$ with $\dim(T)\geqslant m$. We select $S\in\mathrm{SubCS}_m(T)$ and we set $G=[S,\tau(F)]$. By Fact 5.8, the set G is as desired.

Using this property we see that for every positive integer d there exists an injection $\mathcal{T}_d\ni\tau\mapsto G_\tau\in\mathcal{P}(A^{<d+1})$. Therefore, $|\mathcal{T}_d|\leqslant|\mathcal{P}(A^{<d+1})|\leqslant 2^{|A|^{d+1}}$ and the proof of Lemma 5.10 is completed. □

5.2.3. The main result. We are now in a position to state the main result of this section.

THEOREM 5.11. *For every triple k,d,r of positive integers with $k\geqslant 2$ there exists a positive integer N with the following property. If $n\geqslant N$ and A is an alphabet with $|A|=k$, then for every n-dimensional Carlson–Simpson space T of $A^{<\mathbb{N}}$ and every r-coloring of $\mathcal{P}(T)$ there exists $S\in\mathrm{SubCS}_d(T)$ such that every pair of nonempty subsets of S with the same type is monochromatic. The least positive integer with this property will be denoted by* $\mathrm{RamCS}(k,d,r)$.

Moreover, the numbers $\mathrm{RamCS}(k,d,r)$ *are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^6$.*

Theorem 5.11 is optimal, of course, as can be seen by coloring the subsets of T according to their type. The proof of Theorem 5.11 is based on the following fact.

FACT 5.12. *Let k,d,r be positive integers with $k\geqslant 2$. Also let A be an alphabet with $|A|=k$ and T a Carlson–Simpson space of $A^{<\mathbb{N}}$ with*

$$\dim(T)\geqslant \mathrm{CS}(k,2d,d,r^{d+1}). \tag{5.23}$$

Then for every coloring $c\colon T\cup\bigcup_{m=1}^{d}\mathrm{SubCS}_m(T)\to[r]$ there exists $S\in\mathrm{SubCS}_d(T)$ such that: (i) *$c(s_1)=c(s_2)$ for every $s_1,s_2\in S$, and* (ii) *$c(R)=c(R')$ for every $m\in[d]$ and every $R,R'\in\mathrm{SubCS}_m(S)$.*

PROOF. Fix c and define a coloring $C\colon \mathrm{SubCS}_d(T) \to [r^{d+1}]$ by the rule

$$C(U) = \big\langle c\big(U \upharpoonright m+1\big) : m \in \{0,\dots,d\}\big\rangle$$

where $U \upharpoonright m+1 = U(0)\cup\dots\cup U(m)$ for every $U \in \mathrm{SubCS}_d(T)$ and $m \in \{0,\dots,d\}$. By Theorem 4.21 and (5.23), there exists a $2d$-dimensional Carlson–Simpson subspace Y of T such that the set $\mathrm{SubCS}_d(Y)$ is monochromatic with respect to C. In particular, for every $m \in \{0,\dots,d\}$ there exists $r_m \in [r]$ such that $c(U \upharpoonright m+1) = r_m$ for every $U \in \mathrm{SubCS}_d(Y)$.

We set $S = Y \upharpoonright d+1 \in \mathrm{SubCS}_d(T)$ and we claim that S is as desired. Indeed, let $X \in S \cup \bigcup_{m=1}^{d} \mathrm{SubCS}_m(S)$. We set $m=0$ if $X \in S$; otherwise let $m = \dim(X)$. Observe that there exists $U \in \mathrm{SubCS}_d(Y)$ such that $X = U \upharpoonright m+1$. Therefore, $c(X) = c\big(U \upharpoonright m+1\big) = r_m$ and the proof of Fact 5.12 is completed. □

We proceed to the proof of Theorem 5.11.

PROOF OF THEOREM 5.11. It is similar to the proof of Theorem 5.5. Fix a triple k,d,r of positive integers with $k \geqslant 2$ and set $\rho = r^{2^{k^{d+1}}}$. We will show that

$$\mathrm{RamCS}(k,d,r) \leqslant \mathrm{CS}(k,2d,d,\rho^{d+1}). \tag{5.24}$$

By Theorem 4.21, this is enough to complete the proof. To this end, let A be an alphabet with $|A| = k$ and let T be a Carlson–Simpson space of $A^{<\mathbb{N}}$ of dimension at least $\mathrm{CS}(k,2d,d,\rho^{d+1})$. Fix a coloring $c\colon \mathcal{P}(T) \to [r]$ and let $\mathcal{T}_d$ be as in Lemma 5.10. For every $\tau \in \mathcal{T}_d$ we define an r-coloring C_τ of $T \cup \bigcup_{m=1}^{d} \mathrm{SubCS}_m(T)$ as follows. If τ is nonempty and $R \in \mathrm{SubCS}_{|\tau|}(T)$, then we set $C_\tau(R) = c\big([R,\tau]\big)$ where $[R,\tau]$ is as in (5.21). If τ is the empty word and $t \in T$, then we set $C_\tau(t) = c(\{t\})$. In all other cases we define C_τ to be constantly equal to 1.

Now define $C\colon T \cup \bigcup_{m=1}^{d} \mathrm{SubCS}_m(T) \to [r]^{\mathcal{T}_d}$ by $C(X) = \langle C_\tau(X) : \tau \in \mathcal{T}_d\rangle$. By Lemma 5.10, the set $\mathcal{T}_d$ has cardinality at most $2^{k^{d+1}}$ and so C is an ρ-coloring of $T \cup \bigcup_{m=1}^{d} \mathrm{SubCS}_m(T)$. Therefore, by Fact 5.12, there exists $S \in \mathrm{SubCS}_d(T)$ such that: (i) C is constant on S, and (ii) C is constant on $\mathrm{SubCS}_m(S)$ for every $m \in [d]$. We will show that S is as desired. Let G, G' be a pair of nonempty subsets of S with $\tau(G) = \tau(G')$ and set $\tau = \tau(G)$. Notice that $\tau \in \mathcal{T}_d$. Assume that τ is the empty word or, equivalently, that both G and G' are singletons. In this case, using the fact that C is constant of S, we see that $c(G) = c(G')$. Next assume that $|\tau| \geqslant 1$. Set $m = |\tau|$ and observe that $m \in [d]$. By Lemma 5.9, there exist $R_1, R_2 \in \mathrm{SubCS}_m(S)$ such that $[R_1,\tau] = G$ and $[R_2,\tau] = G'$. Hence,

$$c(G) = c\big([R_1,\tau]\big) = C_\tau(R_1) = C_\tau(R_2) = c\big([R_2,\tau]\big) = c(G')$$

and the proof of Theorem 5.11 is completed. □

5.2.4. Infinite-dimensional Carlson–Simpson spaces. Let A be a finite alphabet A with $|A| \geqslant 2$ and, as in Subsection 5.2.2, let

$$\mathcal{T} = \{\tau(F) : F \text{ is a nonempty subset of } A^{<\mathbb{N}}\}.$$

We have the following infinite version of Theorem 5.11.

THEOREM 5.13. *Let A be a finite alphabet with $|A| \geqslant 2$ and let $\tau \in \mathcal{T}$. Also let T be an infinite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$. Then for every finite coloring of $\mathcal{P}(T)$ there exists an infinite-dimensional Carlson–Simpson subspace S of T such that every pair of subsets of S with type τ is monochromatic.*

Of course, by repeated applications of Theorem 5.13, one can deal simultaneously with any nonempty finite subset of $\mathcal{T}$. However, we point out that an exact infinite-dimensional extension of Theorem 5.11 does not hold true, as is shown in the following example.

EXAMPLE 5.3. Fix a finite alphabet A with $|A| \geqslant 2$. For every nonempty subset F of $A^{<\mathbb{N}}$ let $\wedge F$ be the infimum of F. (Recall that $\wedge F$ is the maximal common initial segment of every $w \in F$.) We define a 2-coloring c of $\mathcal{P}(A^{<\mathbb{N}})$ by the rule

$$c(F) = \begin{cases} 0 & \text{if } F \text{ is nonempty and } |\wedge F| < |\tau(F)|, \\ 1 & \text{otherwise.} \end{cases}$$

Fix an infinite subset $\mathcal{S}$ of $\mathcal{T}$ and an infinite-dimensional Carlson–Simpson space T of $A^{<\mathbb{N}}$. We will show that the set $\mathcal{P}_\tau(T) \coloneqq \{F \subseteq T : \tau(F) = \tau\}$ is not monochromatic with respect to c for all but finitely many $\tau \in \mathcal{S}$. Indeed, let $\{\ell_0 < \ell_1 < \cdots\}$ be the increasing enumeration of the level set of T. By Lemma 5.10, the set $\mathcal{T}_{\ell_0} = \{\tau \in \mathcal{T} : |\tau| \leqslant \ell_0\}$ is finite. Let $\tau \in \mathcal{S} \setminus \mathcal{T}_{\ell_0}$ be arbitrary and set $m = |\tau|$. Notice that $m \geqslant 1$ and so $\tau = \tau(F)$ for some $F \subseteq A^{<\mathbb{N}}$ with $|F| \geqslant 2$. We select $a \in A$ and we set $F_0 = [\tau]$ and $F_1 = a^m{}^\frown[\tau]$ where $[\tau]$ is as in (5.20). By Fact 5.8, we have $\tau(F_0) = \tau(F_1) = \tau$. Also notice that $\wedge F_0 = \emptyset$ and $\wedge F_1 = a^m$. Now let $T_0 = T \upharpoonright 2m+1$ and set $G_0 = \mathrm{I}_{T_0}(F_0)$ and $G_1 = \mathrm{I}_{T_0}(F_1)$. By Lemma 5.7, we see that $\tau(G_0) = \tau(G_1) = \tau$. Moreover, $|\wedge G_0| = |\mathrm{I}_{T_0}(\wedge F_0)| = \ell_0 < |\tau|$ and $|\wedge G_2| = |\mathrm{I}_{T_0}(\wedge F_1)| = \ell_m \geqslant m = |\tau|$ which implies that $c(G_0) = 0$ and $c(G_1) = 1$. Hence, the set $\mathcal{P}_\tau(T)$ is not monochromatic.

We proceed to the proof of Theorem 5.13.

PROOF OF THEOREM 5.13. If τ is the empty word, then the result follows from (and is, in fact, equivalent to) the Carlson–Simpson theorem. So assume that τ is nonempty and set $m = |\tau|$. Let $c\colon \mathcal{P}(T) \to [r]$ be a finite coloring and define $C\colon \mathrm{SubCS}_m(T) \to [r]$ by the rule $C(R) = c\big([R,\tau]\big)$. By Theorem 4.23, there exists an infinite-dimensional Carlson–Simpson subspace S of T such that the set $\mathrm{SubCS}_m(S)$ is monochromatic with respect to the coloring C. Let F, F' be two subsets of S with $\tau(F) = \tau(F') = \tau$. By Lemma 5.9, there exist $U, U' \in \mathrm{SubCS}_m(S)$ such that $[U,\tau] = F$ and $[U',\tau] = F'$. Therefore, $c(F) = C(U) = C(U') = c(F')$ and the proof of Theorem 5.13 is completed. $\square$

5.3. Notes and remarks

The material in Section 5.1 is due to Dodos, Kanellopoulos and Tyros and is taken from [**DKT4**]. The analysis in Section 5.2 is new. Closely related results have been also obtained by Furstenberg and Katznelson in [**FK3**].

Part 2

Density theory

CHAPTER 6

Szemerédi's regularity method

In this chapter we will discuss certain aspects of *Szemerédi's regularity method*, a remarkable discovery of Szemerédi [**Sz2**] asserting that dense sets of discrete structures are inherently pseudorandom. The method was first developed in the context of graphs (see, e.g., [**KS**]), but it was realized recently that it can be formulated as an abstract probabilistic principle. This abstraction yields streamlined proofs and, more important, broadly extends the scope of applications of the method. In our exposition we will follow this probabilistic approach.

6.1. Decompositions of random variables

6.1.1. Semirings and their uniformity norms. We are about to present a decomposition of a given random variable into significantly simpler (and, consequently, more manageable) components. To this end, we will need the following slight strengthening of the classical concept of a semiring of sets (see also [**BN**]).

Definition 6.1. *Let Ω be a nonempty set and k a positive integer. Also let $\mathcal{S}$ be a collection of subsets of Ω. We say that $\mathcal{S}$ is a* k-semiring on Ω *if the following properties are satisfied.*

(P1) *We have that $\emptyset, \Omega \in \mathcal{S}$.*
(P2) *For every $S, T \in \mathcal{S}$ we have that $S \cap T \in \mathcal{S}$.*
(P3) *For every $S, T \in \mathcal{S}$ there exist $\ell \in [k]$ and $R_1, \dots, R_\ell \in \mathcal{S}$ which are pairwise disjoint and such that $S \setminus T = R_1 \cup \dots \cup R_\ell$.*

We view every element of a k-semiring $\mathcal{S}$ as a "structured" set and a linear combination of few characteristic functions of elements of $\mathcal{S}$ as a "simple" function. We will use the following norm in order to quantify how far from being "simple" a given function is.

Definition 6.2. *Let (Ω, Σ, μ) be a probability space, k a positive integer and $\mathcal{S}$ a k-semiring on Ω with $\mathcal{S} \subseteq \Sigma$. For every $f \in L_1(\Omega, \Sigma, \mu)$ we set*

$$\|f\|_{\mathcal{S}} = \sup\Big\{ \Big| \int_S f \, d\mu \Big| : S \in \mathcal{S} \Big\}. \tag{6.1}$$

The quantity $\|f\|_{\mathcal{S}}$ will be called the $\mathcal{S}$-uniformity norm *of f.*

Note that, in general, the $\mathcal{S}$-uniformity norm is a seminorm. However, observe that if the k-semiring $\mathcal{S}$ is sufficiently rich, then the function $\|\cdot\|_{\mathcal{S}}$ is indeed a norm. Specifically, the function $\|\cdot\|_{\mathcal{S}}$ is a norm if and only if the family $\{\mathbf{1}_S : S \in \mathcal{S}\}$

separates points in $L_1(\Omega,\Sigma,\mu)$, that is, for every $f,g\in L_1(\Omega,\Sigma,\mu)$ with $f\neq g$ there exists $S\in\mathcal{S}$ with $\int_S f\,d\mu\neq\int_S g\,d\mu$.

The simplest example of a k-semiring on a nonempty set Ω, is an algebra of subsets of Ω. Indeed, notice that a family of subsets of Ω is a 1-semiring if and only if it is an algebra. Another standard example is the collection of all intervals[1] of a linearly ordered set, a family which is easily seen to be a 2-semiring. The following lemma will enable us to construct a variety of k-semirings.

Lemma 6.3. *Let Ω be a nonempty set. Also let $m,k_1,\dots,k_m$ be positive integers and set $k=\sum_{i=1}^m k_i$. If $\mathcal{S}_i$ is a k_i-semiring on Ω for every $i\in[m]$, then the family*

$$\mathcal{S}=\Big\{\bigcap_{i=1}^m S_i : S_i\in\mathcal{S}_i \text{ for every } i\in[m]\Big\} \tag{6.2}$$

is a k-semiring on Ω.

Proof. We may assume, of course, that $m\geqslant 2$. Notice that the family $\mathcal{S}$ satisfies properties (P1) and (P2) in Definition 6.1. To see that property (P3) is also satisfied, fix $S,T\in\mathcal{S}$ and write $S=\bigcap_{i=1}^m S_i$ and $T=\bigcap_{i=1}^m T_i$ where $S_i,T_i\in\mathcal{S}_i$ for every $i\in[m]$. We set $P_1=\Omega\setminus T_1$ and $P_j=T_1\cap\dots\cap T_{j-1}\cap(\Omega\setminus T_j)$ if $j\in\{2,\dots,m\}$. Observe that the sets $P_1,\dots,P_m$ are pairwise disjoint. Moreover,

$$\Omega\setminus\Big(\bigcap_{i=1}^m T_i\Big)=\bigcup_{j=1}^m P_j$$

and so

$$S\setminus T=\Big(\bigcap_{i=1}^m S_i\Big)\setminus\Big(\bigcap_{i=1}^m T_i\Big)=\bigcup_{j=1}^m\Big(\bigcap_{i=1}^m S_i\cap P_j\Big).$$

Let $j\in[m]$ be arbitrary. Since $\mathcal{S}_j$ is a k_j-semiring, there exist $\ell_j\in[k_j]$ and pairwise disjoint sets $R_1^j,\dots,R_{\ell_j}^j\in\mathcal{S}_j$ such that $S_j\setminus T_j=R_1^j\cup\dots\cup R_{\ell_j}^j$. Thus, setting

(a) $B_1=\Omega$ and $B_j=\bigcap_{1\leqslant i<j}(S_i\cap T_i)$ if $j\in\{2,\dots,m\}$,

(b) $C_j=\bigcap_{j<i\leqslant m}S_i$ if $j\in\{1,\dots,m-1\}$ and $C_m=\Omega$,

and invoking the definition of the sets $P_1,\dots,P_m$ we obtain that

$$S\setminus T=\bigcup_{j=1}^m\Big(\bigcup_{n=1}^{\ell_j}\big(B_j\cap R_n^j\cap C_j\big)\Big). \tag{6.3}$$

Now set $I=\bigcup_{j=1}^m\big(\{j\}\times[\ell_j]\big)$ and observe that $|I|\leqslant k$. For every $(j,n)\in I$ let $U_n^j=B_j\cap R_n^j\cap C_j$ and notice that $U_n^j\in\mathcal{S}$, $U_n^j\subseteq R_n^j$ and $U_n^j\subseteq P_j$. This implies, in particular, that the family $\{U_n^j:(j,n)\in I\}$ is contained in $\mathcal{S}$ and consists of pairwise disjoint sets. Moreover, by (6.3), we have

$$S\setminus T=\bigcup_{(j,n)\in I}U_n^j.$$

Hence, the family $\mathcal{S}$ satisfies property (P3) in Definition 6.1 and the proof of Lemma 6.3 is completed. $\square$

[1] Recall that a subset I of a linearly ordered set $(L,<)$ is said to be an *interval* if for every $x,y\in I$ and every $z\in L$ with $x<z<y$ we have $z\in I$.

Next we isolate some basic properties of the $\mathcal{S}$-uniformity norm. Recall that, if (Ω,Σ,μ) is a probability space and Σ' is a sub-σ-algebra of Σ, then for every random variable $f\in L_1(\Omega,\Sigma,\mu)$ by $\mathbb{E}(f\,|\,\Sigma')$ we denote the conditional expectation of f relative to Σ'.

LEMMA 6.4. *Let (Ω,Σ,μ) be a probability space, k a positive integer and $\mathcal{S}$ a k-semiring on Ω with $\mathcal{S}\subseteq\Sigma$. Also let $f\in L_1(\Omega,\Sigma,\mu)$. Then the following hold.*

(a) *We have $\|f\|_{\mathcal{S}}\leqslant\|f\|_{L_1}$.*
(b) *If Σ' is a σ-algebra on Ω with $\Sigma'\subseteq\mathcal{S}$, then $\|\mathbb{E}(f\,|\,\Sigma')\|_{\mathcal{S}}\leqslant\|f\|_{\mathcal{S}}$.*
(c) *If $\mathcal{S}$ is a σ-algebra, then $\|f\|_{\mathcal{S}}\leqslant\|\mathbb{E}(f\,|\,\mathcal{S})\|_{L_1}\leqslant 2\|f\|_{\mathcal{S}}$.*

PROOF. Part (a) is straightforward. For part (b), fix a σ-algebra Σ' on Ω with $\Sigma'\subseteq\mathcal{S}$, and set $P=\{\omega\in\Omega:\mathbb{E}(f\,|\,\Sigma')(\omega)\geqslant 0\}$ and $N=\Omega\setminus P$. Notice that $P,N\in\Sigma'\subseteq\mathcal{S}$. Hence, for every $S\in\mathcal{S}$ we have

$$\begin{aligned}\Big|\int_S\mathbb{E}(f\,|\,\Sigma')\,d\mu\Big| &\leqslant \max\Big\{\int_{P\cap S}\mathbb{E}(f\,|\,\Sigma')\,d\mu,-\int_{N\cap S}\mathbb{E}(f\,|\,\Sigma')\,d\mu\Big\}\\ &\leqslant \max\Big\{\int_{P}\mathbb{E}(f\,|\,\Sigma')\,d\mu,-\int_{N}\mathbb{E}(f\,|\,\Sigma')\,d\mu\Big\}\\ &= \max\Big\{\int_P f\,d\mu,-\int_N f\,d\mu\Big\}\leqslant\|f\|_{\mathcal{S}}\end{aligned}$$

which yields that $\|\mathbb{E}(f\,|\,\Sigma')\|_{\mathcal{S}}\leqslant\|f\|_{\mathcal{S}}$.

Finally, assume that $\mathcal{S}$ is a σ-algebra and observe that $\int_S f\,d\mu=\int_S\mathbb{E}(f\,|\,\mathcal{S})\,d\mu$ for every $S\in\mathcal{S}$. In particular, we have $\|f\|_{\mathcal{S}}\leqslant\|\mathbb{E}(f\,|\,\mathcal{S})\|_{L_1}$. Also let, as above, $P=\{\omega\in\Omega:\mathbb{E}(f\,|\,\mathcal{S})(\omega)\geqslant 0\}$ and $N=\Omega\setminus P$. Since $P,N\in\mathcal{S}$, we obtain that

$$\|\mathbb{E}(f\,|\,\mathcal{S})\|_{L_1}\leqslant 2\cdot\max\Big\{\int_P\mathbb{E}(f\,|\,\mathcal{S})\,d\mu,-\int_N\mathbb{E}(f\,|\,\mathcal{S})\,d\mu\Big\}\leqslant 2\|f\|_{\mathcal{S}}$$

and the proof of Lemma 6.4 is completed. □

We close this subsection by presenting some examples of k-semirings which are relevant from a combinatorial perspective. In the first example the underlying space is the product of a finite sequence of probability spaces. The corresponding k-semirings are closely related to the development of Szemerédi's regularity method for hypergraphs and will be of particular importance in Chapter 7.

EXAMPLE 6.1. Let n be a positive integer and $(\Omega_1,\Sigma_1,\mu_1),\dots,(\Omega_n,\Sigma_n,\mu_n)$ a finite sequence of probability spaces. By $(\boldsymbol{\Omega},\boldsymbol{\Sigma},\boldsymbol{\mu})$ we shall denote their product (see Appendix E). Moreover, if $I\subseteq[n]$ is nonempty, then the product of the spaces $\langle(\Omega_i,\Sigma_i,\mu_i):i\in I\rangle$ will be denoted by $(\boldsymbol{\Omega}_I,\boldsymbol{\Sigma}_I,\boldsymbol{\mu}_I)$. In particular, we have

$$\boldsymbol{\Omega}=\prod_{i=1}^n\Omega_i\ \ \text{and}\ \ \boldsymbol{\Omega}_I=\prod_{i\in I}\Omega_i.$$

(By convention, $\boldsymbol{\Omega}_\emptyset$ stands for the empty set.) Notice that the σ-algebra $\boldsymbol{\Sigma}_I$ is not comparable with $\boldsymbol{\Sigma}$, but it may be "lifted" to the full product $\boldsymbol{\Omega}$ using the natural projection $\pi_I\colon\boldsymbol{\Omega}\to\boldsymbol{\Omega}_I$. Specifically, let

$$\mathcal{B}_I=\big\{\pi_I^{-1}(\mathbf{A}):\mathbf{A}\in\boldsymbol{\Sigma}_I\big\}\tag{6.4}$$

and observe that $\mathcal{B}_I$ is a sub-σ-algebra of $\boldsymbol{\Sigma}$.

Now assume that $n \geqslant 2$ and let $\mathcal{I}$ be a family of nonempty subsets of $[n]$. Set $k = |\mathcal{I}|$ and observe that, by Lemma 6.3, we may associate with the family $\mathcal{I}$ a k-semiring $\mathcal{S}_{\mathcal{I}}$ on $\boldsymbol{\Omega}$ defined by the rule

$$S \in \mathcal{S}_{\mathcal{I}} \Leftrightarrow S = \bigcap_{I \in \mathcal{I}} A_I \text{ where } A_I \in \mathcal{B}_I \text{ for every } I \in \mathcal{I}. \tag{6.5}$$

Note that if the family $\mathcal{I}$ satisfies $[n] \notin \mathcal{I}$ and $\cup\mathcal{I} = [n]$, then it gives rise to a non-trivial semiring whose corresponding uniformity norm is a genuine norm.

It turns out that there is a minimal non-trivial semiring $\mathcal{S}_{\min}$ one can obtain in this way. It corresponds to the family $\mathcal{I}_{\min} = \binom{[n]}{1}$ and is particularly easy to grasp since it consists of all measurable rectangles of $\boldsymbol{\Omega}$. The $\mathcal{S}_{\min}$-uniformity norm is known as the *cut norm* and was introduced by Frieze and Kannan [**FrK**].

At the other extreme, this construction also yields a maximal non-trivial semiring $\mathcal{S}_{\max}$ on $\boldsymbol{\Omega}$. It corresponds to the family $\mathcal{I}_{\max} = \binom{[n]}{n-1}$ and consists of those subsets of the product which can be written as $A_1 \cap \cdots \cap A_n$ where for every $i \in [n]$ the set A_i is measurable and does not depend on the i-th coordinate. The $\mathcal{S}_{\max}$-uniformity norm is known as the *Gowers box norm* and was introduced by Gowers [**Go4, Go5**].

In the second example the underlying space is a combinatorial space of $A^{<\mathbb{N}}$ where A is a finite alphabet with at least two letters. The building blocks of the corresponding k-semirings are the insensitive sets introduced in Subsection 2.1.1.

Example 6.2. Let A be a finite alphabet with $|A| \geqslant 2$. Also let W be a combinatorial space of $A^{<\mathbb{N}}$. We view W as a discrete probability space equipped with the uniform probability measure. For every $a, b \in A$ with $a \neq b$ we set

$$\mathcal{A}_{\{a,b\}} = \{X \subseteq W : X \text{ is } (a,b)\text{-insensitive in } W\}. \tag{6.6}$$

We have already pointed out in Subsection 2.1.1 that the family $\mathcal{A}_{\{a,b\}}$ is an algebra of subsets of W. These algebras can then be used to construct various semirings on W. Specifically, let $\mathcal{I} \subseteq \binom{A}{2}$ and set $k = |\mathcal{I}|$. By Lemma 6.3, we see that the family constructed from the algebras $\{\mathcal{A}_{\{a,b\}} : \{a,b\} \in \mathcal{I}\}$ via formula (6.2) is a k-semiring on W. The maximal semiring obtained in this way corresponds to the family $\binom{A}{2}$. We shall denote it by $\mathcal{S}(W)$. Notice, in particular, that $\mathcal{S}(W)$ is a K-semiring on W where $K = |A|(|A|-1)2^{-1}$. Also observe that if $|A| \geqslant 3$, then the $\mathcal{S}(W)$-uniformity norm is actually a norm.

6.1.2. The main result. First we introduce some terminology and some pieces of notation. We say that a function $F\colon \mathbb{N} \to \mathbb{R}$ is a *growth function* provided that: (i) F is increasing, and (ii) $F(n) \geqslant n+1$ for every $n \in \mathbb{N}$. Moreover, as in Appendix E, for every nonempty set Ω and every finite partition $\mathcal{P}$ of Ω by $\mathcal{A}_{\mathcal{P}}$ we shall denote the finite algebra on Ω generated by $\mathcal{P}$. Recall that the nonempty atoms of $\mathcal{A}_{\mathcal{P}}$ are precisely the members of $\mathcal{P}$, and notice that a finite partition $\mathcal{Q}$ of Ω is a refinement[2] of $\mathcal{P}$ if and only if $\mathcal{A}_{\mathcal{Q}} \supseteq \mathcal{A}_{\mathcal{P}}$.

[2]If $\mathcal{Q}$ and $\mathcal{P}$ are two finite partitions of a nonempty set Ω, then recall that $\mathcal{Q}$ is said to be a *refinement* of $\mathcal{P}$ if for every $Q \in \mathcal{Q}$ there exists $P \in \mathcal{P}$ with $Q \subseteq P$.

Now for every pair k, ℓ of positive integers, every $0 < \sigma \leqslant 1$ and every growth function $F\colon \mathbb{N} \to \mathbb{R}$ we define $h\colon \mathbb{N} \to \mathbb{N}$ recursively by rule

$$\begin{cases} h(0) = 1 \\ h(i+1) = h(i) \cdot (k+1)^{\lceil \sigma^2 F(h(i))^2 \ell \rceil} \end{cases} \tag{6.7}$$

and we set

$$\mathrm{RegSz}(k, \ell, \sigma, F) = h\big(\lceil \sigma^{-2} \ell \rceil\big). \tag{6.8}$$

Observe that if σ is rational and $F\colon \mathbb{N} \to \mathbb{N}$ is a primitive recursive growth function belonging to the class $\mathcal{E}^n$ for some $n \in \mathbb{N}$, then the function h is also primitive recursive and belongs to the class $\mathcal{E}^m$ where $m = \max\{4, n+1\}$.

The following theorem is the main result of this section and is essentially due to Tao [**Tao1, Tao2**]. General facts about the conditional expectation relative to a σ-algebra can be found in Appendix E.

THEOREM 6.5. *Let k, ℓ be positive integers, $0 < \sigma \leqslant 1$ and $F\colon \mathbb{N} \to \mathbb{R}$ a growth function. Also let (Ω, Σ, μ) be a probability space and $\mathcal{S}$ a k-semiring on Ω with $\mathcal{S} \subseteq \Sigma$. Finally, let $\mathcal{F}$ be a family in $L_2(\Omega, \Sigma, \mu)$ such that $\|f\|_{L_2} \leqslant 1$ for every $f \in \mathcal{F}$ and with $|\mathcal{F}| = \ell$. Then there exist*

(i) *a positive integer M with $M \leqslant \mathrm{RegSz}(k, \ell, \sigma, F)$,*
(ii) *a partition $\mathcal{P}$ of Ω with $\mathcal{P} \subseteq \mathcal{S}$ and $|\mathcal{P}| = M$, and*
(iii) *a finite refinement $\mathcal{Q}$ of $\mathcal{P}$ with $\mathcal{Q} \subseteq \mathcal{S}$*

such that for every $f \in \mathcal{F}$, writing $f = f_{\mathrm{str}} + f_{\mathrm{err}} + f_{\mathrm{unf}}$ where

$$f_{\mathrm{str}} = \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{P}}), \quad f_{\mathrm{err}} = \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{Q}}) - \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{P}}) \quad \textit{and} \quad f_{\mathrm{unf}} = f - \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{Q}}), \tag{6.9}$$

we have the estimates

$$\|f_{\mathrm{err}}\|_{L_2} \leqslant \sigma \quad \textit{and} \quad \|f_{\mathrm{unf}}\|_{\mathcal{S}} \leqslant \frac{1}{F(M)}. \tag{6.10}$$

In particular, for every $f \in \mathcal{F}$ the following hold.

(a) *The function f_{str} is constant on each $S \in \mathcal{P}$.*
(b) *The functions f_{str} and $f_{\mathrm{str}} + f_{\mathrm{err}}$ are non-negative if f is non-negative. If, in addition, f is $[0,1]$-valued, then f_{err} and f_{unf} take values in $[-1,1]$ while f_{str} and $f_{\mathrm{str}} + f_{\mathrm{err}}$ take values in $[0,1]$.*
(c) *If Σ' is a sub-σ-algebra of Σ with $\mathcal{S} \subseteq \Sigma'$ and f is Σ'-measurable, then the functions $f_{\mathrm{str}}, f_{\mathrm{err}}$ and f_{unf} are also Σ'-measurable.*

The proof of Theorem 6.5 will be given in the next subsection. We also note that in Section 7.1 we will present a multidimensional version of Theorem 6.5 which will enable us to decompose, simultaneously, a finite family of random variables with respect to an arbitrary finite collection of k-semirings. This additional feature is needed in the context of hypergraphs and related combinatorial structures. However, in all applications in this chapter there will be only one relevant k-semiring.

6.1.3. Proof of Theorem 6.5. First we need to do some preparatory work. Recall that a finite sequence $(f_i)_{i=0}^n$ of integrable random variables on a probability space (Ω,Σ,μ) is said to be a *martingale* if there exists an increasing sequence $(\mathcal{A}_i)_{i=0}^n$ of sub-σ-algebras of Σ such that: (i) $f_i\in L_1(\Omega,\mathcal{A}_i,\mu)$ for every $i\in\{0,\dots,n\}$, and (ii) $f_i=\mathbb{E}(f_{i+1}\,|\,\mathcal{A}_i)$ if $n\geqslant 1$ and $i\in\{0,\dots,n-1\}$. Note that this is equivalent to saying that there exists $f\in L_1(\Omega,\Sigma,\mu)$ such that

$$f_i=\mathbb{E}(f\,|\,\mathcal{A}_i) \tag{6.11}$$

for every $i\in\{0,\dots,n\}$. We have the following basic property of successive differences[3] of square-integrable finite martingales.

Fact 6.6. *Let (Ω,Σ,μ) be a probability space. Also let n be a positive integer and $(\mathcal{A}_i)_{i=0}^n$ an increasing finite sequence of sub-σ-algebras of Σ. Then for every $f\in L_2(\Omega,\Sigma,\mu)$ we have*

$$\Big(\sum_{i=1}^n \|\mathbb{E}(f\,|\,\mathcal{A}_i)-\mathbb{E}(f\,|\,\mathcal{A}_{i-1})\|_{L_2}^2\Big)^{1/2}\leqslant \|\mathbb{E}(f\,|\,\mathcal{A}_n)\|_{L_2}. \tag{6.12}$$

Proof. We set $d_0=\mathbb{E}(f\,|\,\mathcal{A}_0)$ and $d_i=\mathbb{E}(f\,|\,\mathcal{A}_i)-\mathbb{E}(f\,|\,\mathcal{A}_{i-1})$ if $i\in[n]$. By Proposition E.1, the sequence $(d_i)_{i=0}^n$ is orthogonal in $L_2(\Omega,\Sigma,\mu)$. Therefore,

$$\begin{aligned}\Big(\sum_{i=1}^n \|\mathbb{E}(f\,|\,\mathcal{A}_i)-\mathbb{E}(f\,|\,\mathcal{A}_{i-1})\|_{L_2}^2\Big)^{1/2} &\leqslant \Big(\sum_{i=0}^n\|d_i\|_{L_2}^2\Big)^{1/2}\\ &= \Big\|\sum_{i=0}^n d_i\Big\|_{L_2}=\|\mathbb{E}(f\,|\,\mathcal{A}_n)\|_{L_2}\end{aligned}$$

and the proof of Fact 6.6 is completed. □

The following lemma is the first main step of the proof of Theorem 6.5.

Lemma 6.7. *Let k be a positive integer and $0<\delta\leqslant 1$. Also let (Ω,Σ,μ) be a probability space, $\mathcal{S}$ a k-semiring on Ω with $\mathcal{S}\subseteq\Sigma$ and $\mathcal{Q}$ a finite partition of Ω with $\mathcal{Q}\subseteq\mathcal{S}$. Finally, let $f\in L_2(\Omega,\Sigma,\mu)$ be such that $\|f-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})\|_{\mathcal{S}}>\delta$. Then there exists a refinement $\mathcal{R}$ of $\mathcal{Q}$ such that:* (i) $\mathcal{R}\subseteq\mathcal{S}$, (ii) $|\mathcal{R}|\leqslant|\mathcal{Q}|(k+1)$, *and* (iii) $\|\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{R}})-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})\|_{L_2}>\delta$.

Proof. By our assumptions, there exists $S\in\mathcal{S}$ such that

$$\Big|\int_S\big(f-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})\big)\,d\mu\Big|>\delta. \tag{6.13}$$

Since $\mathcal{S}$ is a k-semiring on Ω, there exists a refinement $\mathcal{R}$ of $\mathcal{Q}$ such that: (i) $\mathcal{R}\subseteq\mathcal{S}$, (ii) $|\mathcal{R}|\leqslant|\mathcal{Q}|(k+1)$, and (iii) $S\in\mathcal{A}_{\mathcal{R}}$. In particular, it follows that

$$\int_S\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{R}})\,d\mu=\int_S f\,d\mu. \tag{6.14}$$

[3]Successive differences of martingales are known as *martingale difference sequences*.

Therefore, by the monotonicity of the L_p norms, we obtain that

$$\begin{aligned}\delta \overset{(6.13)}{<} \Big|\int_S \big(f-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})\big)\,d\mu\Big| \overset{(6.14)}{=} & \Big|\int_S \big(\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{R}})-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})\big)\,d\mu\Big| \\ \leqslant & \ \|\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{R}})-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})\|_{L_1} \\ \leqslant & \ \|\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{R}})-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})\|_{L_2}\end{aligned}$$

and the proof of Lemma 6.7 is completed. □

Let k be a positive integer, Ω a nonempty set and $\mathcal{S}$ a k-semiring on Ω. For every $N\in\mathbb{N}$ and every finite partition $\mathcal{P}$ of Ω with $\mathcal{P}\subseteq\mathcal{S}$ by $\Pi^N_{\mathcal{S}}(\mathcal{P})$ we shall denote the set of all refinements $\mathcal{Q}$ of $\mathcal{P}$ which satisfy $\mathcal{Q}\subseteq\mathcal{S}$ and $|\mathcal{Q}|\leqslant|\mathcal{P}|(k+1)^N$. We proceed with the following lemma.

Lemma 6.8. *Let k,ℓ be positive integers, $0<\delta,\sigma\leqslant 1$ and set $N=\lceil\sigma^2\delta^{-2}\ell\rceil$. Also let (Ω,Σ,μ) be a probability space, $\mathcal{S}$ a k-semiring on Ω with $\mathcal{S}\subseteq\Sigma$, $\mathcal{P}$ a finite partition of Ω with $\mathcal{P}\subseteq\mathcal{S}$ and $\mathcal{F}$ a family in $L_2(\Omega,\Sigma,\mu)$ with $|\mathcal{F}|=\ell$. Then there exists $\mathcal{Q}\in\Pi^N_{\mathcal{S}}(\mathcal{P})$ such that either*

(a) *$\|\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{P}})\|_{L_2}>\sigma$ for some $f\in\mathcal{F}$, or*

(b) *$\|\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{P}})\|_{L_2}\leqslant\sigma$ and $\|f-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})\|_{\mathcal{S}}\leqslant\delta$ for every $f\in\mathcal{F}$.*

Proof. Assume that there is no $\mathcal{Q}\in\Pi^N_{\mathcal{S}}(\mathcal{P})$ which satisfies the first part of the lemma. Observe that this is equivalent to saying that

(H1) $\|\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{P}})\|_{L_2}\leqslant\sigma$ for every $\mathcal{Q}\in\Pi^N_{\mathcal{S}}(\mathcal{P})$ and every $f\in\mathcal{F}$.

We will use hypothesis (H1) to show that there exists $\mathcal{Q}\in\Pi^N_{\mathcal{S}}(\mathcal{P})$ which satisfies the second part of the lemma.

To this end we will argue by contradiction. Let $\mathcal{Q}\in\Pi^N_{\mathcal{S}}(\mathcal{P})$ be arbitrary. By hypothesis (H1) and our assumption that part (b) does not hold true, there exists $f\in\mathcal{F}$ (possibly depending on the partition $\mathcal{Q}$) such that $\|f-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})\|_{\mathcal{S}}>\delta$. Hence, by Lemma 6.7, we obtain that

(H2) for every $\mathcal{Q}\in\Pi^N_{\mathcal{S}}(\mathcal{P})$ there exist $\mathcal{R}\in\Pi^1_{\mathcal{S}}(\mathcal{Q})$ and $f\in\mathcal{F}$ such that $\|\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{R}})-\mathbb{E}(f\,|\,\mathcal{A}_{\mathcal{Q}})\|_{L_2}>\delta$.

Recursively and using hypothesis (H2), we select a finite sequence $\mathcal{P}_0,\dots,\mathcal{P}_N$ of finite partitions of Ω with $\mathcal{P}_0=\mathcal{P}$ and a finite sequence $f_1,\dots,f_N$ in $\mathcal{F}$ such that for every $i\in[N]$ we have $\mathcal{P}_i\in\Pi^1_{\mathcal{S}}(\mathcal{P}_{i-1})$ and $\|\mathbb{E}(f_i\,|\,\mathcal{A}_{\mathcal{P}_i})-\mathbb{E}(f_i\,|\,\mathcal{A}_{\mathcal{P}_{i-1}})\|_{L_2}>\delta$. The first property implies that $\mathcal{P}_j\in\Pi^N_{\mathcal{S}}(\mathcal{P}_i)$ for every $i,j\in\{0,\dots,N\}$ with $i<j$. In particular, we have $\mathcal{P}_N\in\Pi^N_{\mathcal{S}}(\mathcal{P})$. On the other hand, by the fact that $|\mathcal{F}|=\ell$ and the classical pigeonhole principle, there exist $g\in\mathcal{F}$ and $I\subseteq[N]$ with $|I|\geqslant N/\ell$ and such that $g=f_i$ for every $i\in I$. Hence, $\|\mathbb{E}(g\,|\,\mathcal{A}_{\mathcal{P}_i})-\mathbb{E}(g\,|\,\mathcal{A}_{\mathcal{P}_{i-1}})\|_{L_2}>\delta$ for every $i\in I$. By Fact 6.6 applied to the random variable "$f=g-\mathbb{E}(g\,|\,\mathcal{A}_{\mathcal{P}})$" and the sequence $(\mathcal{A}_{\mathcal{P}_i})_{i=0}^N$, we obtain that

$$\sigma\leqslant\delta(N/\ell)^{1/2}\leqslant\delta|I|^{1/2}<\|\mathbb{E}(g\,|\,\mathcal{A}_{\mathcal{P}_N})-\mathbb{E}(g\,|\,\mathcal{A}_{\mathcal{P}})\|_{L_2}. \tag{6.15}$$

Summing up we see that $\mathcal{P}_N\in\Pi^N_{\mathcal{S}}(\mathcal{P})$ and $\|\mathbb{E}(g\,|\,\mathcal{A}_{\mathcal{P}_N})-\mathbb{E}(g\,|\,\mathcal{A}_{\mathcal{P}})\|_{L_2}>\sigma$ which contradicts hypothesis (H1). The proof of Lemma 6.8 is thus completed. □

The last step of the proof of Theorem 6.5 is the content of the following lemma.

LEMMA 6.9. *Let k, ℓ be positive integers, $0 < \sigma \leqslant 1$ and $F\colon \mathbb{N} \to \mathbb{R}$ a growth function. Set $L = \lceil \sigma^{-2}\ell \rceil$ and define two sequences (N_i) and (M_i) in $\mathbb{N}$ recursively by the rule*

$$\begin{cases} N_0 = 0 \text{ and } M_0 = 1, \\ N_{i+1} = \lceil \sigma^2 F(M_i)^2 \ell \rceil \text{ and } M_{i+1} = M_i (k+1)^{N_{i+1}}. \end{cases} \tag{6.16}$$

Let (Ω, Σ, μ) be a probability space, $\mathcal{S}$ a k-semiring on Ω with $\mathcal{S} \subseteq \Sigma$ and $\mathcal{F}$ a family in $L_2(\Omega, \Sigma, \mu)$ such that $\|f\|_{L_2} \leqslant 1$ for every $f \in \mathcal{F}$ and with $|F| = \ell$. Then there exist $i \in \{0, \dots, L-1\}$ and two finite partitions $\mathcal{P}$ and $\mathcal{Q}$ of Ω with: (i) *$\mathcal{P} \subseteq \mathcal{S}$,* (ii) *$|\mathcal{P}| \leqslant M_i$,* (iii) *$\mathcal{Q} \in \Pi_{\mathcal{S}}^{N_{i+1}}(\mathcal{P})$, and* (iv) *$\|\mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{Q}}) - \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{P}})\|_{L_2} \leqslant \sigma$ and $\|f - \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{Q}})\|_{\mathcal{S}} \leqslant 1/F(M_i)$ for every $f \in \mathcal{F}$.*

PROOF. It is similar to the proof of Lemma 6.8. Let $i \in \{0, \dots, L-1\}$ and let $\mathcal{P}$ be a finite partition of Ω with $\mathcal{P} \subseteq \mathcal{S}$ and $|\mathcal{P}| \leqslant M_i$. By Lemma 6.8, we see that one of the following alternatives is satisfied.

(A1) There exists $\mathcal{Q} \in \Pi_{\mathcal{S}}^{N_{i+1}}(\mathcal{P})$ such that $\|\mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{Q}}) - \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{P}})\|_{L_2} \leqslant \sigma$ and $\|f - \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{Q}})\|_{\mathcal{S}} \leqslant 1/F(M_i)$ for every $f \in \mathcal{F}$.

(A2) There exist $\mathcal{Q} \in \Pi_{\mathcal{S}}^{N_{i+1}}(\mathcal{P})$ and $f \in \mathcal{F}$ with $\|\mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{Q}}) - \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{P}})\|_{L_2} > \sigma$.

Of course, the proof of the lemma will be completed once we show that the first alternative holds true for some $i \in \{0, \dots, L-1\}$ and some finite partition $\mathcal{P}$ as described above.

Assume, towards a contradiction, that such a pair cannot be found. Recursively and invoking alternative (A2), we select a finite sequence $\mathcal{P}_0, \dots, \mathcal{P}_L$ of finite partitions of Ω with $\mathcal{P}_0 = \{\Omega\}$ and a finite sequence $f_1, \dots, f_L$ in $\mathcal{F}$ such that for every $i \in [L]$ we have $\mathcal{P}_i \in \Pi_{\mathcal{S}}^{N_{i+1}}(\mathcal{P}_{i-1})$ and $\|\mathbb{E}(f_i \,|\, \mathcal{A}_{\mathcal{P}_i}) - \mathbb{E}(f_i \,|\, \mathcal{A}_{\mathcal{P}_{i-1}})\|_{L_2} > \sigma$. By the classical pigeonhole principle, there exist $g \in \mathcal{F}$ and $I \subseteq [L]$ with $|I| \geqslant L/\ell \geqslant \sigma^{-2}$ and such that $g = f_i$ for every $i \in I$. By Fact 6.6, the previous discussion and the fact that $\|g\|_{L_2} \leqslant 1$, we conclude that

$$1 \leqslant \sigma |I|^{1/2} < \|\mathbb{E}(g \,|\, \mathcal{A}_{\mathcal{P}_L})\|_{L_2} \leqslant \|g\|_{L_2} \leqslant 1$$

which is clearly a contradiction. The proof of Lemma 6.9 is completed. □

We are now ready to complete the proof of Theorem 6.5.

PROOF OF THEOREM 6.5. Fix the positive integers k and ℓ, the constant σ and the growth function F. Set $L = \lceil \sigma^{-2}\ell \rceil$. Also let $i \in \{0, \dots, L-1\}$ and $\mathcal{P}, \mathcal{Q}$ be as in Lemma 6.9, and set $M = |\mathcal{P}|$. We will show that the positive integer M and the finite partitions $\mathcal{P}$ and $\mathcal{Q}$ are as desired.

To this end, let h be the function defined in (6.7) for the fixed data k, ℓ, σ and F. By (6.16), we have $M_i = h(i)$ for every $i \in \mathbb{N}$ and so

$$M = |\mathcal{P}| \leqslant M_i \leqslant M_{\lceil \sigma^{-2}\ell \rceil} = h(\lceil \sigma^{-2}\ell \rceil) \overset{(6.8)}{=} \mathrm{RegSz}(k, \ell, \sigma, F). \tag{6.17}$$

Since $\mathcal{P}, \mathcal{Q} \subseteq \mathcal{S}$ and $\mathcal{Q}$ is a finite refinement of $\mathcal{P}$, we see that $M, \mathcal{P}$ and $\mathcal{Q}$ satisfy the requirements of the theorem. Next, let $f \in \mathcal{F}$ be arbitrary and set

$$f_{\mathrm{str}} = \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{P}}), \quad f_{\mathrm{err}} = \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{Q}}) - \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{P}}) \ \text{ and } \ f_{\mathrm{unf}} = f - \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{Q}}).$$

Clearly, it is enough to show that the random variables f_{err} and f_{unf} obey the estimates in (6.10). Indeed, by the choice of $\mathcal{P}$ and $\mathcal{Q}$ in Lemma 6.9, we have

$$\|f_{\mathrm{err}}\|_{L_2} = \|\mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{Q}}) - \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{P}})\|_{L_2} \leqslant \sigma. \tag{6.18}$$

Moreover, invoking Lemma 6.9 once again and using the fact that the function $F\colon \mathbb{N} \to \mathbb{R}$ is increasing, we conclude that

$$\|f_{\mathrm{unf}}\|_{\mathcal{S}} = \|f - \mathbb{E}(f \,|\, \mathcal{A}_{\mathcal{Q}})\|_{\mathcal{S}} \leqslant \frac{1}{F(M_i)} \leqslant \frac{1}{F(M)}. \tag{6.19}$$

The proof of Theorem 6.5 is completed. □

6.1.4. Uniform partitions. In this subsection we will obtain a consequence of Theorem 6.5 which is more akin to the graph-theoretic versions of Szemerédi's regularity method and is somewhat easier to use in a combinatorial setting. To this end, we give the following definition.

DEFINITION 6.10. *Let (Ω, Σ, μ) be a probability space, k a positive integer and $\mathcal{S}$ a k-semiring on Ω with $\mathcal{S} \subseteq \Sigma$. Also let $f \in L_1(\Omega, \Sigma, \mu)$, $0 < \eta \leqslant 1$ and $S \in \mathcal{S}$. We say that the set S is* $(f, \mathcal{S}, \eta)$-uniform *if for every $T \subseteq S$ with $T \in \mathcal{S}$ we have*

$$\Big| \int_T \big(f - \mathbb{E}(f \,|\, S)\big)\, d\mu \Big| \leqslant \eta \cdot \mu(S). \tag{6.20}$$

Moreover, for every $\mathcal{C} \subseteq \mathcal{S}$ we set $\mathrm{Unf}(\mathcal{C}, f, \eta) = \{C \in \mathcal{C} : C$ is $(f, \mathcal{S}, \eta)$-uniform$\}$.

Notice that if $S \in \mathcal{S}$ with $\mu(S) = 0$, then the set S is $(f, \mathcal{S}, \eta)$-uniform for every $0 < \eta \leqslant 1$. The same remark of course applies if the function f is constant on S. Also note that the concept of $(f, \mathcal{S}, \eta)$-uniformity is closely related to the $\mathcal{S}$-uniformity norm introduced in Definition 6.2. Indeed, let $S \in \mathcal{S}$ with $\mu(S) > 0$ and observe that the set S is $(f, \mathcal{S}, \eta)$-uniform if and only if the function $f - \mathbb{E}(f \,|\, S)$, viewed as a random variable in $L_1(\Omega, \Sigma, \mu_S)$, has $\mathcal{S}$-uniformity norm less than or equal to η. In particular, the set Ω is $(f, \mathcal{S}, \eta)$-uniform if and only if $\|f - \mathbb{E}(f)\|_{\mathcal{S}} \leqslant \eta$.

We have the following proposition (see [**TV**, Section 11.6]).

PROPOSITION 6.11. *For every pair k, ℓ of positive integers and every $0 < \eta \leqslant 1$ there exists a positive integer $\mathrm{U}(k, \ell, \eta)$ with the following property. Let (Ω, Σ, μ) be a probability space, $\mathcal{S}$ a k-semiring on Ω with $\mathcal{S} \subseteq \Sigma$ and $\mathcal{F}$ a family in $L_2(\Omega, \Sigma, \mu)$ such that $\|f\|_{L_2} \leqslant 1$ for every $f \in \mathcal{F}$ and with $|\mathcal{F}| = \ell$. Then there exist a positive integer $M \leqslant \mathrm{U}(k, \ell, \eta)$ and a partition $\mathcal{P}$ of Ω with $\mathcal{P} \subseteq \mathcal{S}$ and $|\mathcal{P}| = M$ such that*

$$\sum_{S \in \mathrm{Unf}(\mathcal{P}, f, \eta)} \mu(S) \geqslant 1 - \eta \tag{6.21}$$

for every $f \in \mathcal{F}$.

The following lemma will enable us to reduce Proposition 6.11 to Theorem 6.5.

LEMMA 6.12. *Let (Ω, Σ, μ) be a probability space, k a positive integer and $\mathcal{S}$ a k-semiring on Ω with $\mathcal{S} \subseteq \Sigma$. Also let $\mathcal{C}$ be a nonempty finite subfamily of $\mathcal{S}$ consisting of pairwise disjoint sets, $f \in L_1(\Omega, \Sigma, \mu)$ and $0 < \eta \leqslant 1$. Assume that f admits a decomposition $f = f_{\mathrm{str}} + f_{\mathrm{err}} + f_{\mathrm{unf}}$ into integrable random variables* *such that f_{str} is constant on each $S \in \mathcal{C}$ and the functions f_{err} and f_{unf} obey the estimates $\|f_{\mathrm{err}}\|_{L_1} \leqslant \eta^2/8$ and $\|f_{\mathrm{unf}}\|_{\mathcal{S}} \leqslant (\eta^2/8)|\mathcal{C}|^{-1}$. Then we have*

$$\sum_{S \notin \mathrm{Unf}(\mathcal{C}, f, \eta)} \mu(S) \leqslant \eta. \tag{6.22}$$

Proof. Fix $S \notin \mathrm{Unf}(\mathcal{C}, f, \eta)$. We select $T \subseteq S$ with $T \in \mathcal{S}$ such that

$$\eta \cdot \mu(S) < \Big| \int_T \big(f - \mathbb{E}(f \,|\, S)\big)\, d\mu \Big|. \tag{6.23}$$

The function f_{str} is constant on S and so, by (6.23), we see that

$$\eta \cdot \mu(S) < \Big| \int_T \big(f_{\mathrm{err}} - \mathbb{E}(f_{\mathrm{err}} \,|\, S)\big)\, d\mu \Big| + \Big| \int_T \big(f_{\mathrm{unf}} - \mathbb{E}(f_{\mathrm{unf}} \,|\, S)\big)\, d\mu \Big|. \tag{6.24}$$

Next observe that

$$\Big| \int_T \big(f_{\mathrm{err}} - \mathbb{E}(f_{\mathrm{err}} \,|\, S)\big)\, d\mu \Big| \leqslant 2\mathbb{E}(|f_{\mathrm{err}}| \,|\, S) \cdot \mu(S) \tag{6.25}$$

and

$$\Big| \int_T \big(f_{\mathrm{unf}} - \mathbb{E}(f_{\mathrm{unf}} \,|\, S)\big)\, d\mu \Big| \leqslant 2\|f_{\mathrm{unf}}\|_{\mathcal{S}}. \tag{6.26}$$

Finally, notice that $\mu(S) > 0$ since $S \notin \mathrm{Unf}(\mathcal{C}, f, \eta)$. Thus, setting

$$\mathcal{A} = \{S \in \mathcal{C} : \mathbb{E}(|f_{\mathrm{err}}| \,|\, S) \geqslant \eta/4\} \ \text{ and } \ \mathcal{B} = \{S \in \mathcal{C} : \mu(S) \leqslant 4\eta^{-1}\|f_{\mathrm{unf}}\|_{\mathcal{S}}\}$$

and invoking (6.24)–(6.26), we obtain that $\mathcal{C} \setminus \mathrm{Unf}(\mathcal{C}, f, \eta) \subseteq \mathcal{A} \cup \mathcal{B}$.

Now recall that the family $\mathcal{C}$ consists of pairwise disjoint sets. Hence,

$$\sum_{S \in \mathcal{A}} \mu(S) \leqslant \frac{4}{\eta}\Big(\sum_{S \in \mathcal{A}} \int_S |f_{\mathrm{err}}|\, d\mu \Big) \leqslant \frac{4}{\eta}\|f_{\mathrm{err}}\|_{L_1} \leqslant \frac{\eta}{2}. \tag{6.27}$$

Moreover,

$$\sum_{S \in \mathcal{B}} \mu(S) \leqslant \frac{4\|f_{\mathrm{unf}}\|_{\mathcal{S}}}{\eta} \cdot |\mathcal{B}| \leqslant \frac{4\|f_{\mathrm{unf}}\|_{\mathcal{S}}}{\eta} \cdot |\mathcal{C}| \leqslant \frac{\eta}{2}. \tag{6.28}$$

By (6.27) and (6.28) and using the inclusion $\mathcal{C} \setminus \mathrm{Unf}(\mathcal{C}, f, \eta) \subseteq \mathcal{A} \cup \mathcal{B}$, we conclude that the estimate in (6.22) is satisfied. The proof of Lemma 6.12 is completed. □

We proceed to the proof of Proposition 6.11.

Proof of Proposition 6.11. Fix k, ℓ and η. We set $\sigma = \eta^2/8$ and we define $F\colon \mathbb{N} \to \mathbb{R}$ by the rule $F(n) = n/\sigma + 1 = 8n/\eta^2 + 1$ for every $n \in \mathbb{N}$. Notice that F is a growth function. We set $\mathrm{U}(k, \ell, \eta) = \mathrm{RegSz}(k, \ell, \sigma, F)$ and we claim that with this choice the result follows. Indeed, let (Ω, Σ, μ) be a probability space and $\mathcal{S}$ a k-semiring on Ω with $\mathcal{S} \subseteq \Sigma$. Also let $\mathcal{F}$ be a family in $L_2(\Omega, \Sigma, \mu)$ such that $\|f\|_{L_2} \leqslant 1$ for every $f \in \mathcal{F}$ and with $|\mathcal{F}| = \ell$. By Theorem 6.5, there exist a positive integer $M \leqslant \mathrm{U}(k, \ell, \eta)$, a partition $\mathcal{P}$ of Ω with $\mathcal{P} \subseteq \mathcal{S}$ and $|\mathcal{P}| = M$, and for every $f \in \mathcal{F}$ a decomposition $f = f_{\mathrm{str}} + f_{\mathrm{err}} + f_{\mathrm{unf}}$ into integrable random variables such that f_{str} is constant on each $S \in \mathcal{P}$, $\|f_{\mathrm{err}}\|_{L_2} \leqslant \sigma$ and $\|f_{\mathrm{unf}}\|_{\mathcal{S}} \leqslant 1/F(M)$. By the monotonicity of the L_p norms, we have $\|f_{\mathrm{err}}\|_{L_1} \leqslant \sigma$. Hence, by the choice of σ and F and applying Lemma 6.12 for "$\mathcal{C} = \mathcal{P}$", we conclude that the estimate in (6.21) is satisfied for every $f \in \mathcal{F}$. The proof of Proposition 6.11 is completed. □

We close this section by presenting an application of Proposition 6.11 in the context of the Hales–Jewett theorem (see also [**Tao4**]). Let A be a finite alphabet with $|A| \geqslant 2$ and set $K = |A|(|A|-1)2^{-1}$. Also let W be a combinatorial space of $A^{<\mathbb{N}}$. As in Example 6.2, we view W as a discrete probability space equipped with the uniform probability measure and we denote by $\mathcal{S}(W)$ the K-semiring on W consisting of all subsets S of W which are written as

$$S = \bigcap_{\{a,b\} \in \binom{A}{2}} X_{\{a,b\}} \tag{6.29}$$

where $X_{\{a,b\}}$ is (a,b)-insensitive in W for every $\{a,b\} \in \binom{A}{2}$.

Now let D be a subset of W, $0 < \varepsilon \leqslant 1$ and $S \in \mathcal{S}(W)$. Notice that the set S is $(\mathbf{1}_D, \mathcal{S}(W), \varepsilon^2)$-uniform if and only if for every $T \subseteq S$ with $T \in \mathcal{S}(W)$ we have

$$|\mathrm{dens}_T(D) - \mathrm{dens}_S(D)| \cdot \mathrm{dens}_W(T) \leqslant \varepsilon^2 \cdot \mathrm{dens}_W(S). \tag{6.30}$$

In particular, if S is $(\mathbf{1}_D, \mathcal{S}(W), \varepsilon^2)$-uniform, then for every $T \subseteq S$ with $T \in \mathcal{S}(W)$ and $|T| \geqslant \varepsilon|S|$ we have $|\mathrm{dens}_T(D) - \mathrm{dens}_S(D)| \leqslant \varepsilon$. Thus, by Proposition 6.11 and taking into account these remarks, we obtain the following corollary.

Corollary 6.13. *For every $k \in \mathbb{N}$ with $k \geqslant 2$ and every $0 < \varepsilon \leqslant 1$ there exists a positive integer $N(k,\varepsilon)$ with the following property. If A is an alphabet with $|A| = k$, W is a combinatorial space of $A^{<\mathbb{N}}$ and D is a subset of W, then there exist a positive integer $M \leqslant N(k,\varepsilon)$, a partition $\mathcal{P}$ of W with $\mathcal{P} \subseteq \mathcal{S}(W)$ and $|\mathcal{P}| = M$, and a subfamily $\mathcal{P}' \subseteq \mathcal{P}$ with $\mathrm{dens}_W(\cup\mathcal{P}') \geqslant 1 - \varepsilon$ such that*

$$|\mathrm{dens}_T(D) - \mathrm{dens}_S(D)| \leqslant \varepsilon \tag{6.31}$$

for every $S \in \mathcal{P}'$ and every $T \subseteq S$ with $T \in \mathcal{S}(W)$ and $|T| \geqslant \varepsilon|S|$.

6.2. Szemerédi's regularity lemma

Let $G = (V,E)$ be a finite graph and X, Y two nonempty disjoint subsets of V. The *edge density* $d(X,Y)$ *between* X *and* Y is the quantity defined by

$$d(X,Y) = \frac{|E \cap (X \times Y)|}{|X| \cdot |Y|}. \tag{6.32}$$

Also let $0 < \varepsilon \leqslant 1$. The pair (X,Y) is said to be ε-*regular* (*with respect to* G) if for every $X' \subseteq X$ and every $Y' \subseteq Y$ with $|X'| \geqslant \varepsilon|X|$ and $|Y'| \geqslant \varepsilon|Y|$ we have

$$|d(X',Y') - d(X,Y)| \leqslant \varepsilon. \tag{6.33}$$

Otherwise, the pair (X,Y) is said to be ε-*irregular*.

The following result is known as *Szemerédi's regularity lemma* and is due to Szemerédi [**Sz2**].

Theorem 6.14. *For every $0 < \varepsilon \leqslant 1$ and every integer $m \geqslant 1$ there exist two positive integers $t_{\mathrm{Sz}}(\varepsilon,m)$ and $K_{\mathrm{Sz}}(\varepsilon,m)$ with the following property. If $G = (V,E)$ is a finite graph with $|V| \geqslant t_{\mathrm{Sz}}(\varepsilon,m)$, then there exists an integer K with*

$$m \leqslant K \leqslant K_{\mathrm{Sz}}(\varepsilon,m) \tag{6.34}$$

and a partition $\mathcal{V} = \{V_1, \dots, V_K\}$ of V such that: (i) *$\mathcal{V}$ is equitable in the sense that $\big||V_i| - |V_j|\big| \leqslant 1$ for every $i, j \in [K]$, and* (ii) *the pair (V_i, V_j) is ε-regular for all but at most $\varepsilon\binom{K}{2}$ of the pairs $1 \leqslant i < j \leqslant K$.*

Szemerédi's regularity lemma is one of the most important structural results about large dense graphs and has had a huge impact on the development of extremal combinatorics. Several of its applications are discussed in [**KS, KSSS**].

We will present a proof of Szemerédi's regularity lemma using Theorem 6.5 as a main tool. In particular, for every finite graph $G = (V, E)$ we view the set $V \times V$ as a discrete probability space equipped with the uniform probability measure. By $\mathcal{S}(G)$ we shall denote the set of all rectangles of $V \times V$, that is,

$$\mathcal{S}(G) = \{X \times Y : X, Y \subseteq V\}.$$

Notice that, by Lemma 6.3, the family $\mathcal{S}(G)$ is a 2-semiring.

We will also need the following facts. The first one relates the notion of uniformity introduced in Definition 6.10 with the graph-theoretic concept of regularity.

FACT 6.15. *Let $G = (V, E)$ be a finite graph. Also let $0 < \varepsilon \leqslant 1$ and X, Y two nonempty disjoint subsets of V. If the set $X \times Y$ is $(\mathbf{1}_E, \mathcal{S}(G), \varepsilon^3)$-uniform, then the pair (X, Y) is ε-regular with respect to G.*

PROOF. Let $X' \subseteq X$ and $Y' \subseteq Y$ with $|X'| \geqslant \varepsilon|X|$ and $|Y'| \geqslant \varepsilon|Y|$ be arbitrary. Notice that $X' \times Y' \in \mathcal{S}(G)$ and $\mathrm{dens}_{V\times V}(X' \times Y') \geqslant \varepsilon^2 \cdot \mathrm{dens}_{V\times V}(X \times Y)$. By our assumption that the set $X \times Y$ is $(\mathbf{1}_E, \mathcal{S}(G), \varepsilon^3)$-uniform, we obtain that

$$|\mathrm{dens}_{X'\times Y'}(E) - \mathrm{dens}_{X\times Y}(E)| \cdot \mathrm{dens}_{V\times V}(X' \times Y') \leqslant \varepsilon^3 \cdot \mathrm{dens}_{V\times V}(X \times Y).$$

Since $\mathrm{dens}_{X'\times Y'}(E) = d(X', Y')$ and $\mathrm{dens}_{X\times Y}(E) = d(X, Y)$, we conclude that $|d(X', Y') - d(X, Y)| \leqslant \varepsilon$ and the proof of Fact 6.15 is completed. $\square$

The second fact is a general stability property of uniform sets.

FACT 6.16. *Let (Ω, Σ, μ) be a probability space, k a positive integer and $\mathcal{S}$ a k-semiring on Ω with $\mathcal{S} \subseteq \Sigma$. Also let f be a $[0,1]$-valued random variable, $0 < \eta, \delta < 1$ and $S \in \mathcal{S}$ with $\eta + 3\delta \leqslant 1$ and $\mu(S) > 0$. Assume that the set S is $(f, \mathcal{S}, \eta)$-uniform. If $S' \in \mathcal{S}$ is such that $S \subseteq S'$ and $\mu(S') \leqslant (1+\delta)\mu(S)$, then the set S' is $(f, \mathcal{S}, \eta + 3\delta)$-uniform.*

PROOF. We fix $S' \in \mathcal{S}$ with $S \subseteq S'$ and $\mu(S') \leqslant (1+\delta)\mu(S)$. Let $T' \subseteq S'$ with $T' \in \mathcal{S}$ be arbitrary and set $T = T' \cap S$. Notice that $T \in \mathcal{S}$ and $T \subseteq S$. Moreover, the fact that f takes values in $[0,1]$ implies that $|f - \mathbb{E}(f \,|\, S')| \leqslant 1$. Thus, by our assumptions and the triangle inequality, we obtain that

$$\begin{aligned} \Big|\int_{T'} \big(f - \mathbb{E}(f \,|\, S')\big)\, d\mu\Big| &\leqslant \Big|\int_{T'\setminus T} \big(f - \mathbb{E}(f \,|\, S')\big)\, d\mu\Big| + \Big|\int_{T} \big(f - \mathbb{E}(f \,|\, S')\big)\, d\mu\Big| \\ &\leqslant \delta\mu(S) + \eta\mu(S) + \int_T |\mathbb{E}(f \,|\, S) - \mathbb{E}(f \,|\, S')|\, d\mu. \end{aligned}$$

Invoking once again the fact that the random variable f is $[0,1]$-valued, we see that $|\mathbb{E}(f \mid S) - \mathbb{E}(f \mid S')| \leqslant 2\mu(S' \setminus S)/\mu(S')$. Therefore,

$$\int_T |\mathbb{E}(f \mid S) - \mathbb{E}(f \mid S')| \, d\mu \leqslant \frac{2\mu(S' \setminus S)}{\mu(S')} \cdot \mu(T) \leqslant 2\mu(S' \setminus S) \leqslant 2\delta\mu(S).$$

Summing up, we conclude that

$$\Big| \int_{T'} \big(f - \mathbb{E}(f \mid S')\big) \, d\mu \Big| \leqslant \delta\mu(S) + \eta\mu(S) + 2\delta\mu(S) \leqslant (\eta + 3\delta)\, \mu(S')$$

and the proof of Fact 6.16 is completed. □

The third, and last, fact will enable us to produce an equitable partition of the vertex set V of a finite graph from a given partition of $V \times V$.

FACT 6.17. *Let M be a positive integer and $0 < \theta \leqslant 1/2$. Also let V be a finite set with $|V| \geqslant 4^M \theta^{-3}$ and $\mathcal{P}$ a partition of $V \times V$ into M sets of the form $X \times Y$ where $X, Y \subseteq V$. Then there exist a positive integer K with*

$$(1-\theta)\theta^{-1} 4^M \leqslant K \leqslant 2\theta^{-1} 4^M, \tag{6.35}$$

a family $\mathcal{U} = \{U_1, \dots, U_K\}$ of pairwise disjoint subsets of V and a partition $\mathcal{V} = \{V_1, \dots, V_K\}$ of V such that: (i) $|U_i| = \lfloor \theta |V| \cdot 4^{-M} \rfloor$ *for every $i \in [K]$,* (ii) *the set $U_i \times U_j$ is contained in a* (*necessarily unique*) *element of $\mathcal{P}$ for every $i, j \in [K]$,* (iii) $\big| |V_i| - |V_j| \big| \leqslant 1$ *for every $i, j \in [K]$,* (iv) *$U_i \subseteq V_i$ for every $i \in [K]$, and* (v) $|V_i| \leqslant (1+3\theta)|U_i|$ *for every $i \in [K]$.*

PROOF. We write the partition $\mathcal{P}$ as $\{X_i \times Y_i : i \in [M]\}$. Let $\mathcal{A}$ and $\mathcal{B}$ be the sets of all nonempty atoms of the algebras generated by the families $\{X_i : i \in [M]\}$ and $\{Y_i : i \in [M]\}$ respectively. We set $\mathcal{R}_0 = \{A \cap B : A \in \mathcal{A} \text{ and } B \in \mathcal{B}\}$ and we observe that $\mathcal{R}_0$ is a partition of V with $|\mathcal{R}_0| \leqslant 4^M$ and such that the family $\mathcal{P}' \coloneqq \{X \times Y : X, Y \in \mathcal{R}_0\}$ is a refinement of $\mathcal{P}$. Next, we partition every $X \in \mathcal{R}_0$ into disjoint sets of size $N = \lfloor \theta |V| \cdot 4^{-M} \rfloor$ plus an error set of size at most N. (Since $|V| \geqslant 4^M \theta^{-3}$ we have $N \geqslant 1$, and so such a partition is possible.) Let E_0 be the union of all the error sets and let $U_1, \dots, U_K$ be the remaining sets. Notice that

$$|E_0| \leqslant |\mathcal{R}_0| N \leqslant 4^M N \leqslant \theta |V|.$$

In particular, we have

$$|V| - \theta|V| \leqslant KN \leqslant |V| \tag{6.36}$$

which is easily seen to imply the estimate on K in (6.35). Also observe that for every $i, j \in [K]$ there exist $X, Y \in \mathcal{R}_0$ such that $U_i \subseteq X$ and $U_j \subseteq Y$. This implies, of course, that $U_i \times U_j \subseteq X \times Y \in \mathcal{P}'$. Using the fact that $\mathcal{P}'$ is a refinement of $\mathcal{P}$, we obtain that $U_i \times U_j$ is contained in a unique element of $\mathcal{P}$. Thus, the family $\mathcal{U} \coloneqq \{U_1, \dots, U_K\}$ satisfies parts (i) and (ii).

We proceed to define the partition $\mathcal{V}$. We break up the set E_0 arbitrarily into K sets $E_1, \dots, E_K$ such that $\big| |E_i| - |E_j| \big| \leqslant 1$ for every $i, j \in [K]$ and we set $V_i = U_i \cup E_i$ for every $i \in [K]$. It is then clear that the partition $\mathcal{V} \coloneqq \{V_1, \dots, V_K\}$ satisfies parts (iii) and (iv). To see that part (v) is also satisfied, let $x = \min\{|E_i| : i \in [K]\}$ and observe that $x \leqslant |E_i| \leqslant x+1$ for every $i \in [K]$. Notice that the family $\{E_1, \dots, E_K\}$ is a partition of E_0 and recall that $0 < \theta \leqslant 1/2$. Hence,

$$Kx \leqslant \theta|V| \overset{(6.36)}{\leqslant} \theta(1-\theta)^{-1}KN \leqslant 2\theta KN.$$

which implies that $0 \leqslant x \leqslant 2\theta N$. Using again the fact that $0 < \theta \leqslant 1/2$ we see that $\theta^{-2} - \theta^{-1} > 1$ and, in particular, that $\theta^{-1} \leqslant \lfloor \theta^{-2} \rfloor$. Since $|V| \geqslant 4^M\theta^{-3}$ we have $\theta N = \theta\lfloor \theta|V| \cdot 4^{-M}\rfloor \geqslant \theta\lfloor\theta^{-2}\rfloor \geqslant 1$. Therefore, for every $i \in [K]$ we obtain that

$$|V_i| = |U_i| + |V_i \setminus U_i| = N + |E_i| \leqslant N + x + 1 \leqslant N + 3\theta N = (1+3\theta)|U_i|.$$

The proof of Fact 6.17 is completed. $\square$

We are ready to give the proof of Theorem 6.14.

Proof of Theorem 6.14. We follow the proof from [**Tao2**]. Fix $0 < \varepsilon \leqslant 1$ and a positive integer m. Let

$$\eta = \frac{\varepsilon^3}{64}, \quad \theta = \frac{\eta}{45m} \quad \text{and} \quad \sigma = \frac{\eta^2}{8} \tag{6.37}$$

and define $F\colon \mathbb{N} \to \mathbb{R}$ by the rule $F(n) = \sigma^{-1}(2\theta^{-1}4^n)^2$ for every $n \in \mathbb{N}$. Notice that F is a growth function. Finally, let

$$M_0 = \mathrm{RegSz}(2, 1, \sigma, F) \tag{6.38}$$

and set

$$t_{\mathrm{Sz}}(\varepsilon, m) = \lceil 4^{M_0}\theta^{-3}\rceil \quad \text{and} \quad K_{\mathrm{Sz}}(\varepsilon, m) = \lfloor 2\theta^{-1}4^{M_0}\rfloor. \tag{6.39}$$

We will show that $t_{\mathrm{Sz}}(\varepsilon, m)$ and $K_{\mathrm{Sz}}(\varepsilon, m)$ are as desired.

Let $G = (V, E)$ be an arbitrary finite graph with $|V| \geqslant t_{\mathrm{Sz}}(\varepsilon, m)$. By Theorem 6.5 applied for the 2-semiring $\mathcal{S}(G)$ and the family $\mathcal{F} = \{\mathbf{1}_E\}$, there exist a positive integer $M \leqslant M_0$, a partition $\mathcal{P}$ of $V \times V$ with $\mathcal{P} \subseteq \mathcal{S}(G)$ and $|\mathcal{P}| = M$, and a decomposition $\mathbf{1}_E = f_{\mathrm{str}} + f_{\mathrm{err}} + f_{\mathrm{unf}}$ such that f_{str} is constant on each $S \in \mathcal{P}$, $\|f_{\mathrm{err}}\|_{L_2} \leqslant \sigma$ and $\|f_{\mathrm{unf}}\|_{\mathcal{S}} \leqslant 1/F(M)$. By the choice of $t_{\mathrm{Sz}}(\varepsilon, m)$ in (6.39) and the fact that $|\mathcal{P}| = M \leqslant M_0$, we have $|V| \geqslant 4^M\theta^{-3}$. Thus, by Fact 6.17, there exist a positive integer K with

$$(1-\theta)\theta^{-1}4^M \leqslant K \leqslant 2\theta^{-1}4^M, \tag{6.40}$$

a family $\mathcal{U} = \{U_1, \dots, U_K\}$ of pairwise disjoint subsets of V and a partition $\mathcal{V} = \{V_1, \dots, V_K\}$ of V satisfying parts (i)–(v) of Fact 6.17.

We claim that K and $\mathcal{V}$ satisfy the requirements of the theorem. Indeed, notice first that, by (6.40) and the choice of θ and $K_{\mathrm{Sz}}(\varepsilon, m)$ in (6.37) and (6.39), we have $m \leqslant K \leqslant K_{\mathrm{Sz}}(\varepsilon, m)$. Also let

$$\mathcal{D} = \{V_i \times V_j : 1 \leqslant i < j \leqslant K\} \quad \text{and} \quad \mathcal{C} = \{U_i \times U_j : 1 \leqslant i < j \leqslant K\}.$$

As in Definition 6.10, by $\mathrm{Unf}(\mathcal{D}, \mathbf{1}_E, 2\eta)$ we denote the set of all $V_i \times V_j \in \mathcal{D}$ such that $V_i \times V_j$ is $(\mathbf{1}_E, \mathcal{S}(G), 2\eta)$-uniform. Respectively, $\mathrm{Unf}(\mathcal{C}, \mathbf{1}_E, \eta)$ stands for set of all $U_i \times U_j \in \mathcal{C}$ such that $U_i \times U_j$ is $(\mathbf{1}_E, \mathcal{S}(G), \eta)$-uniform.

Let $1 \leqslant i < j \leqslant K$ and assume that the pair (V_i, V_j) is ε-irregular. By (6.37), we have $2\eta \leqslant \varepsilon^3$ and so, by Fact 6.15, we obtain that $V_i \times V_j \notin \mathrm{Unf}(\mathcal{D}, \mathbf{1}_E, 2\eta)$. Next observe that, by the choice of $\mathcal{U}$ and $\mathcal{V}$, we have $U_i \times U_j \subseteq V_i \times V_j$ and

$|V_i \times V_j| \leqslant (1+15\theta)|U_i \times U_j|$. On the other hand, by (6.37), we have $45\theta \leqslant \eta$ and $\eta + 45\theta \leqslant 1$. Thus, by Fact 6.16 and the previous discussion, we conclude that $U_i \times U_j \notin \mathrm{Unf}(\mathcal{C}, \mathbf{1}_E, \eta)$. Therefore, it is enough to show that

$$|\mathcal{C} \setminus \mathrm{Unf}(\mathcal{C}, \mathbf{1}_E, \eta)| \leqslant \varepsilon \binom{K}{2}. \tag{6.41}$$

To this end, notice that $\mathcal{C} \subseteq \mathcal{S}(G)$ is a collection of pairwise disjoint sets. Moreover, for every $U_i \times U_j \in \mathcal{C}$ we have

$$\mathrm{dens}_{V\times V}(U_i \times U_j) = \frac{\lfloor \theta|V| \cdot 4^{-M} \rfloor^2}{|V|^2} \geqslant \Big(\frac{\theta 4^{-M}}{2}\Big)^2 \overset{(6.40)}{\geqslant} \Big(\frac{1-\theta}{2K}\Big)^2 \overset{(6.37)}{\geqslant} (4K)^{-2}.$$

Invoking the definition of the growth function F, we also have

$$|\mathcal{C}| \leqslant K^2 \overset{(6.40)}{\leqslant} (2\theta^{-1}4^M)^2 = \sigma F(M)$$

and so $\|f_{\mathrm{unf}}\|_{\mathcal{S}} \leqslant 1/F(M) \leqslant \sigma|\mathcal{C}|^{-1} = (\eta^2/8)|\mathcal{C}|^{-1}$. On the other hand, observe that $\|f_{\mathrm{err}}\|_{L_1} \leqslant \|f_{\mathrm{err}}\|_{L_2} \leqslant \sigma = \eta^2/8$. Hence, by Lemma 6.12 and the previous estimates, we obtain that

$$|\mathcal{C} \setminus \mathrm{Unf}(\mathcal{C}, \mathbf{1}_E, \eta)| \cdot (4K)^{-2} \leqslant \sum_{U_i \times U_j \notin \mathrm{Unf}(\mathcal{C}, \mathbf{1}_E, \eta)} \mathrm{dens}_{V\times V}(U_i \times U_j) \leqslant \eta. \tag{6.42}$$

Finally, by (6.37) and (6.40), we see that $K \geqslant 2$ which implies that $K^2 \leqslant 4\binom{K}{2}$. Therefore, we conclude that

$$|\mathcal{C} \setminus \mathrm{Unf}(\mathcal{C}, \mathbf{1}_E, \eta)| \overset{(6.42)}{\leqslant} 16\eta K^2 \overset{(6.37)}{\leqslant} \frac{\varepsilon}{4} K^2 \leqslant \varepsilon \binom{K}{2}.$$

The proof of Theorem 6.14 is completed. □

6.3. A concentration inequality for product spaces

6.3.1. The main result. In this section we will present a concentration inequality for product spaces which asserts that every square-integrable random variable defined on the product of sufficiently many probability spaces exhibits pseudorandom behavior. Combinatorial applications will be discussed in Subsections 6.3.2 and 6.3.3.

First we introduce some notation concerning product spaces. Let n be a positive integer and let $(\Omega_1, \Sigma_1, \mu_1), \dots, (\Omega_n, \Sigma_n, \mu_n)$ be a finite sequence of probability spaces. As in Example 6.1, by $(\mathbf{\Omega}, \mathbf{\Sigma}, \boldsymbol{\mu})$ we shall denote their product, while for every nonempty $I \subseteq [n]$ by $(\mathbf{\Omega}_I, \mathbf{\Sigma}_I, \boldsymbol{\mu}_I)$ we shall denote the product of the spaces $\langle (\Omega_i, \Sigma_i, \mu_i) : i \in I \rangle$. (Recall that, by convention, $\mathbf{\Omega}_\emptyset$ stands for the empty set.) If $I \subseteq [n]$, $\mathbf{x} \in \mathbf{\Omega}_I$ and $\mathbf{y} \in \mathbf{\Omega}_{[n]\setminus I}$, then by $(\mathbf{x}, \mathbf{y})$ we shall denote the unique element $\mathbf{z}$ of $\mathbf{\Omega}$ such that $\pi_I(\mathbf{z}) = \mathbf{x}$ and $\pi_{[n]\setminus I}(\mathbf{z}) = \mathbf{y}$. (Here, $\pi_I \colon \mathbf{\Omega} \to \mathbf{\Omega}_I$ and $\pi_{[n]\setminus I} \colon \mathbf{\Omega} \to \mathbf{\Omega}_{[n]\setminus I}$ are the natural projections.) Finally, for every function $f \colon \mathbf{\Omega} \to \mathbb{R}$ and every $\mathbf{x} \in \mathbf{\Omega}_I$ by $f_{\mathbf{x}} \colon \mathbf{\Omega}_{[n]\setminus I} \to \mathbb{R}$ we shall denote the map defined by $f_{\mathbf{x}}(\mathbf{y}) = f\big((\mathbf{x}, \mathbf{y})\big)$. Notice, in particular, that for every subset A of $\mathbf{\Omega}$ the function $(\mathbf{1}_A)_{\mathbf{x}}$ coincides with the characteristic function of the section $A_{\mathbf{x}} = \{\mathbf{y} \in \mathbf{\Omega}_{[n]\setminus I} : (\mathbf{x}, \mathbf{y}) \in A\}$ of A at $\mathbf{x}$.

Now for every positive integer ℓ, every $0 < \sigma \leqslant 1$ and every $F\colon \mathbb{N} \to \mathbb{N}$ we set

$$\mathrm{ConcProd}(\ell, \sigma, F) = F^{(\lceil \sigma^{-3}\ell \rceil)}(1). \tag{6.43}$$

The following theorem is the main result of this section.

THEOREM 6.18. *Let ℓ be a positive integer, $0 < \sigma \leqslant 1$ and $F\colon \mathbb{N} \to \mathbb{N}$ such that $F(m) \geqslant m+1$ for every $m \in \mathbb{N}$. Also let n be a positive integer with $n \geqslant F\big(\mathrm{ConcProd}(\ell,\sigma,F)\big)+1$ and let $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ be the product of a finite sequence $(\Omega_1, \Sigma_1, \mu_1), \dots, (\Omega_n, \Sigma_n, \mu_n)$ of probability spaces. If $\mathcal{F}$ is a family in $L_2(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ such that $\|f\|_{L_2} \leqslant 1$ for every $f \in \mathcal{F}$ and with $|\mathcal{F}| = \ell$, then there exists a positive integer M with*

$$M \leqslant \mathrm{ConcProd}(\ell, \sigma, F) \tag{6.44}$$

such that for every nonempty $I \subseteq \{M+1, \dots, F(M)\}$ and every $f \in \mathcal{F}$ we have

$$\boldsymbol{\mu}_I\big(\{\mathbf{x} \in \boldsymbol{\Omega}_I : |\mathbb{E}(f_{\mathbf{x}}) - \mathbb{E}(f)| \leqslant \sigma\}\big) \geqslant 1 - \sigma. \tag{6.45}$$

The proof of Theorem 6.18 is also based on Fact 6.6. Specifically, let $m \in [n]$ and recall that $(\boldsymbol{\Omega}_{[m]}, \boldsymbol{\Sigma}_{[m]}, \boldsymbol{\mu}_{[m]})$ stands for the product of the probability spaces $(\Omega_1, \Sigma_1, \mu_1), \dots, (\Omega_m, \Sigma_m, \mu_m)$. As in Example 6.1, we may "extend" the σ-algebra $\boldsymbol{\Sigma}_{[m]}$ to the full product $\boldsymbol{\Omega}$ using the projection $\pi_{[m]}$. Indeed, for every $m \in [n]$ let

$$\mathcal{B}_m = \big\{\pi^{-1}_{[m]}(\mathbf{A}) : \mathbf{A} \in \boldsymbol{\Sigma}_{[m]}\big\} \tag{6.46}$$

and observe that $\mathcal{B}_m = \{\mathbf{A} \times \boldsymbol{\Omega}_{[n]\setminus[m]} : \mathbf{A} \in \boldsymbol{\Sigma}_{[m]}\}$ if $m < n$ while $\mathcal{B}_n = \boldsymbol{\Sigma}$. It follows that $(\mathcal{B}_m)_{m=1}^n$ is an increasing finite sequence of sub-σ-algebras of $\boldsymbol{\Sigma}$, and so for every $f \in L_2(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ with $\|f\|_{L_2} \leqslant 1$ the sequence $\mathbb{E}(f \,|\, \mathcal{B}_1), \dots, \mathbb{E}(f \,|\, \mathcal{B}_n)$ is a finite martingale which is contained in the unit ball of $L_2(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$. We have the following property which is satisfied by all finite martingales of this form.

LEMMA 6.19. *Let l be a positive integer, $0 < \eta \leqslant 1$ and $\Phi\colon \mathbb{N} \to \mathbb{N}$ such that $\Phi(m) \geqslant m+1$ for every $m \in \mathbb{N}$. Also let $n \in \mathbb{N}$ with $n \geqslant \Phi^{(\lceil \eta^{-1} l \rceil + 1)}(1)$. Finally, let (Ω, Σ, μ) be a probability space, $(\mathcal{A}_m)_{m=1}^n$ an increasing finite sequence of sub-σ-algebras of Σ and $\mathcal{F}$ a family in $L_2(\Omega, \Sigma, \mu)$ such that $\|f\|_{L_2} \leqslant 1$ for every $f \in \mathcal{F}$ and with $|\mathcal{F}| = l$. Then there exists a positive integer M with*

$$M \leqslant \Phi^{(\lceil \eta^{-1} l \rceil)}(1) \tag{6.47}$$

such that for every $f \in \mathcal{F}$ we have

$$\|\mathbb{E}(f \,|\, \mathcal{A}_{\Phi(M)}) - \mathbb{E}(f \,|\, \mathcal{A}_M)\|^2_{L_2} \leqslant \eta. \tag{6.48}$$

PROOF. Assume that there is no positive integer M which satisfies (6.47) and (6.48). This implies that for every $M \in \{1, \dots, \Phi^{(\lceil \eta^{-1} l \rceil)}(1)\}$ there exists $f \in \mathcal{F}$ such that $\|\mathbb{E}(f \,|\, \mathcal{A}_{\Phi(M)}) - \mathbb{E}(f \,|\, \mathcal{A}_M)\|^2_{L_2} > \eta$. Therefore, for every $i \in \{1, \dots, \lceil \eta^{-1} l \rceil\}$ we may select $f_i \in \mathcal{F}$ such that $\|\mathbb{E}(f_i \,|\, \mathcal{A}_{\Phi(M_i)}) - \mathbb{E}(f_i \,|\, \mathcal{A}_{M_i})\|^2_{L_2} > \eta$ where $M_1 = 1$ and $M_i = \Phi(M_{i-1}) = \Phi^{(i-1)}(1)$ if $i \geqslant 2$. By the classical pigeonhole principle, there exist $g \in \mathcal{F}$ and a subset I of $\{1, \dots, \lceil \eta^{-1} l \rceil\}$ with $|I| \geqslant \lceil \eta^{-1} l \rceil / |\mathcal{F}|$ and such that $f_i = g$ for every $i \in I$. Hence, by Fact 6.6, we obtain that

$$1 \leqslant \eta |I| < \|\mathbb{E}(g \,|\, \mathcal{A}_n)\|^2_{L_2} \leqslant \|g\|^2_{L_2} \leqslant 1$$

which is clearly a contradiction. The proof of Lemma 6.19 is completed. $\square$

We will also need the following lemma. In its proof, and in the rest of this subsection, we will follow the common practice when proving inequalities and we will ignore measurability issues since they can be easily resolved with standard arguments.

LEMMA 6.20. *Let $I \subseteq [n]$ and assume that both I and $[n] \setminus I$ are nonempty. Then for every $g, h \in L_2(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ we have*

$$\int \|g_{\mathbf{x}} - h_{\mathbf{x}}\|_{L_1}^2 \, d\boldsymbol{\mu}_I \leqslant \|g - h\|_{L_2}^2. \tag{6.49}$$

PROOF. By Fubini's theorem, we see that

$$\|g - h\|_{L_2}^2 = \int \Big(\int |g_{\mathbf{x}} - h_{\mathbf{x}}|^2 \, d\boldsymbol{\mu}_{[n]\setminus I} \Big) \, d\boldsymbol{\mu}_I. \tag{6.50}$$

On the other hand, by Jensen's inequality, for every $\mathbf{x} \in \boldsymbol{\Omega}_I$ we have

$$\|g_{\mathbf{x}} - h_{\mathbf{x}}\|_{L_1}^2 = \Big(\int |g_{\mathbf{x}} - h_{\mathbf{x}}| \, d\boldsymbol{\mu}_{[n]\setminus I} \Big)^2 \leqslant \int |g_{\mathbf{x}} - h_{\mathbf{x}}|^2 \, d\boldsymbol{\mu}_{[n]\setminus I} \tag{6.51}$$

and so, taking the average over all $\mathbf{x} \in \boldsymbol{\Omega}_I$ and using (6.50), we conclude that the estimate in (6.49) is satisfied. The proof of Lemma 6.20 is completed. □

We are ready to give the proof of Theorem 6.18.

PROOF OF THEOREM 6.18. Fix a family $\mathcal{F}$ in $L_2(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ with $\|f\|_{L_2} \leqslant 1$ for every $f \in \mathcal{F}$ and $|\mathcal{F}| = \ell$. We apply Lemma 6.19 to the sequence $(\mathcal{B}_m)_{m=1}^n$ defined in (6.46), the family $\mathcal{F}$ and the constant "$\eta = \sigma^3$", and we obtain a positive integer $M \leqslant \mathrm{ConcProd}(\ell, \sigma, F)$ such that

$$\|\mathbb{E}(f \,|\, \mathcal{B}_{F(M)}) - \mathbb{E}(f \,|\, \mathcal{B}_M)\|_{L_2}^2 \leqslant \sigma^3 \tag{6.52}$$

for every $f \in \mathcal{F}$. We will show that the positive integer M is as desired.

Notice, first, that the estimate in (6.44) is satisfied. Next, fix a nonempty subset I of $\{M+1, \dots, F(M)\}$ and let $f \in \mathcal{F}$ be arbitrary. We set

$$g = \mathbb{E}(f \,|\, \mathcal{B}_{F(M)}) \ \text{ and } \ h = \mathbb{E}(f \,|\, \mathcal{B}_M). \tag{6.53}$$

We have the following claim.

CLAIM 6.21. *For every $\mathbf{x} \in \boldsymbol{\Omega}_I$ we have $\mathbb{E}(g_{\mathbf{x}}) = \mathbb{E}(f_{\mathbf{x}})$ and $\mathbb{E}(h_{\mathbf{x}}) = \mathbb{E}(f)$.*

PROOF OF CLAIM 6.21. Fix $\mathbf{x} \in \boldsymbol{\Omega}_I$. Since $I \subseteq [F(M)]$, by (6.53) and Fubini's theorem, we see that for every $\mathbf{y} \in \boldsymbol{\Omega}_{[F(M)]\setminus I}$ the function $g_{(\mathbf{x},\mathbf{y})} \colon \boldsymbol{\Omega}_{[n]\setminus[F(M)]} \to \mathbb{R}$ is constant and equal to $\mathbb{E}(f_{(\mathbf{x},\mathbf{y})})$. Therefore,

$$\begin{aligned}
\mathbb{E}(g_{\mathbf{x}}) = \int g_{\mathbf{x}} \, d\boldsymbol{\mu}_{[n]\setminus I} &= \int \Big(\int g_{(\mathbf{x},\mathbf{y})} \, d\boldsymbol{\mu}_{[n]\setminus[F(M)]} \Big) \, d\boldsymbol{\mu}_{[F(M)]\setminus I} \\
&= \int \mathbb{E}(f_{(\mathbf{x},\mathbf{y})}) \, d\boldsymbol{\mu}_{[F(M)]\setminus I} \\
&= \int \Big(\int f_{(\mathbf{x},\mathbf{y})} \, d\boldsymbol{\mu}_{[n]\setminus[F(M)]} \Big) \, d\boldsymbol{\mu}_{[F(M)]\setminus I} \\
&= \int f_{\mathbf{x}} \, d\boldsymbol{\mu}_{[n]\setminus I} = \mathbb{E}(f_{\mathbf{x}}).
\end{aligned}$$

We proceed to show that $\mathbb{E}(h_{\mathbf{x}}) = \mathbb{E}(f)$. As above we observe that, by (6.53) and Fubini's theorem, for every $\mathbf{z} \in \mathbf{\Omega}_{[M]}$ the function $h_{\mathbf{z}} \colon \mathbf{\Omega}_{[n]\setminus[M]} \to \mathbb{R}$ is constant and equal to $\mathbb{E}(f_{\mathbf{z}})$. Since $I \cap [M] = \emptyset$, the function $h_{(\mathbf{x},\mathbf{z})} \colon \mathbf{\Omega}_{[n]\setminus(I\cup[M])} \to \mathbb{R}$ is also constant and equal to $\mathbb{E}(f_{\mathbf{z}})$. Hence,

$$\begin{aligned}\mathbb{E}(h_{\mathbf{x}}) = \int h_{\mathbf{x}}\, d\boldsymbol{\mu}_{[n]\setminus I} &= \int \Big(\int h_{(\mathbf{x},\mathbf{z})}\, d\boldsymbol{\mu}_{[n]\setminus(I\cup[M])}\Big)\, d\boldsymbol{\mu}_{[M]} \\ &= \int \mathbb{E}(f_{\mathbf{z}})\, d\boldsymbol{\mu}_{[M]} = \mathbb{E}(f).\end{aligned}$$

The proof of Claim 6.21 is completed. □

By Claim 6.21, for every $\mathbf{x} \in \mathbf{\Omega}_I$ we have

$$|\mathbb{E}(f_{\mathbf{x}}) - \mathbb{E}(f)| = \Big| \int (g_{\mathbf{x}} - h_{\mathbf{x}})\, d\boldsymbol{\mu}_{[n]\setminus I}\Big| \leqslant \|g_{\mathbf{x}} - h_{\mathbf{x}}\|_{L_1}$$

and so, by (6.52), (6.53) and Lemma 6.20, we obtain that

$$\int |\mathbb{E}(f_{\mathbf{x}}) - \mathbb{E}(f)|^2\, d\boldsymbol{\mu}_I \leqslant \sigma^3. \tag{6.54}$$

By Markov's inequality, we conclude that

$$\boldsymbol{\mu}_I\big(\{\mathbf{x} \in \mathbf{\Omega}_I : |\mathbb{E}(f_{\mathbf{x}}) - \mathbb{E}(f)| \leqslant \sigma\}\big) \geqslant 1 - \sigma \tag{6.55}$$

and the proof of Theorem 6.18 is completed. □

6.3.2. Combinatorial spaces. In the last two subsections we will present two combinatorial applications of Theorem 6.18. The first one is in the context of the Hales–Jewett theorem and asserts that every dense subset of a high-dimensional hypercube can be effectively modeled as a family of measurable events indexed by the elements of another hypercube of smaller, but large enough, dimension. Related results will also be obtained in Chapters 8 and 9.

We proceed to the details. Let k, ℓ, m be positive integers with $k \geqslant 2$ and $0 < \varepsilon \leqslant 1$. We set

$$\sigma = \min\{\varepsilon, k^{-m}/2\} \tag{6.56}$$

and we define $F \colon \mathbb{N} \to \mathbb{N}$ by the rule $F(n) = n + m$ for every $n \in \mathbb{N}$. Finally, let

$$\mathrm{RegSp}(k, \ell, m, \varepsilon) = F\big(\mathrm{ConcProd}(\ell, \sigma, F)\big) + 1. \tag{6.57}$$

We have the following lemma.

Lemma 6.22. *Let k, ℓ, m be positive integers with $k \geqslant 2$ and $0 < \varepsilon \leqslant 1$. Also let A be an alphabet with $|A| = k$ and $n \geqslant \mathrm{RegSp}(k, \ell, m, \varepsilon)$. If $\mathcal{F}$ is a family of subsets of A^n with $|\mathcal{F}| = \ell$, then there exists an interval $I \subseteq \{l \in \mathbb{N} : l < n\}$ with $|I| = m$ such that for every $D \in \mathcal{F}$ and every $t \in A^I$ we have*

$$|\mathrm{dens}(D_t) - \mathrm{dens}(D)| \leqslant \varepsilon \tag{6.58}$$

where $D_t = \{s \in A^{\{l\in\mathbb{N}\colon l<n\}\setminus I} : (t, s) \in D\}$ is the section of D at t.

PROOF. We view the sets A and A^n as discrete probability spaces equipped with their uniform probability measures. Notice, in particular, that the probability space A^n is the product of n many copies of A. Hence, by (6.57) and Theorem 6.18, if $\mathcal{F}$ is a family of subsets of A^n with $|\mathcal{F}| = \ell$, then there exists an interval $I \subseteq \{l \in \mathbb{N} : l < n\}$ with $|I| = m$ such that for every $D \in \mathcal{F}$ we have

$$\mathrm{dens}\big(\{t \in A^I : |\mathrm{dens}(D_t) - \mathrm{dens}(D)| \leqslant \sigma\}\big) \geqslant 1 - \sigma \overset{(6.56)}{\geqslant} 1 - k^{-m}/2$$

which implies, of course, that $|\mathrm{dens}(D_t) - \mathrm{dens}(D)| \leqslant \sigma$ for every $t \in A^I$. Since $\sigma \leqslant \varepsilon$ we conclude that the estimate in (6.58) is satisfied and the proof of Lemma 6.22 is completed. □

6.3.3. Carlson–Simpson spaces. We proceed to discuss the analogue of Lemma 6.22 in the context of the Carlson–Simpson theorem. To this end, we need the following definition.

DEFINITION 6.23. *Let A be a finite alphabet with $|A| \geqslant 2$ and $\mathcal{F}$ a family of subsets of $A^{<\mathbb{N}}$. Also let $0 < \varepsilon \leqslant 1$ and J a nonempty finite subset of $\mathbb{N}$. We say that the family $\mathcal{F}$ is* (ε, J)-regular *provided that for every $D \in \mathcal{F}$, every $n \in J$, every (possibly empty) $I \subseteq \{j \in J : j < n\}$ and every $t \in A^I$ we have*

$$|\mathrm{dens}\big(\{s \in A^{\{l\in\mathbb{N}:\, l<n\}\setminus I} : (t,s) \in D\}\big) - \mathrm{dens}_{A^n}(D)| \leqslant \varepsilon. \tag{6.59}$$

Notice that for every $t \in A^I$ the set $\{s \in A^{\{l\in\mathbb{N}:\, l<n\}\setminus I} : (t,s) \in D\}$ is just the section of the set $D \cap A^n$ at t. Therefore, Definition 6.23 guarantees that for every $D \in \mathcal{F}$, every $n \in J$ and every $I \subseteq \{j \in J : j < n\}$ the density of the sections of $D \cap A^n$ along elements of A^I are essentially equal to the density of $D \cap A^n$.

We also need to introduce some numerical invariants. Specifically, let k, ℓ, m be positive integers with $k \geqslant 2$ and $0 < \varepsilon \leqslant 1$. We define $h\colon \mathbb{N} \to \mathbb{N}$ recursively by the rule

$$\begin{cases} h(0) = h(1) = 1, \\ h(i+1) = \mathrm{RegSp}(k, \ell, h(i), \varepsilon) \end{cases} \tag{6.60}$$

where $\mathrm{RegSp}(k, \ell, h(i), \varepsilon)$ is as in (6.57), and we set

$$\mathrm{RegCS}(k, \ell, m, \varepsilon) = h(m). \tag{6.61}$$

The following lemma is the main result of this subsection.

LEMMA 6.24. *Let k, ℓ, m be positive integers with $k \geqslant 2$ and $0 < \varepsilon \leqslant 1$. Also let A be an alphabet with $|A| = k$ and let J_0 be a finite subset of $\mathbb{N}$ with $|J_0| \geqslant \mathrm{RegCS}(k, \ell, m, \varepsilon)$. If $\mathcal{F}$ a family of subsets of $A^{<\mathbb{N}}$ with $|\mathcal{F}| = \ell$, then there exists a subset J of J_0 with $|J| = m$ such that the family $\mathcal{F}$ is (ε, J)-regular.*

The proof of Lemma 6.24 is based on the following refinement of Lemma 6.22.

SUBLEMMA 6.25. *Let k, ℓ, r be positive integers with $k \geqslant 2$ and $0 < \varepsilon \leqslant 1$. Also let A be an alphabet with $|A| = k$ and let J be a finite subset of $\mathbb{N}$ with $|J| \geqslant \mathrm{RegSp}(k, \ell, r, \varepsilon)$. We set $p = \max(J)$. If $\mathcal{F}$ is a family of subsets of A^p with*

$|\mathcal{F}| = \ell$, then there exists a subset J' of $J \setminus \{p\}$ with $|J'| = r$ such that for every $D \in \mathcal{F}$, every subset I of J' and every $t \in A^I$ we have

$$|\mathrm{dens}\big(\{s \in A^{\{l\in\mathbb{N}\colon l<p\}\setminus I} : (t,s) \in D\}\big) - \mathrm{dens}_{A^p}(D)| \leqslant \varepsilon. \tag{6.62}$$

Proof. First, recall that $\mathrm{RegSp}(k,\ell,r,\varepsilon) = F\big(\mathrm{ConcProd}(\ell,\sigma,F)\big) + 1$ where $\sigma = \min\{\varepsilon, k^{-r}/2\}$ and $F\colon \mathbb{N} \to \mathbb{N}$ is defined by $F(i) = i + r$ for every $i \in \mathbb{N}$. Next let $n = |J|$ and observe that

$$n \geqslant \mathrm{RegSp}(k,\ell,r,\varepsilon) = F\big(\mathrm{ConcProd}\big(\ell,\sigma,F\big)\big) + 1. \tag{6.63}$$

In particular, we have $n \geqslant 2$. Let $j_1 < \cdots < j_n$ be the increasing enumeration of J and set $J^* = J \setminus \{p\} = \{j_1, \dots, j_{n-1}\}$. We view the sets $A^p, A^{\{j_i\}}, \dots, A^{\{j_{n-1}\}}$ and $A^{\{l\in\mathbb{N}\colon l<p\}\setminus J^*}$ as discrete probability spaces equipped with their uniform probability measures and, as in the proof of Lemma 6.22, we observe that the probability space A^p is naturally identified with the product of the spaces $A^{\{j_i\}}, \dots, A^{\{j_{n-1}\}}$ and $A^{\{l\in\mathbb{N}\colon l<p\}\setminus J^*}$. By Theorem 6.18 and (6.63), if $\mathcal{F}$ is a family of subsets of A^p with $|\mathcal{F}| = \ell$, then there exists $J' \subseteq \{j_1, \dots, j_{n-1}\} = J \setminus \{p\}$ with $|J'| = r$ such that for every $D \in \mathcal{F}$ and every nonempty $I \subseteq J'$, the set of all $t \in A^I$ satisfying

$$|\mathrm{dens}\big(\{s \in A^{\{l\in\mathbb{N}\colon l<p\}\setminus I} : (t,s) \in D\}\big) - \mathrm{dens}_{A^p}(D)| \leqslant \sigma.$$

has density at least $1-\sigma$. Using the fact that $\sigma = \min\{\varepsilon, k^{-r}/2\}$, we conclude that for every $D \in \mathcal{F}$, every nonempty $I \subseteq J'$ and every $t \in A^I$ we have

$$|\mathrm{dens}\big(\{s \in A^{\{l\in\mathbb{N}\colon l<p\}\setminus I} : (t,s) \in D\}\big) - \mathrm{dens}_{A^p}(D)| \leqslant \varepsilon.$$

Since the estimate in (6.62) is automatically satisfied if I is empty, the proof of Sublemma 6.25 is completed. □

We are ready to give the proof of Lemma 6.24.

Proof of Lemma 6.24. Clearly, we may assume that $m \geqslant 2$. We fix a family $\mathcal{F}$ of subsets of $A^{<\mathbb{N}}$ with $|\mathcal{F}| = \ell$. Recursively and using Sublemma 6.25, we select a finite sequence $J_1, \dots, J_{m-1}$ of subsets of J_0 such that for every $i \in [m-1]$ the following are satisfied.

- (C1) We have $|J_i| = h(m-i)$ where h is as in (6.60).
- (C2) The set J_i is a subset of $J_{i-1} \setminus \{\max(J_{i-1})\}$.
- (C3) For every $D \in \mathcal{F}$, every subset I of J_i and every $t \in A^I$ we have

$$|\mathrm{dens}\big(\{s \in A^{\{l\in\mathbb{N}\colon l<p_{i-1}\}\setminus I} : (t,s) \in D\}\big) - \mathrm{dens}_{A^{p_{i-1}}}(D)| \leqslant \varepsilon$$

where $p_{i-1} = \max(J_{i-1})$.

We set $J = \big\{\max(J_{m-1}), \dots, \max(J_0)\big\}$. Using (C2) and (C3), we see that J is a subset of J_0 with $|J| = m$ and such that the family $\mathcal{F}$ is (ε, J)-regular. The proof of Lemma 6.24 is thus completed. □

6.4. Notes and remarks

6.4.1. As we have already mentioned, Theorem 6.5 is due to Tao [**Tao1, Tao2**]. His approach, however, is somewhat different since he works with σ-algebras instead of k-semirings. Our presentation follows [**DKKa**].

The idea to obtain uniformity estimates with respect to an arbitrary growth function appeared first in [**AFKS**] and has been also considered by several other authors (see, e.g., [**LS, RSc1**]). This particular feature of Theorem 6.5 is essential when one needs to iterate this structural decomposition.

6.4.2. A weaker version of Theorem 6.14, restricted to bipartite graphs, was first introduced in [**Sz1**] where it was used as a tool in the proof of Szemerédi's theorem on arithmetic progressions. The current form of Szemerédi's regularity lemma was obtained somewhat later in [**Sz2**].

We also note that the proof of Theorem 6.14 that we presented yields that the numbers $K_{\mathrm{Sz}}(\varepsilon, m)$ are upper bounded by a primitive recursive function belonging to the class $\mathcal{E}^4$. It is possible to obtain slightly better estimates by proceeding with a more direct—but still probabilistic—argument (see, e.g., [**ASp**]). We point out, however, that this tower-type dependence is actually necessary. Specifically, in [**Go2**] examples were given of graphs for which any equitable partition satisfying the second part of Theorem 6.14 has cardinality at least γ, where γ is a tower of twos of height proportional to $\varepsilon^{-1/16}$.

6.4.3. Theorem 6.18 is an abstract version of Lemma 3.2 in [**DKT3**]. Lemmas 6.22 and 6.24 are taken from [**DKT3**].

6.4.4. We remark that Theorems 6.5 and 6.18 can be extended to finite families of random variables in L_p for any $p > 1$. (Of course, the corresponding bounds will also depend upon the choice p.) These extensions are based on inequalities for martingale difference sequences in L_p spaces (see [**DKKa**] and [**DKT5**] for details).

CHAPTER 7

The removal lemma

The *removal lemma* is a powerful result with several consequences in Ramsey theory. It originates from the work of Ruzsa and Szemerédi in [**RS**], though its full combinatorial strength was obtained much later by Gowers [**Go5**] and, independently, by Nagle, Rödl, Schacht and Skokan [**NRS, RSk**]. However, as in the case of Szemerédi's regularity lemma, it also has a probabilistic formulation which refers to the following natural measure-theoretic structures.

Definition 7.1. *A* hypergraph system *is a triple*

$$\mathscr{H} = (n, \langle(\Omega_i, \Sigma_i, \mu_i) : i \in [n]\rangle, \mathcal{H}) \tag{7.1}$$

where n is a positive integer, $\langle(\Omega_i, \Sigma_i, \mu_i) : i \in [n]\rangle$ is a finite sequence of probability spaces and $\mathcal{H}$ is a hypergraph on $[n]$. If the hypergraph $\mathcal{H}$ is r-uniform, then $\mathscr{H}$ will be called an r-uniform hypergraph system.

For every hypergraph system $\mathscr{H} = (n, \langle(\Omega_i, \Sigma_i, \mu_i) : i \in [n]\rangle, \mathcal{H})$ by $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ we shall denote the product of the spaces $\langle(\Omega_i, \Sigma_i, \mu_i) : i \in [n]\rangle$. Moreover, as in Example 6.1, for every nonempty $e \subseteq [n]$ let $(\boldsymbol{\Omega}_e, \boldsymbol{\Sigma}_e, \boldsymbol{\mu}_e)$ be the product of the spaces $\langle(\Omega_i, \Sigma_i, \mu_i) : i \in e\rangle$ and recall that the σ-algebra $\boldsymbol{\Sigma}_e$ can be "lifted" to the full product $\boldsymbol{\Omega}$ via the formula

$$\mathcal{B}_e = \big\{\pi_e^{-1}(\mathbf{A}) : \mathbf{A} \in \boldsymbol{\Sigma}_e\big\} \tag{7.2}$$

where $\pi_e \colon \boldsymbol{\Omega} \to \boldsymbol{\Omega}_e$ is the natural projection. By convention, we set $\mathcal{B}_\emptyset = \{\emptyset, \boldsymbol{\Omega}\}$. The following theorem is the main result of this chapter and is due to Tao [**Tao1**].

Theorem 7.2. *For every $n, r \in \mathbb{N}$ with $n \geqslant r \geqslant 2$ and every $0 < \varepsilon \leqslant 1$ there exist a strictly positive constant $\delta(n, r, \varepsilon)$ and a positive integer $K(n, r, \varepsilon)$ with the following property. Let $\mathscr{H} = (n, \langle(\Omega_i, \Sigma_i, \mu_i) : i \in [n]\rangle, \mathcal{H})$ be an r-uniform hypergraph system and for every $e \in \mathcal{H}$ let $E_e \in \mathcal{B}_e$ such that*

$$\boldsymbol{\mu}\Big(\bigcap_{e \in \mathcal{H}} E_e\Big) \leqslant \delta(n, r, \varepsilon). \tag{7.3}$$

Then for every $e \in \mathcal{H}$ there exists $F_e \in \mathcal{B}_e$ with

$$\boldsymbol{\mu}(E_e \setminus F_e) \leqslant \varepsilon \quad \text{and} \quad \bigcap_{e \in \mathcal{H}} F_e = \emptyset. \tag{7.4}$$

Moreover, there exists a collection $\langle \mathcal{P}_{e'} : e' \subseteq e \text{ for some } e \in \mathcal{H}\rangle$ of partitions of $\boldsymbol{\Omega}$ such that: (i) *$\mathcal{P}_{e'} \subseteq \mathcal{B}_{e'}$ and $|\mathcal{P}_{e'}| \leqslant K(n, r, \varepsilon)$ for every $e' \subseteq e \in \mathcal{H}$, and* (ii) *for every $e \in \mathcal{H}$ the set F_e belongs to the algebra generated by the family $\bigcup_{e' \subsetneq e} \mathcal{P}_{e'}$.*

The proof of Theorem 7.2 is effective and follows a method first implemented by Ruzsa and Szemerédi [**RS**]. It consists of two parts. The first part is a regularity lemma for hypergraph systems which follows from a multidimensional version of Theorem 6.5. The second part is a "counting lemma" which enables us to estimate the probability of various events similar to those that appear in (7.3). These preparatory steps are presented in Sections 7.1, 7.2 and 7.3. The proof of Theorem 7.2 is completed in Section 7.4 while in Section 7.5 we discuss applications.

7.1. A multidimensional version of Theorem 6.5

Let Ω be a nonempty set. As in Appendix E, for every finite partition $\mathcal{P}$ of Ω by $\mathcal{A}_{\mathcal{P}}$ we shall denote the algebra on Ω generated by $\mathcal{P}$. Moreover, for every finite tuple $\boldsymbol{\mathcal{C}} = (\mathcal{C}_1, \dots, \mathcal{C}_d)$ of families of subsets of Ω and every nonempty $I \subseteq [d]$ set

$$\boldsymbol{\mathcal{C}}_I = \Big\{ \bigcap_{i\in I} C_i : C_i \in \mathcal{C}_i \text{ for every } i \in I \Big\}. \tag{7.5}$$

Note that, by Lemma 6.3, if $\boldsymbol{\mathcal{S}} = (\mathcal{S}_1, \dots, \mathcal{S}_d)$ consists of k-semirings on Ω (where k is a positive integer), then $\boldsymbol{\mathcal{S}}_I$ is a $(k \cdot |I|)$-semiring on Ω. On the other hand, if $\boldsymbol{\mathcal{P}} = (\mathcal{P}_1, \dots, \mathcal{P}_d)$ is a tuple of finite partitions of Ω with $\mathcal{P}_i \subseteq \mathcal{S}_i$ for every $i \in [d]$, then the family $\boldsymbol{\mathcal{P}}_I$ is also a finite partition of Ω which is contained in $\boldsymbol{\mathcal{S}}_I$.

The following theorem is a multidimensional version of Theorem 6.5 and is the main result in this section. Recall that a growth function is an increasing function $F\colon \mathbb{N} \to \mathbb{R}$ which satisfies $F(n) \geqslant n+1$ for every $n \in \mathbb{N}$.

THEOREM 7.3. *Let k, d, ℓ be positive integers, $0 < \sigma \leqslant 1$ and $F\colon \mathbb{N} \to \mathbb{R}$ a growth function. Also let (Ω, Σ, μ) be a probability space and $\boldsymbol{\mathcal{S}} = (\mathcal{S}_1, \dots, \mathcal{S}_d)$ a d-tuple of k-semirings on Ω with $\mathcal{S}_i \subseteq \Sigma$ for every $i \in [d]$. Finally, let $\mathcal{F}$ be a family in $L_2(\Omega, \Sigma, \mu)$ with $\|f\|_{L_2} \leqslant 1$ for every $f \in \mathcal{F}$ and $|\mathcal{F}| = \ell$. Then there exist*

(a) *a positive integer M with $M \leqslant \mathrm{RegSz}(k, \ell \cdot 2^d, \sigma, F)$,*
(b) *a d-tuple $\boldsymbol{\mathcal{P}} = (\mathcal{P}_1, \dots, \mathcal{P}_d)$ of finite partitions of Ω with $\mathcal{P}_i \subseteq \mathcal{S}_i$ and $|\mathcal{P}_i| \leqslant M$ for every $i \in [d]$, and*
(c) *a d-tuple $\boldsymbol{\mathcal{Q}} = (\mathcal{Q}_1, \dots, \mathcal{Q}_d)$ where $\mathcal{Q}_i$ is a finite refinement of $\mathcal{P}_i$ with $\mathcal{Q}_i \subseteq \mathcal{S}_i$ for every $i \in [d]$*

with the following property. For every $f \in \mathcal{F}$ and every nonempty subset I of $[d]$, letting $\boldsymbol{\mathcal{S}}_I$, $\boldsymbol{\mathcal{P}}_I$ and $\boldsymbol{\mathcal{Q}}_I$ be as in (7.5) and writing $f = f^I_{\mathrm{str}} + f^I_{\mathrm{err}} + f^I_{\mathrm{unf}}$ where

$$f^I_{\mathrm{str}} = \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{P}}_I}),\ f^I_{\mathrm{err}} = \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{Q}}_I}) - \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{P}}_I}) \text{ and } f^I_{\mathrm{unf}} = f - \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{Q}}_I}), \tag{7.6}$$

we have the estimates

$$\|f^I_{\mathrm{err}}\|_{L_2} \leqslant \sigma \ \text{ and } \ \|f^I_{\mathrm{unf}}\|_{\boldsymbol{\mathcal{S}}_I} \leqslant \frac{1}{F(M)} \tag{7.7}$$

where $\|\cdot\|_{\boldsymbol{\mathcal{S}}_I}$ is the uniformity norm associated with the $(k \cdot |I|)$-semiring $\boldsymbol{\mathcal{S}}_I$.

The main point in Theorem 7.3 is that one can decompose a finite family of random variables by using two collections of partitions which are implicated in the same way as the corresponding semirings. This additional coherence property is needed for the proof of Theorem 7.2.

In spite of being stronger, Theorem 7.3 follows the same strategy as the proof of Theorem 6.5. In particular, the reader is advised to review at this point the material in Subsection 6.1.3.

It is convenient to introduce the following notation. Let k, d be positive integers, Ω a nonempty set and $\boldsymbol{\mathcal{S}} = (\mathcal{S}_1, \dots, \mathcal{S}_d)$ a d-tuple of k-semirings on Ω. Also let $\boldsymbol{\mathcal{P}} = (\mathcal{P}_1, \dots, \mathcal{P}_d)$ be a d-tuple of finite partitions of Ω with $\mathcal{P}_i \subseteq \mathcal{S}_i$ for every $i \in [d]$. For every $N \in \mathbb{N}$ by $\Pi^N_{\boldsymbol{\mathcal{S}}}(\boldsymbol{\mathcal{P}})$ we shall denote the set of all d-tuples $\boldsymbol{\mathcal{Q}} = (\mathcal{Q}_1, \dots, \mathcal{Q}_d)$ such that for every $i \in [d]$ we have: (i) $\mathcal{Q}_i \subseteq \mathcal{S}_i$, (ii) $\mathcal{Q}_i$ is a refinement of $\mathcal{P}_i$, and (iii) $|\mathcal{Q}_i| \leqslant |\mathcal{P}_i|(k+1)^N$. We have the following variant of Lemma 6.7.

Lemma 7.4. *Let k, d be positive integers and $0 < \delta \leqslant 1$. Also let (Ω, Σ, μ) be a probability space, $\boldsymbol{\mathcal{S}} = (\mathcal{S}_1, \dots, \mathcal{S}_d)$ a d-tuple of k-semirings on Ω with $\mathcal{S}_i \subseteq \Sigma$ for every $i \in [d]$ and $\boldsymbol{\mathcal{Q}} = (\mathcal{Q}_1, \dots, \mathcal{Q}_d)$ a d-tuple of finite partitions of Ω with $\mathcal{Q}_i \subseteq \mathcal{S}_i$ for every $i \in [d]$. Finally, let $f \in L_2(\Omega, \Sigma, \mu)$ and let $I \subseteq [d]$ be nonempty, and assume that $\|f - \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{Q}}_I})\|_{\boldsymbol{\mathcal{S}}_I} > \delta$. Then there exists $\boldsymbol{\mathcal{R}} \in \Pi^1_{\boldsymbol{\mathcal{S}}}(\boldsymbol{\mathcal{Q}})$ such that $\|\mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{R}}_I}) - \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{Q}}_I})\|_{L_2} > \delta$.*

Proof. We select $S \in \boldsymbol{\mathcal{S}}_I$ such that

$$\Big| \int_S \big(f - \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{Q}}_I})\big)\, d\mu \Big| > \delta. \tag{7.8}$$

By (7.5), for every $i \in I$ there exists $S_i \in \mathcal{S}_i$ such that

$$S = \bigcap_{i \in I} S_i.$$

The d-tuple $\boldsymbol{\mathcal{S}} = (\mathcal{S}_1, \dots, \mathcal{S}_d)$ consists of k-semirings on Ω, and so there exists $\boldsymbol{\mathcal{R}} = (\mathcal{R}_1, \dots, \mathcal{R}_d) \in \Pi^1_{\boldsymbol{\mathcal{S}}}(\boldsymbol{\mathcal{Q}})$ such that $S_i \in \mathcal{A}_{\mathcal{R}_i}$ for every $i \in I$. Therefore, we have $S \in \mathcal{A}_{\boldsymbol{\mathcal{R}}_I}$ and, consequently,

$$\int_S \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{R}}_I})\, d\mu = \int_S f\, d\mu. \tag{7.9}$$

By (7.8), (7.9) and the monotonicity of the L_p norms, we conclude that

$$\|\mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{R}}_I}) - \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{Q}}_I})\|_{L_2} > \delta$$

and the proof of Lemma 7.4 is completed. □

The next lemma follows arguing precisely as in the proof of Lemma 6.8 and using Lemma 7.4 instead of Lemma 6.7.

Lemma 7.5. *Let k, d, ℓ be positive integers, $0 < \delta, \sigma \leqslant 1$ and set $N = \lceil \sigma^2 \delta^{-2} \ell\, 2^d \rceil$. Also let (Ω, Σ, μ) be a probability space, $\boldsymbol{\mathcal{S}} = (\mathcal{S}_1, \dots, \mathcal{S}_d)$ a d-tuple of k-semirings on Ω with $\mathcal{S}_i \subseteq \Sigma$ for every $i \in [d]$ and $\boldsymbol{\mathcal{P}} = (\mathcal{P}_1, \dots, \mathcal{P}_d)$ a d-tuple of finite partitions of Ω with $\mathcal{P}_i \subseteq \mathcal{S}_i$ for every $i \in [d]$. Finally, let $\mathcal{F}$ be a family in $L_2(\Omega, \Sigma, \mu)$ with $|\mathcal{F}| = \ell$. Then there exists $\boldsymbol{\mathcal{Q}} \in \Pi^N_{\boldsymbol{\mathcal{S}}}(\boldsymbol{\mathcal{P}})$ such that either*

(a) *$\|\mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{Q}}_I}) - \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{P}}_I})\|_{L_2} > \sigma$ for some $f \in \mathcal{F}$ and some nonempty subset I of $[d]$, or*

(b) *$\|\mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{Q}}_I}) - \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{P}}_I})\|_{L_2} \leqslant \sigma$ and $\|f - \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{Q}}_I})\|_{\boldsymbol{\mathcal{S}}_I} \leqslant \delta$ for every $f \in \mathcal{F}$ and every nonempty subset I of $[d]$.*

The last step towards the proof of Theorem 7.3 is the following analogue of Lemma 6.9. Its proof is identical to that of Lemma 6.9.

LEMMA 7.6. *Let k, d, ℓ be positive integers, $0 < \sigma \leqslant 1$ and $F\colon \mathbb{N} \to \mathbb{R}$ a growth function. Set $L = \lceil \sigma^{-2}\ell\, 2^d \rceil$ and define two sequences (N_j) and (M_j) in $\mathbb{N}$ recursively by the rule*

$$\begin{cases} N_0 = 0 \text{ and } M_0 = 1, \\ N_{j+1} = \lceil \sigma^2 F(M_j)^2 \ell\, 2^d \rceil \text{ and } M_{j+1} = M_j (k+1)^{N_{j+1}}. \end{cases} \tag{7.10}$$

Let (Ω, Σ, μ) be a probability space and $\boldsymbol{\mathcal{S}} = (\mathcal{S}_1, \dots, \mathcal{S}_d)$ a d-tuple of k-semirings on Ω with $\mathcal{S}_i \subseteq \Sigma$ for every $i \in [d]$. Also let $\mathcal{F}$ be a family in $L_2(\Omega, \Sigma, \mu)$ such that $\|f\|_{L_2} \leqslant 1$ for every $f \in \mathcal{F}$ and with $|\mathcal{F}| = \ell$. Then there exist $j \in \{0, \dots, L-1\}$, a d-tuple $\boldsymbol{\mathcal{P}} = (\mathcal{P}_1, \dots, \mathcal{P}_d)$ of partitions of Ω with $\mathcal{P}_i \subseteq \mathcal{S}_i$ and $|\mathcal{P}_i| \leqslant M_j$ for every $i \in [d]$, and $\boldsymbol{\mathcal{Q}} \in \Pi^{N_{j+1}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{\mathcal{P}})$ such that $\|\mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{Q}}_I}) - \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{P}}_I})\|_{L_2} \leqslant \sigma$ and $\|f - \mathbb{E}(f \,|\, \mathcal{A}_{\boldsymbol{\mathcal{Q}}_I})\|_{\boldsymbol{\mathcal{S}}_I} \leqslant 1/F(M_j)$ for every $f \in \mathcal{F}$ and every nonempty $I \subseteq [d]$.

We are ready to give the proof of Theorem 7.3.

PROOF OF THEOREM 7.3. Fix the positive integers k, d, ℓ, the constant σ and the growth function F. Let $j, \boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ be as in Lemma 7.6 and set $M = M_j$. We claim that $M, \boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ are as desired. Indeed, let $h\colon \mathbb{N} \to \mathbb{N}$ be the map defined in (6.7) for the parameters $k, \ell \cdot 2^d, \sigma$ and F. By (7.10), we have $M_j = h(j)$ and so

$$M = M_j \leqslant M_{\lceil \sigma^{-2}\ell\, 2^d \rceil} = h(\lceil \sigma^{-2}\ell\, 2^d \rceil) \stackrel{(6.8)}{=} \mathrm{RegSz}(k, \ell \cdot 2^d, \sigma, F).$$

Therefore, by Lemma 7.6, we see that $M, \boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ satisfy the requirements of the theorem. Finally, notice the estimate in (7.7) is an immediate consequence of (7.6) and Lemma 7.6. The proof of Theorem 7.3 is completed. □

7.2. A regularity lemma for hypergraph systems

In this section we will present the first main step of the proof of Theorem 7.2. Specifically, given a uniform hypergraph system $\mathscr{H}$ and a finite collection $\mathcal{C}$ of measurable events of its product space $\boldsymbol{\Omega}$, we will produce a sequence of finite partitions which can be considered as "approximations" of the given events. These partitions are contained in gradually smaller σ-algebras (hence, we improve upon the measurability of the members of $\mathcal{C}$), while at the same time we keep strong control on the error terms of the "approximations".

We proceed to the details. For every hypergraph $\mathcal{G}$ and every $e \in \mathcal{G}$ let

$$\partial e = \{e' \subseteq e : |e'| = |e| - 1\} \tag{7.11}$$

and set

$$\partial \mathcal{G} = \{e' : e' \in \partial e \text{ for some } e \in \mathcal{G}\}. \tag{7.12}$$

Clearly, if $\mathcal{G}$ is k-uniform for some $k \geqslant 1$, then $\partial\mathcal{G}$ is $(k-1)$-uniform. Next, for every r-uniform hypergraph system $\mathscr{H} = (n, \langle (\Omega_i, \Sigma_i, \mu_i) : i \in [n] \rangle, \mathcal{H})$ we define a finite sequence $\mathcal{H}_0, \dots, \mathcal{H}_r$ of hypergraphs on $[n]$ recursively by the rule

$$\mathcal{H}_r = \mathcal{H} \ \text{ and } \ \mathcal{H}_j = \partial \mathcal{H}_{j+1}. \tag{7.13}$$

Notice that $\mathcal{H}_j$ is j-uniform for every $j \in \{0, \dots, r\}$. Moreover, if $r \geqslant 2$, then as in Example 6.1 for every $j \in \{2, \dots, r\}$ and every $e \in \mathcal{H}_j$ let

$$\mathcal{S}_{\partial e} = \Big\{ \bigcap_{e' \in \partial e} A_{e'} : A_{e'} \in \mathcal{B}_{e'} \text{ for every } e' \in \partial e \Big\} \tag{7.14}$$

and recall that $\mathcal{S}_{\partial e}$ is a j-semiring on $\mathbf{\Omega}$. Note that $\mathcal{S}_{\partial e} \subseteq \mathcal{B}_e$.

We also need to introduce some numerical invariants. Let n be an integer with $n \geqslant 2$ and let $F : \mathbb{N} \to \mathbb{R}$ be a growth function. First, for every $m \in \mathbb{N}$ we define a function $G_{m,F} : \mathbb{N} \to \mathbb{R}$ by setting $G_{m,F}(x) = F(F(m)+x)$ for every $x \in \mathbb{N}$. Notice that $G_{m,F}$ is a growth function. Next, for every $r \in \{2, \dots, n\}$ and every $0 < \sigma \leqslant 1$ we define a growth function $\phi_{n,r,\sigma,F} : \mathbb{N} \to \mathbb{R}$ by

$$\phi_{n,r,\sigma,F}(m) = F(m) + \mathrm{RegSz}\big(1, m \cdot n^r \cdot 2^{n^{r-1}}, \sigma, G_{m,F}\big) \tag{7.15}$$

where $\mathrm{RegSz}(1, m \cdot n^r \cdot 2^{n^{r-1}}, \sigma, G_{m,F})$ is as in (6.8). Finally, we define two finite sequences of growth functions $(F_{n,r})_{r=2}^n$ and $(\Phi_{n,r,F})_{r=2}^n$ recursively by the rule

$$\begin{cases} F_{n,2}(m) = F^{(2)}(m) \text{ and } \Phi_{n,2,F}(m) = \phi_{n,2,1/F(m),F_{n,2}}(m), \\ F_{n,r}(m) = F^{(2)}\big(\Phi_{n,r-1,F}(m)\big), \\ \Phi_{n,r,F}(m) = \Phi_{n,r-1,F}\big(\phi_{n,r,1/F(m),F_{n,r}}(m)\big). \end{cases} \tag{7.16}$$

We have the following lemma.

LEMMA 7.7. *Let $n, r \in \mathbb{N}$ with $n \geqslant r \geqslant 2$ and $F : \mathbb{N} \to \mathbb{R}$ a growth function. Also let $\mathscr{H} = (n, \langle(\Omega_i, \Sigma_i, \mu_i) : i \in [n]\rangle, \mathcal{H})$ be an r-uniform hypergraph system and let $\mathcal{H}_0, \dots, \mathcal{H}_r$ be as in* (7.13). *Finally, let M_r be a positive integer and for every $e \in \mathcal{H}_r$ let $\mathcal{P}_e$ be a partition of $\mathbf{\Omega}$ with $\mathcal{P}_e \subseteq \mathcal{B}_e$ and $|\mathcal{P}_e| \leqslant M_r$. Then there exist:* (i) *a finite sequence $(M_j)_{j=1}^{r-1}$ of positive integers with*

$$M_r \leqslant F(M_r) \leqslant M_{r-1} \leqslant F(M_{r-1}) \leqslant \cdots \leqslant M_1 \leqslant \Phi_{n,r,F}(M_r), \tag{7.17}$$

(ii) *for every $j \in [r-1]$ and every $e' \in \mathcal{H}_j$ a partition $\mathcal{P}_{e'}$ of $\mathbf{\Omega}$ with $\mathcal{P}_{e'} \subseteq \mathcal{B}_{e'}$ and $|\mathcal{P}_{e'}| \leqslant M_j$, and* (iii) *for every $j \in [r-1]$ and every $e' \in \mathcal{H}_j$ a finite refinement $\mathcal{Q}_{e'}$ of $\mathcal{P}_{e'}$ with $\mathcal{Q}_{e'} \subseteq \mathcal{B}_{e'}$, such that the following hold. Fix $j \in \{2, \dots, r\}$ and $e \in \mathcal{H}_j$. Let $\mathcal{S}_{\partial e}$ be the j-semiring defined in* (7.14) *and consider the finite partitions*

$$\mathcal{P}_{\partial e} = \Big\{ \bigcap_{e' \in \partial e} A_{e'} : A_{e'} \in \mathcal{P}_{e'} \textit{ for every } e' \in \partial e \Big\} \tag{7.18}$$

and

$$\mathcal{Q}_{\partial e} = \Big\{ \bigcap_{e' \in \partial e} A_{e'} : A_{e'} \in \mathcal{Q}_{e'} \textit{ for every } e' \in \partial e \Big\}. \tag{7.19}$$

Then for every $A \in \mathcal{P}_e$, writing $\mathbf{1}_A = \mathrm{s}_A + \mathrm{b}_A + \mathrm{u}_A$ with

$$\mathrm{s}_A = \mathbb{E}(\mathbf{1}_A \,|\, \mathcal{A}_{\mathcal{P}_{\partial e}}), \quad \mathrm{b}_A = \mathbb{E}(\mathbf{1}_A \,|\, \mathcal{A}_{\mathcal{Q}_{\partial e}}) - \mathbb{E}(\mathbf{1}_A \,|\, \mathcal{A}_{\mathcal{P}_{\partial e}}) \\ \textit{and} \;\; \mathrm{u}_A = \mathbf{1}_A - \mathbb{E}(\mathbf{1}_A \,|\, \mathcal{A}_{\mathcal{Q}_{\partial e}}), \tag{7.20}$$

we have the estimates

$$\|\mathrm{b}_A\|_{L_2} \leqslant \frac{1}{F(M_j)} \;\; \textit{and} \;\; \|\mathrm{u}_A\|_{\mathcal{S}_{\partial e}} \leqslant \frac{1}{F(M_0)} \tag{7.21}$$

where $M_0 = F(M_1)$.

PROOF. By induction on r. The initial case "$r=2$" is identical to the general one, and so let $r \geqslant 3$ and assume that the result has been proved up to $r-1$. Let n be a positive integer with $n \geqslant r$ and fix an r-uniform hypergraph system $\mathscr{H} = (n, \langle(\Omega_i, \Sigma_i, \mu_i) : i \in [n]\rangle, \mathcal{H})$. Also let F be a growth function, M_r a positive integer and for every $e \in \mathcal{H}_r$ let $\mathcal{P}_e$ be a partition of $\boldsymbol{\Omega}$ with $\mathcal{P}_e \subseteq \mathcal{B}_e$ and $|\mathcal{P}_e| \leqslant M_r$.

Set $\mathcal{F} = \{\mathbf{1}_A : A \in \mathcal{P}_e \text{ for some } e \in \mathcal{H}_r\}$ and observe that $\mathcal{F}$ is a family in $L_2(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ with $\|f\|_{L_2} \leqslant 1$ for every $f \in \mathcal{F}$ and $|\mathcal{F}| \leqslant M_r \cdot |\mathcal{H}_r| \leqslant M_r \cdot n^r$. Next, let $\boldsymbol{\mathcal{B}} = \langle \mathcal{B}_{e'} : e' \in \mathcal{H}_{r-1}\rangle$ and notice that $\boldsymbol{\mathcal{B}}$ is a collection of 1-semirings on $\boldsymbol{\Omega}$; moreover, $|\mathcal{H}_{r-1}| \leqslant n^{r-1}$. Finally, let $F_{n,r}$ be as in (7.16) and recall that $G_{M_r, F_{n,r}}$ stands for the growth function defined by $G_{M_r,F_{n,r}}(x) = F_{n,r}(F_{n,r}(M_r) + x)$. We apply Theorem 7.3 for $\sigma = 1/F(M_r)$, the growth function $G_{M_r,F_{n,r}}$, the collection $\boldsymbol{\mathcal{B}}$ and the family $\mathcal{F}$ and we obtain: (i) a positive integer M with

$$M \leqslant \mathrm{RegSz}\big(1, M_r \cdot n^r \cdot 2^{n^{r-1}}, 1/F(M_r), G_{M_r,F_{n,r}}\big), \tag{7.22}$$

(ii) two collections $\langle \mathcal{P}_{e'} : e' \in \mathcal{H}_{r-1}\rangle$ and $\langle \mathcal{Q}_{e'} : e' \in \mathcal{H}_{r-1}\rangle$ of finite partitions of $\boldsymbol{\Omega}$, and (iii) for every $f \in \mathcal{F}$ and every nonempty $I \subseteq \mathcal{H}_{r-1}$ a decomposition $f = f^I_{\mathrm{str}} + f^I_{\mathrm{err}} + f^I_{\mathrm{unf}}$ as described Theorem 7.3. Note that if $e \in \mathcal{H}_r$ and $A \in \mathcal{P}_e$, then the family $\boldsymbol{\mathcal{B}}_{\partial e}$ defined in (7.5) for $I = \partial e$ coincides with the r-semiring $\mathcal{S}_{\partial e}$ defined in (7.14) and, moreover, $(\mathbf{1}_A)^{\partial e}_{\mathrm{str}} = \mathrm{s}_A$, $(\mathbf{1}_A)^{\partial e}_{\mathrm{err}} = \mathrm{b}_A$ and $(\mathbf{1}_A)^{\partial e}_{\mathrm{unf}} = \mathrm{u}_A$. Set

$$M_{r-1} = F_{n,r}(M_r) + M. \tag{7.23}$$

Since $G_{M_r,F_{n,r}}(M) = F_{n,r}(M_{r-1})$, it follows from the previous discussion and (7.7) that for every $e \in \mathcal{H}_r$ and every $A \in \mathcal{P}_e$ we have

$$\|\mathrm{b}_A\|_{L_2} \leqslant \frac{1}{F(M_r)} \ \text{ and } \ \|\mathrm{u}_A\|_{\mathcal{S}_{\partial e}} \leqslant \frac{1}{F_{n,r}(M_{r-1})}. \tag{7.24}$$

On the other hand, by (7.15), (7.22) and (7.23), we obtain that

$$M_r \leqslant F(M_r) \leqslant M_{r-1} \leqslant \phi_{n,r,1/F(M_r),F_{n,r}}(M_r). \tag{7.25}$$

Next, we apply our inductive assumptions to the $(r-1)$-uniform hypergraph system $\mathscr{H}_{r-1} = (n, \langle(\Omega_i, \Sigma_i, \mu_i) : i \in [n]\rangle, \mathcal{H}_{r-1})$, the positive integer M_{r-1} and the collection of partitions $\langle \mathcal{P}_{e'} : e' \in \mathcal{H}_{r-1}\rangle$ and we obtain: (iv) a finite sequence $(M_j)_{j=1}^{r-2}$ of positive integers with

$$M_{r-1} \leqslant F(M_{r-1}) \leqslant \cdots \leqslant M_1 \leqslant \Phi_{n,r-1,F}(M_{r-1}), \tag{7.26}$$

(v) two collections $\langle \mathcal{P}_{e'} : e' \in \mathcal{H}_1 \cup \cdots \cup \mathcal{H}_{r-2}\rangle$ and $\langle \mathcal{Q}_{e'} : e' \in \mathcal{H}_1 \cup \cdots \cup \mathcal{H}_{r-2}\rangle$ of finite partitions of $\boldsymbol{\Omega}$, and (vi) for every $j \in \{2, \dots, r-1\}$, every $e' \in \mathcal{H}_j$ and every $A \in \mathcal{P}_{e'}$ a decomposition $\mathbf{1}_A = \mathrm{s}_A + \mathrm{b}_A + \mathrm{u}_A$ where s_A, b_A, and u_A are as in (7.20) and satisfy the estimates in (7.21). By (7.24) and (7.26), it is enough to show that $F(M_0) \leqslant F_{n,r}(M_{r-1})$ and $M_1 \leqslant \Phi_{n,r,F}(M_r)$. Indeed, since F is increasing, we have

$$F(M_0) = F^{(2)}(M_1) \overset{(7.26)}{\leqslant} F^{(2)}\big(\Phi_{n,r-1,F}(M_{r-1})\big) \overset{(7.16)}{=} F_{n,r}(M_{r-1}).$$

Moreover,

$$M_1 \overset{(7.26)}{\leqslant} \Phi_{n,r-1,F}(M_{r-1}) \overset{(7.25)}{\leqslant} \Phi_{n,r-1,F}\big(\phi_{n,r,1/F(M_r),F_{n,r}}(M_r)\big) \overset{(7.16)}{=} \Phi_{n,r,F}(M_r)$$

and the proof of Lemma 7.7 is completed. $\square$

7.3. A counting lemma for hypergraph systems

First we introduce some terminology and some pieces of notation. We say that a hypergraph $\mathcal{G}$ is *closed under set inclusion* if for every $g \in \mathcal{G}$ and every $g' \subseteq g$ we have $g' \in \mathcal{G}$. Moreover, we define the *downwards closure* of $\mathcal{G}$ by the rule

$$\widehat{\mathcal{G}} = \{g' : g' \subseteq g \text{ for some } g \in \mathcal{G}\}. \tag{7.27}$$

That is, $\widehat{\mathcal{G}}$ is the smallest hypergraph containing $\mathcal{G}$ and closed under set inclusion.

Now let $n, r \in \mathbb{N}$ with $n \geqslant r \geqslant 2$ and $\mathscr{H} = (n, \langle(\Omega_i, \Sigma_i, \mu_i) : i \in [n]\rangle, \mathcal{H})$ an r-uniform hypergraph system. Write $\widehat{\mathcal{H}} = \mathcal{H}_0 \cup \cdots \cup \mathcal{H}_r$ where $\mathcal{H}_0, \ldots, \mathcal{H}_r$ are as in (7.13), and let $\mathcal{P}_\emptyset = \{\boldsymbol{\Omega}\}$. *These data will be fixed throughout this section.*

Assume that we are given a growth function F, a positive integer M_r and for every $e \in \mathcal{H}_r$ a partition $\mathcal{P}_e$ of $\boldsymbol{\Omega}$ with $\mathcal{P}_e \subseteq \mathcal{B}_e$ and $|\mathcal{P}_e| \leqslant M_r$. Then, by applying Lemma 7.7 to F, M_r and $\langle \mathcal{P}_e : e \in \mathcal{H}_r\rangle$, we obtain: (i) a finite sequence $(M_j)_{j=1}^{r-1}$ of positive integers, (ii) a collection $\langle \mathcal{P}_{e'} : e' \in \mathcal{H}_1 \cup \cdots \cup \mathcal{H}_{r-1}\rangle$ of partitions of $\boldsymbol{\Omega}$, and (iii) for every $j \in \{2, \ldots, r\}$, every $e \in \mathcal{H}_j$ and every $A \in \mathcal{P}_e$ a decomposition $\mathbf{1}_A = \mathrm{s}_A + \mathrm{b}_A + \mathrm{u}_A$ as described in Lemma 7.7. We enlarge the collection in (ii) by attaching the initial partitions $\langle \mathcal{P}_e : e \in \mathcal{H}_r\rangle$ and $\mathcal{P}_\emptyset$, and we obtain a new collection $\langle \mathcal{P}_e : e \in \widehat{\mathcal{H}}\rangle$ which is indexed by the downwards closure of $\mathcal{H}$. This new collection in turn generates a finite partition of $\boldsymbol{\Omega}$ whose atoms are of the form

$$\bigcap_{e \in \widehat{\mathcal{H}}} A_e \tag{7.28}$$

where $A_e \in \mathcal{P}_e$ for every $e \in \widehat{\mathcal{H}}$. Our goal in this section is to estimate the size of these atoms. Note that this task is particularly easy if the decomposition in Lemma 7.7 was perfect, that is, if $\mathrm{b}_{A_e} = \mathrm{u}_{A_e} = 0$ for every $j \in \{2, \ldots, r\}$ and every $e \in \mathcal{H}_j$. Indeed, proceeding by induction and using (7.20), in this case we have

$$\boldsymbol{\mu}\Big(\bigcap_{e \in \widehat{\mathcal{H}}} A_e\Big) = \prod_{j=1}^{r} \prod_{e \in \mathcal{H}_j} \mathbb{E}\Big(\mathbf{1}_{A_e} \,\Big|\, \bigcap_{e' \in \partial e} A_{e'}\Big). \tag{7.29}$$

(This identity also follows from the proof of Lemma 7.9 below.) Although the decomposition in Lemma 7.7 does have error terms, we will see that an approximate version of (7.29) holds true provided that the atom in (7.28) is not degenerate in the sense of the following definition.

Definition 7.8. *Fix a growth function F, a positive integer M_r and for every $e \in \mathcal{H}_r$ a partition $\mathcal{P}_e$ of $\boldsymbol{\Omega}$ with $\mathcal{P}_e \subseteq \mathcal{B}_e$ and $|\mathcal{P}_e| \leqslant M_r$. Let $(M_j)_{j=1}^{r-1}$, $\langle \mathcal{P}_e : e \in \widehat{\mathcal{H}}\rangle$ and $\langle \mathrm{s}_A, \mathrm{b}_A, \mathrm{u}_A : A \in \mathcal{P}_e$ for some $e \in \mathcal{H}_2 \cup \cdots \cup \mathcal{H}_r\rangle$ be as in Lemma 7.7 when applied to F, M_r and $\langle \mathcal{P}_e : e \in \mathcal{H}_r\rangle$. Also let $\langle A_e : e \in \widehat{\mathcal{H}}\rangle$ with $A_e \in \mathcal{P}_e$ for every $e \in \widehat{\mathcal{H}}$. We say that the family $\langle A_e : e \in \widehat{\mathcal{H}}\rangle$ is* good *provided that the following conditions are satisfied.*

(C1) *For every $j \in [r]$ and every $e \in \mathcal{H}_j$ we have*

$$\mathbb{E}\Big(\mathbf{1}_{A_e} \,\Big|\, \bigcap_{e' \in \partial e} A_{e'}\Big) \geqslant \frac{1}{\log F(M_j)}. \tag{7.30}$$

(C2) *For every $j \in \{2, \dots, r\}$ and every $e \in \mathcal{H}_j$ we have*

$$\mathbb{E}\Big(\mathrm{b}^2_{A_e} \,\Big|\, \bigcap_{e' \subsetneq e} A_{e'}\Big) \leqslant \frac{1}{F(M_j)}. \tag{7.31}$$

Observe that a good family $\langle A_e : e \in \widehat{\mathcal{H}}\rangle$ does not necessarily represent an atom via formula (7.28), since in Definition 7.8 we do not demand that the intersection of the members of the family is nonempty. Note, however, that condition (C1) implies that for every $j \in \{2, \dots, r\}$ and every $e \in \mathcal{H}_j$ the constant value of the function s_{A_e} on $\bigcap_{e' \in \partial e} A_{e'}$ is at least $1/\log F(M_j)$.

Now let $\zeta \colon \mathbb{N} \to \mathbb{R}$ be defined by $\zeta(\ell) = \sup\{x^{-1/4}(\log x)^\ell : x \geqslant 1\}$ and notice that $\zeta(\ell) \leqslant (4\ell)^\ell$ for every $\ell \geqslant 1$. Moreover, for every pair m, k of positive integers and every $\varrho, \ell \in \mathbb{N}$ we define $c(m, k, \varrho, \ell)$ and $C(m, k, \varrho, \ell)$ recursively by the rules

$$\begin{cases} c(m,k,0,\ell) = c(m,k,1,\ell) = 0, \\ c(m,k,\varrho+2,0) = c(m,2k,\varrho+1,2k^{\varrho+1}), \\ c(m,k,\varrho+2,\ell+1) = c(m,k,\varrho+2,\ell) + \big(1 + c(m,k,\varrho+2,0)\big)\cdot m^{-1/4}\cdot \zeta(\ell+1) \end{cases} \tag{7.32}$$

and

$$\begin{cases} C(m,k,0,\ell) = C(m,k,1,\ell) = 0, \\ C(m,k,\varrho+2,0) = C(m,2k,\varrho+1,2k^{\varrho+1}), \\ C(m,k,\varrho+2,\ell+1) = C(m,k,\varrho+2,\ell) + C(m,k,\varrho+2,0)\cdot m^{2^k} + 1. \end{cases} \tag{7.33}$$

We isolate, for future use, the following basic properties.

(P1) For every $k \geqslant 1$ and every $\varrho, \ell \in \mathbb{N}$ we have $c(m,k,\varrho,\ell) \to 0$ as $m \to +\infty$.

(P2) For every $m \geqslant 1$ and every $\varrho \in \mathbb{N}$ we have $c(m,k,\varrho,\ell) \leqslant c(m,k',\varrho,\ell')$ and $C(m,k,\varrho,\ell) \leqslant C(m,k',\varrho,\ell')$ whenever $1 \leqslant k \leqslant k'$ and $\ell \leqslant \ell'$.

The following lemma is the second main step of the proof of Theorem 7.2.

Lemma 7.9. *Fix a growth function F, a positive integer M_r and for every $e \in \mathcal{H}_r$ a partition $\mathcal{P}_e$ of $\mathbf{\Omega}$ with $\mathcal{P}_e \subseteq \mathcal{B}_e$ and $|\mathcal{P}_e| \leqslant M_r$. Let $(M_j)_{j=1}^{r-1}$, $\langle \mathcal{P}_e : e \in \widehat{\mathcal{H}}\rangle$ and $\langle \mathrm{s}_A, \mathrm{b}_A, \mathrm{u}_A : A \in \mathcal{P}_e$ for some $e \in \mathcal{H}_2 \cup \dots \cup \mathcal{H}_r\rangle$ be as in Lemma 7.7 when applied to F, M_r and $\langle \mathcal{P}_e : e \in \mathcal{H}_r\rangle$. Also let $\langle A_e : e \in \widehat{\mathcal{H}}\rangle$ with $A_e \in \mathcal{P}_e$ for every $e \in \widehat{\mathcal{H}}$ and assume that the family $\langle A_e : e \in \widehat{\mathcal{H}}\rangle$ is good. Set $p_\emptyset = 1$ and for every nonempty $e \in \widehat{\mathcal{H}}$ let*

$$p_e = \mathbb{E}\Big(\mathbf{1}_{A_e} \,\Big|\, \bigcap_{e' \in \partial e} A_{e'}\Big). \tag{7.34}$$

Then, setting $M_0 = F(M_1)$, we have

$$\Big|\boldsymbol{\mu}\Big(\bigcap_{e \in \widehat{\mathcal{H}}} A_e\Big) - \prod_{e \in \widehat{\mathcal{H}}} p_e\Big| \leqslant c(M_r, n, r, n^r)\cdot \prod_{e \in \widehat{\mathcal{H}}} p_e + \frac{C(M_0, n, r, n^r)}{F(M_0)}. \tag{7.35}$$

7.3.1. Proof of Lemma 7.9. As above, let $n, r \in \mathbb{N}$ with $n \geqslant r \geqslant 2$ and $\mathscr{H} = (n, \langle(\Omega_i, \Sigma_i, \mu_i) : i \in [n]\rangle, \mathcal{H})$ an r-uniform hypergraph system. Also write $\widehat{\mathcal{H}} = \mathcal{H}_0 \cup \dots \cup \mathcal{H}_r$ where $\mathcal{H}_0, \dots, \mathcal{H}_r$ are as in (7.13), and let $\mathcal{P}_\emptyset = \{\mathbf{\Omega}\}$. *Again we emphasize that these data will be fixed in what follows.*

Although for the proof of Theorem 7.2 we need precisely the estimate in (7.35), we will actually prove a slightly stronger estimate which is much more amenable to an inductive argument. To this end, we introduce the following definition.

DEFINITION 7.10. *A* hypergraph bundle over $\mathcal{H}$ *is a triple* $(k, \mathcal{G}, \varphi)$ *where* k *is a positive integer,* $\mathcal{G}$ *is a nonempty, closed under set inclusion hypergraph on* $[k]$ *and* $\varphi\colon [k] \to [n]$ *is a homomorphism from* $\mathcal{G}$ *into* $\widehat{\mathcal{H}}$*, that is, for every nonempty* $g \in \mathcal{G}$ *the restriction of* φ *on* g *is an injection and, moreover,* $\varphi(g) \in \widehat{\mathcal{H}}$*.*

With every hypergraph bundle $(k, \mathcal{G}, \varphi)$ over $\mathcal{H}$ we associate a new hypergraph system $\mathscr{G} = (k, \langle(\Omega'_j, \Sigma'_j, \mu'_j) : j \in [k]\rangle, \mathcal{G})$ where

$$(\Omega'_j, \Sigma'_j, \mu'_j) = (\Omega_{\varphi(j)}, \Sigma_{\varphi(j)}, \mu_{\varphi(j)}) \tag{7.36}$$

for every $j \in [k]$. Attached to the hypergraph system $\mathscr{G}$, we have the product $(\boldsymbol{\Omega}', \boldsymbol{\Sigma}', \boldsymbol{\mu}')$ of the spaces $\langle(\Omega'_j, \Sigma'_j, \mu'_j) : j \in [k]\rangle$ as well as the product $(\boldsymbol{\Omega}'_g, \boldsymbol{\Sigma}'_g, \boldsymbol{\mu}'_g)$ of the spaces $\langle(\Omega'_j, \Sigma'_j, \mu'_j) : j \in g\rangle$ for every nonempty $g \subseteq [k]$. Recall that the σ-algebra $\boldsymbol{\Sigma}'_g$ can be "lifted" to the full product $\boldsymbol{\Omega}'$ via the natural projection $\pi'_g\colon \boldsymbol{\Omega}' \to \boldsymbol{\Omega}'_g$. Specifically, for every nonempty $g \subseteq [k]$ let

$$\mathcal{B}'_g = \big\{(\pi'_g)^{-1}(\mathbf{A}') : \mathbf{A}' \in \boldsymbol{\Sigma}'_g\big\} \tag{7.37}$$

and note that $\mathcal{B}'_g$ is a sub-σ-algebra of $\boldsymbol{\Sigma}'$. Next, let $g \in \mathcal{G}$ be nonempty and define

$$\mathrm{I}_g\colon (\boldsymbol{\Omega}'_g, \boldsymbol{\Sigma}'_g, \boldsymbol{\mu}'_g) \to (\boldsymbol{\Omega}_{\varphi(g)}, \boldsymbol{\Sigma}_{\varphi(g)}, \boldsymbol{\mu}_{\varphi(g)}) \tag{7.38}$$

by setting $\mathrm{I}_g\big((\omega'_j)_{j\in g}\big) = (\omega_i)_{i\in\varphi(g)}$ where $\omega_i = \omega'_j$ if $i = \varphi(j)$. Since the restriction of φ on g is an injection, by (7.36), we see that the map I_g is an isomorphism. That is, I_g is a bijection, both I_g and I_g^{-1} are measurable and $\boldsymbol{\mu}_{\varphi(g)}(\mathbf{A}) = \boldsymbol{\mu}'_g\big(\mathrm{I}_g^{-1}(\mathbf{A})\big)$ for every $\mathbf{A} \in \boldsymbol{\Sigma}_{\varphi(g)}$. These isomorphisms will be used to transfer information from the hypergraph system $\mathscr{H}$ to the hypergraph system $\mathscr{G}$ as follows.

Fix a nonempty $g \in \mathcal{G}$. Let $A \in \mathcal{B}_{\varphi(g)}$ be arbitrary and notice that, by (7.2), we have $A = \pi^{-1}_{\varphi(g)}\big(\pi_{\varphi(g)}(A)\big)$. Thus, setting

$$\mathbf{A} = \pi_{\varphi(g)}(A) \in \boldsymbol{\Sigma}_{\varphi(g)}, \quad \mathbf{A}' = \mathrm{I}_g^{-1}(\mathbf{A}) \in \boldsymbol{\Sigma}'_g \ \text{ and } \ A' = (\pi'_g)^{-1}(\mathbf{A}') \in \mathcal{B}'_g, \tag{7.39}$$

we see that the map $\mathcal{B}_{\varphi(g)} \ni A \mapsto A' \in \mathcal{B}'_g$ is a bijection and, moreover,

$$\boldsymbol{\mu}(A) = \boldsymbol{\mu}_{\varphi(g)}(\mathbf{A}) = \boldsymbol{\mu}'_g(\mathbf{A}') = \boldsymbol{\mu}'(A'). \tag{7.40}$$

More generally, let $f \in L_1(\boldsymbol{\Omega}, \mathcal{B}_{\varphi(g)}, \boldsymbol{\mu})$. Also let $\mathbf{f}$ be the unique random variable in $L_1(\boldsymbol{\Omega}_{\varphi(g)}, \boldsymbol{\Sigma}_{\varphi(g)}, \boldsymbol{\mu}_{\varphi(g)})$ such that $f = \mathbf{f} \circ \pi_{\varphi(g)}$ and set

$$\mathbf{f}' = \mathbf{f} \circ \mathrm{I}_g \in L_1(\boldsymbol{\Omega}'_g, \boldsymbol{\Sigma}'_g, \boldsymbol{\mu}'_g) \ \text{ and } \ f' = \mathbf{f}' \circ \pi'_g \in L_1(\boldsymbol{\Omega}', \mathcal{B}'_g, \boldsymbol{\mu}'). \tag{7.41}$$

Observe that for every $A \in \mathcal{B}_{\varphi(g)}$ we have

$$\int_A f \, d\boldsymbol{\mu} = \int_{\mathbf{A}} \mathbf{f} \, d\boldsymbol{\mu}_{\varphi(g)} = \int_{\mathbf{A}'} \mathbf{f}' \, d\boldsymbol{\mu}'_g = \int_{A'} f' \, d\boldsymbol{\mu}'. \tag{7.42}$$

Hence, the map $f \mapsto f'$ is a linear isometry from $L_1(\boldsymbol{\Omega}, \mathcal{B}_{\varphi(g)}, \boldsymbol{\mu})$ onto $L_1(\boldsymbol{\Omega}', \mathcal{B}'_g, \boldsymbol{\mu}')$.

We are now in a position to state the aforementioned variant of Lemma 7.9.

LEMMA 7.9′. *Fix a growth function F, a positive integer M_r and for every $e\in\mathcal{H}_r$ a partition $\mathcal{P}_e$ of $\boldsymbol{\Omega}$ with $\mathcal{P}_e\subseteq\mathcal{B}_e$ and $|\mathcal{P}_e|\leqslant M_r$. Let $(M_j)_{j=1}^{r-1}$, $\langle\mathcal{P}_e:e\in\widehat{\mathcal{H}}\rangle$ and $\langle \mathrm{s}_A,\mathrm{b}_A,\mathrm{u}_A : A\in\mathcal{P}_e$ for some $e\in\mathcal{H}_2\cup\dots\cup\mathcal{H}_r\rangle$ be as in Lemma 7.7 when applied to F,M_r and $\langle\mathcal{P}_e:e\in\mathcal{H}_r\rangle$. Also let $\langle A_e:e\in\widehat{\mathcal{H}}\rangle$ with $A_e\in\mathcal{P}_e$ for every $e\in\widehat{\mathcal{H}}$ and assume that the family $\langle A_e:e\in\widehat{\mathcal{H}}\rangle$ is good. Set $p_\emptyset=1$ and for every nonempty $e\in\widehat{\mathcal{H}}$ let p_e be as in (7.34).*

Finally, let $(k,\mathcal{G},\varphi)$ be a hypergraph bundle over $\mathcal{H}$. Let $r'=\max\{|g|:g\in\mathcal{G}\}$ be the order of $\mathcal{G}$ and set $\ell=|\{g\in\mathcal{G}:|g|=r'\}|$. Also set $A'_\emptyset=\boldsymbol{\Omega}'$ and for every nonempty $g\in\mathcal{G}$ let $A'_g\in\mathcal{B}'_g$ be as in (7.39) for the set $A_{\varphi(g)}\in\mathcal{P}_{\varphi(g)}\subseteq\mathcal{B}_{\varphi(g)}$. Then, setting $M_0=F(M_1)$, we have

$$\Big|\boldsymbol{\mu}'\Big(\bigcap_{g\in\mathcal{G}}A'_g\Big)-\prod_{g\in\mathcal{G}}p_{\varphi(g)}\Big|\leqslant c(M_r,k,r',\ell)\cdot\prod_{g\in\mathcal{G}}p_{\varphi(g)}+\frac{C(M_0,k,r',\ell)}{F(M_0)}. \tag{7.43}$$

It is clear that Lemma 7.9 follows by applying Lemma 7.9′ to the hypergraph bundle $(n,\widehat{\mathcal{H}},\mathrm{Id})$ over $\mathcal{H}$. We proceed to the proof.

PROOF OF LEMMA 7.9′. First observe that the cases "$r'=0$" and "$r'=1$" are straightforward. Indeed, if $r'=0$, then $\mathcal{G}$ consists only of the empty set. On the other hand, if $r'=1$, then the family $\langle A'_g:g\in\mathcal{G}\rangle$ is independent and satisfies $p_{\varphi(g)}=\boldsymbol{\mu}'(A'_g)$ for every $g\in\mathcal{G}$; these facts imply, of course, the estimate in (7.43).

We now enter the main part of the proof which proceeds by double induction. Specifically, fix $r'\in\{2,\dots,r\}$ and assume that

(A1) the estimate in (7.43) has been proved for every hypergraph bundle over $\mathcal{H}$ of order at most $r'-1$.

Next, let k,ℓ be positive integers with $r'\leqslant k$ and $1\leqslant\ell\leqslant\binom{k}{r'}$ and assume that

(A2) the estimate in (7.43) has been proved for every hypergraph bundle $(k,\mathcal{Z},\psi)$ over $\mathcal{H}$ of order at most r' and with $|\{z\in\mathcal{Z}:|z|=r'\}|\leqslant\ell-1$.

Finally, let $(k,\mathcal{G},\varphi)$ be a hypergraph bundle over $\mathcal{H}$ of order r' and satisfying $|\{g\in\mathcal{G}:|g|=r'\}|=\ell$. We need to show that (7.43) is satisfied for $(k,\mathcal{G},\varphi)$.

To this end, fix $g_0\in\mathcal{G}$ with $|g_0|=r'$. Since $A_{\varphi(g_0)}\in\mathcal{B}_{\varphi(g_0)}$, by (7.20), we see that $\mathrm{s}_{A_{\varphi(g_0)}},\mathrm{b}_{A_{\varphi(g_0)}},\mathrm{u}_{A_{\varphi(g_0)}}\in L_1(\boldsymbol{\Omega},\mathcal{B}_{\varphi(g_0)},\boldsymbol{\mu})$. Let

$$\mathrm{s}'_{g_0},\mathrm{b}'_{g_0},\mathrm{u}'_{g_0}\in L_1(\boldsymbol{\Omega}',\mathcal{B}'_{g_0},\boldsymbol{\mu}') \tag{7.44}$$

be as in (7.41) for the random variables $\mathrm{s}_{A_{\varphi(g_0)}}$, $\mathrm{b}_{A_{\varphi(g_0)}}$ and $\mathrm{u}_{A_{\varphi(g_0)}}$ respectively. Then, by Lemma 7.7, Definition 7.8, (7.41) and (7.42), the following hold.

(P3) We have $\mathbf{1}_{A'_{g_0}}=\mathrm{s}'_{g_0}+\mathrm{b}'_{g_0}+\mathrm{u}'_{g_0}$.

(P4) The function s'_{g_0} is constant on the set $\bigcap_{g\in\partial g_0}A'_g$ and equals to $p_{\varphi(g_0)}$.

(P5) The functions b'_{g_0} and u'_{g_0} are $\mathcal{B}'_{g_0}$-measurable.

(P6) If $\|\cdot\|_{\mathcal{S}'_{\partial g_0}}$ is the uniformity norm associated with the r'-semiring

$$\mathcal{S}'_{\partial g_0}=\Big\{\bigcap_{g\in\partial g_0}B'_g : B'_g\in\mathcal{B}'_g \text{ for every } g\in\partial g_0\Big\},$$

then we have $\|\mathrm{b}'_{g_0}\|_{L_2}\leqslant 1/F(M_{r'})$ and $\|\mathrm{u}'_{g_0}\|_{\mathcal{S}'_{\partial g_0}}\leqslant 1/F(M_0)$.

(P7) We have $p_{\varphi(g_0)}\geqslant 1/\log F(M_{r'})$ and $\mathbb{E}\big(|\mathrm{b}'_{g_0}|^2\mid\bigcap_{g\subsetneq g_0}A'_g\big)\leqslant 1/F(M_{r'})$.

We set

$$D = \big|\boldsymbol{\mu}'\Big(\bigcap_{g\in\mathcal{G}} A'_g\Big) - \prod_{g\in\mathcal{G}} p_{\varphi(g)}\big|.$$

By property (P3), we have

$$\mathbf{1}_{\bigcap_{g\in\mathcal{G}} A'_g} = \Big(\prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A'_g}\Big)\cdot(\mathrm{s}'_{g_0}+\mathrm{b}'_{g_0}+\mathrm{u}'_{g_0})$$

and so, setting

$$D_1 = \Big|\int\Big(\prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A'_g}\Big)\cdot \mathrm{s}'_{g_0}\, d\boldsymbol{\mu}' - \prod_{g\in\mathcal{G}} p_{\varphi(g)}\Big|,$$

$$R_1 = \int\Big(\prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A'_g}\Big)\cdot |\mathrm{b}'_{g_0}|\, d\boldsymbol{\mu}' \ \text{ and } \ R_2 = \Big|\int\Big(\prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A'_g}\Big)\cdot \mathrm{u}'_{g_0}\, d\boldsymbol{\mu}'\Big|,$$

we see that $D \leqslant D_1 + R_1 + R_2$. Thus, it suffices to estimate D_1, R_1 and R_2. We will first deal with D_1 and R_2, and then we will concentrate on R_1 which requires some more work.

Before we proceed to the details, we need to introduce some pieces of notation. For every nonempty $g, h \subseteq [k]$ with $g \subseteq h$ by $\pi'_{h,g}\colon \boldsymbol{\Omega}'_h \to \boldsymbol{\Omega}'_g$ we shall denote the natural projection map. (Notice, in particular, that $\pi'_{[k],g} = \pi'_g$ for every nonempty $g \subseteq [k]$.) Moreover, as in Subsection 6.3.1, we write $\boldsymbol{\Omega}' = \boldsymbol{\Omega}'_{[k]\setminus g_0} \times \boldsymbol{\Omega}'_{g_0}$ and for every $\boldsymbol{\omega}' \in \boldsymbol{\Omega}'$ let $\boldsymbol{\omega}' = (\mathbf{x}, \mathbf{y})$ where $\mathbf{x} = \pi'_{[k]\setminus g_0}(\boldsymbol{\omega}')$ and $\mathbf{y} = \pi'_{g_0}(\boldsymbol{\omega}')$. We will use the representation of $\boldsymbol{\Omega}'$ as the product $\boldsymbol{\Omega}'_{[k]\setminus g_0} \times \boldsymbol{\Omega}'_{g_0}$ in order to apply Fubini's theorem. In these applications, we will follow the convention in Subsection 6.3.1 and we will ignore issues related to the measurability of sections of sets and functions, since they can be easily resolved with standard arguments.

Estimation of D_1. By Definition 7.10, the hypergraph $\mathcal{G}$ is closed under set inclusion. It follows that $\partial g_0 \subseteq \mathcal{G}\setminus\{g_0\}$ which implies, by property (P4), that

$$\Big(\prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A'_g}\Big)\cdot \mathrm{s}'_{g_0} = \Big(\prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A'_g}\Big)\cdot p_{\varphi(g_0)}.$$

Consequently,

$$D_1 = \big|\boldsymbol{\mu}'\Big(\bigcap_{g\in\mathcal{G}\setminus\{g_0\}} A'_g\Big) - \prod_{g\in\mathcal{G}\setminus\{g_0\}} p_{\varphi(g)}\big|\cdot p_{\varphi(g_0)}$$

and so, by (A2), property (P2) and the fact that $0 < p_{\varphi(g_0)} \leqslant 1$, we conclude that

$$D_1 \leqslant c(M_r, k, r', \ell-1)\cdot \prod_{g\in\mathcal{G}} p_{\varphi(g)} + \frac{C(M_0, k, r', \ell-1)}{F(M_0)}. \tag{7.45}$$

Estimation of R_2. This step is based on the fact that $\|\mathrm{u}'_{g_0}\|_{\mathcal{S}'_{\partial g_0}} \leqslant 1/F(M_0)$. We will assume that $[k]\setminus g_0$ is nonempty. (If $[k] = g_0$, then the proof is similar.) By Fubini's theorem, we have

$$\int\Big(\prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{A'_g}\Big)\cdot \mathrm{u}'_{g_0}\, d\boldsymbol{\mu}' = \int\Big(\int\big(\prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{(A'_g)_{\mathbf{x}}}\big)\cdot(\mathrm{u}'_{g_0})_{\mathbf{x}}\, d\boldsymbol{\mu}'_{g_0}\Big)\, d\boldsymbol{\mu}'_{[k]\setminus g_0}$$

where $(A'_g)_{\mathbf{x}}$ and $(\mathrm{u}'_{g_0})_{\mathbf{x}}$ are the sections at $\mathbf{x}$ of A'_g and u'_{g_0} respectively. We will show that for every $\mathbf{x} \in \mathbf{\Omega}'_{[k]\setminus g_0}$ there exists a family $\langle B'_h : h \in \partial g_0\rangle$ (possibly depending on $\mathbf{x}$) with $B'_h \in \mathcal{B}'_h$ for every $h \in \partial g_0$, and such that

$$\int \Big(\prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{(A'_g)_{\mathbf{x}}}\Big) \cdot (\mathrm{u}'_{g_0})_{\mathbf{x}}\, d\boldsymbol{\mu}'_{g_0} = \int \Big(\prod_{h\in\partial g_0} \mathbf{1}_{B'_h}\Big) \cdot \mathrm{u}'_{g_0}\, d\boldsymbol{\mu}'. \tag{7.46}$$

Once this is done then, by property (P6) and taking the average over all $\mathbf{x} \in \mathbf{\Omega}'_{[k]\setminus g_0}$, we obtain that

$$R_2 \leqslant \frac{1}{F(M_0)}. \tag{7.47}$$

So, let $\mathbf{x} \in \mathbf{\Omega}'_{[k]\setminus g_0}$ be arbitrary. The selection of the family $\langle B'_h : h \in \partial g_0\rangle$ relies on the following claim.

Claim 7.11. *For every $g \in \mathcal{G} \setminus \{g_0\}$ there exist $h \in \partial g_0$ and $\mathbf{C}'_{g,h} \in \mathbf{\Sigma}'_h$ (not necessarily unique) such that $(A'_g)_{\mathbf{x}} = (\pi'_{g_0,h})^{-1}(\mathbf{C}'_{g,h})$.*

Proof of Claim 7.11. If $g = \emptyset$ or, more generally, if $g \cap g_0 = \emptyset$, then we have $(A'_g)_{\mathbf{x}} = \mathbf{\Omega}'_{g_0}$. In these cases we select an arbitrary $h \in \partial g_0$ and we set $\mathbf{C}'_{g,h} = \mathbf{\Omega}'_h$.

Next, assume that $g \subsetneq g_0$ is nonempty. By (7.39), there exists $\mathbf{A}'_g \in \mathbf{\Sigma}'_g$ such that $A'_g = (\pi'_g)^{-1}(\mathbf{A}'_g)$. Notice that $(A'_g)_{\mathbf{x}} = (\pi'_{g_0,g})^{-1}(\mathbf{A}'_g)$. Fix $h \in \partial g_0$ with $g \subseteq h$ and observe that $(\pi'_{g_0,g})^{-1}(\mathbf{A}'_g) = (\pi'_{g_0,h})^{-1}\big((\pi'_{h,g})^{-1}(\mathbf{A}'_g)\big)$. Thus, the edge h and the set $\mathbf{C}'_{g,h} \coloneqq (\pi'_{h,g})^{-1}(\mathbf{A}'_g) \in \mathbf{\Sigma}'_h$ satisfy the requirements of the claim.

Finally, assume that $g \setminus g_0$ and $g \cap g_0$ are nonempty. Write $\mathbf{\Omega}'_g = \mathbf{\Omega}'_{g\cap g_0} \times \mathbf{\Omega}'_{g\setminus g_0}$ and let $\mathbf{x}' = \pi'_{[k]\setminus g_0, g\setminus g_0}(\mathbf{x}) \in \mathbf{\Omega}'_{g\setminus g_0}$. Also let $(\mathbf{A}'_g)_{\mathbf{x}'}$ be the section of $\mathbf{A}'_g$ at $\mathbf{x}'$, and note that $(\mathbf{A}'_g)_{\mathbf{x}'} \in \mathbf{\Sigma}'_{g\cap g_0}$ and $(A'_g)_{\mathbf{x}} = (\pi'_{g_0,g\cap g_0})^{-1}\big((\mathbf{A}'_g)_{\mathbf{x}'}\big)$. On the other hand, since $g \neq g_0$ and $|g_0| = \max\{|g'| : g' \in \mathcal{G}\}$, we see that $g \cap g_0$ is a nonempty proper subset of g_0. We select $h \in \partial g_0$ with $g \cap g_0 \subseteq h$ and we set $\mathbf{C}'_{g,h} \coloneqq (\pi'_{h,g\cap g_0})^{-1}\big((\mathbf{A}'_g)_{\mathbf{x}'}\big) \in \mathbf{\Sigma}'_h$. Clearly, h and $\mathbf{C}'_{g,h}$ are as desired. The proof of Claim 7.11 is completed. □

We are ready to define the family $\langle B'_h : h \in \partial g_0\rangle$. Specifically, by Claim 7.11, for every $h \in \partial g_0$ there exists $\mathbf{C}''_h \in \mathbf{\Sigma}'_h$ such that

$$\bigcap_{g\in\mathcal{G}\setminus\{g_0\}} (A'_g)_{\mathbf{x}} = \bigcap_{h\in\partial g_0} (\pi'_{g_0,h})^{-1}(\mathbf{C}''_h). \tag{7.48}$$

Set $B'_h = (\pi'_h)^{-1}(\mathbf{C}''_h)$ and observe that $B'_h \in \mathcal{B}'_h$ and $\mathbf{1}_{(\pi'_{g_0,h})^{-1}(\mathbf{C}''_h)} \circ \pi'_{g_0} = \mathbf{1}_{B'_h}$. Moreover, by property (P5), let $\mathbf{u}'_{g_0}$ be the unique $\mathbf{\Sigma}'_{g_0}$-measurable function such that $\mathrm{u}'_{g_0} = \mathbf{u}'_{g_0} \circ \pi'_{g_0}$ and notice that $(\mathrm{u}'_{g_0})_{\mathbf{x}} = \mathbf{u}'_{g_0}$. Then we have

$$\int \Big(\prod_{g\in\mathcal{G}\setminus\{g_0\}} \mathbf{1}_{(A'_g)_{\mathbf{x}}}\Big) \cdot (\mathrm{u}'_{g_0})_{\mathbf{x}}\, d\boldsymbol{\mu}'_{g_0} \overset{(7.48)}{=} \int \Big(\prod_{h\in\partial g_0} \mathbf{1}_{(\pi'_{g_0,h})^{-1}(\mathbf{C}''_h)}\Big) \cdot \mathbf{u}'_{g_0}\, d\boldsymbol{\mu}'_{g_0} =$$

$$= \int \Big(\big(\prod_{h\in\partial g_0} \mathbf{1}_{(\pi'_{g_0,h})^{-1}(\mathbf{C}''_h)}\big) \cdot \mathbf{u}'_{g_0}\Big) \circ \pi'_{g_0}\, d\boldsymbol{\mu}' = \int \Big(\prod_{h\in\partial g_0} \mathbf{1}_{B'_h}\Big) \cdot \mathrm{u}'_{g_0}\, d\boldsymbol{\mu}'$$

and the proof of (7.46) is completed.

Estimation of R_1. We set

$$\mathcal{G}_{<g_0} = \{g \in \mathcal{G} : g \subsetneq g_0\} \ \text{ and } \ \mathcal{G}' = \{g \in \mathcal{G} \setminus \mathcal{G}_{<g_0} : |g| \leqslant r' - 1\}.$$

As before, to simplify the exposition, we will assume that $\mathcal{G}'$ is nonempty. (If $\mathcal{G}'$ is empty, then the proof is simpler since Claim 7.13 below is superfluous.) Let $A'_{<g_0} = \bigcap_{g \in \mathcal{G}_{<g_0}} A'_g$ and observe that $A'_{<g_0} \in \mathcal{B}'_{g_0}$. In particular, there exists a unique $\mathbf{A}'_{<g_0} \in \boldsymbol{\Sigma}'_{g_0}$ with $A'_{<g_0} = (\pi'_{g_0})^{-1}(\mathbf{A}'_{<g_0})$. The set $\mathbf{A}'_{<g_0}$ can also be written as the intersection of a family of events indexed by $\mathcal{G}_{<g_0}$. Indeed, we have

$$\mathbf{A}'_{<g_0} = \bigcap_{g \in \mathcal{G}_{<g_0}} \pi'_{g_0}(A'_g) = \bigcap_{g \in \mathcal{G}_{<g_0}} (\pi'_{g_0,g})^{-1}(\mathbf{A}'_g) \tag{7.49}$$

where, as above, $\mathbf{A}'_g = \pi'_g(A'_g)$ for every nonempty $g \in \mathcal{G}_{<g_0}$ and, by convention, $\mathbf{A}'_\emptyset = \boldsymbol{\Omega}'_{g_0}$. In the following claim we obtain a first estimate for R_1 by eliminating the contribution of b'_{g_0}.

Claim 7.12. *We have*

$$R_1 \leqslant \Big(\frac{\boldsymbol{\mu}'(A'_{<g_0})}{F(M_{r'})}\Big)^{1/2} \cdot \Big(\int \mathbf{1}_{\mathbf{A}'_{<g_0}} \cdot \big(\int (\prod_{g \in \mathcal{G}'} \mathbf{1}_{A'_g})_{\mathbf{y}} \, d\boldsymbol{\mu}'_{[k]\setminus g_0}\big)^2 \, d\boldsymbol{\mu}'_{g_0}\Big)^{1/2} \tag{7.50}$$

where for every $\mathbf{y} \in \boldsymbol{\Omega}'_{g_0}$ by $(\prod_{g \in \mathcal{G}'} \mathbf{1}_{A'_g})_{\mathbf{y}}$ we denote the section of $\prod_{g \in \mathcal{G}'} \mathbf{1}_{A'_g}$ at $\mathbf{y}$.

Proof of Claim 7.12. Since $\mathcal{G}_{<g_0} \cup \mathcal{G}' \subseteq \mathcal{G} \setminus \{g_0\}$ we have

$$R_1 \leqslant \int \Big(\prod_{g \in \mathcal{G}_{<g_0} \cup \mathcal{G}'} \mathbf{1}_{A'_g}\Big) \cdot |\mathrm{b}'_{g_0}| \, d\boldsymbol{\mu}'$$

and so, by Fubini's theorem,

$$R_1 \leqslant \int \Big(\int (\prod_{g \in \mathcal{G}_{<g_0} \cup \mathcal{G}'} \mathbf{1}_{A'_g})_{\mathbf{y}} \cdot (|\mathrm{b}'_{g_0}|)_{\mathbf{y}} \, d\boldsymbol{\mu}'_{[k]\setminus g_0}\Big) \, d\boldsymbol{\mu}'_{g_0}. \tag{7.51}$$

We will write (7.51) in a more manageable form. Fix $\mathbf{y} \in \boldsymbol{\Omega}'_{g_0}$ and, by property (P5), let $\mathbf{b}'_{g_0}$ be the unique $\boldsymbol{\Sigma}'_{g_0}$-measurable function such that $\mathrm{b}'_{g_0} = \mathbf{b}'_{g_0} \circ \pi'_{g_0}$. Notice that $(\mathrm{b}'_{g_0})_{\mathbf{y}}(\mathbf{x}) = \mathrm{b}'_{g_0}\big((\mathbf{x},\mathbf{y})\big) = (\mathbf{b}'_{g_0} \circ \pi'_{g_0})(\mathbf{x},\mathbf{y}) = \mathbf{b}'_{g_0}(\mathbf{y})$ for every $\mathbf{x} \in \boldsymbol{\Omega}'_{[k]\setminus g_0}$. Hence, $(\mathrm{b}'_{g_0})_{\mathbf{y}}$ is constantly equal to $\mathbf{b}'_{g_0}(\mathbf{y})$. Moreover, $(\mathbf{1}_{A'_g})_{\mathbf{y}}(\mathbf{x}) = \mathbf{1}_{(\pi'_{g_0,g})^{-1}(\mathbf{A}'_g)}(\mathbf{y})$ for every nonempty $g \in \mathcal{G}_{<g_0}$ and every $\mathbf{x} \in \boldsymbol{\Omega}'_{[k]\setminus g_0}$ and so, by (7.49),

$$(\prod_{g \in \mathcal{G}_{<g_0}} \mathbf{1}_{A'_g})_{\mathbf{y}}(\mathbf{x}) = \mathbf{1}_{\mathbf{A}'_{<g_0}}(\mathbf{y}).$$

Thus, if $G \colon \boldsymbol{\Omega}'_{g_0} \to \mathbb{R}$ is the random variable defined by the rule

$$G(\mathbf{y}) = \int (\prod_{g \in \mathcal{G}'} \mathbf{1}_{A'_g})_{\mathbf{y}} \, d\boldsymbol{\mu}'_{[k]\setminus g_0},$$

then we may write (7.51) as

$$R_1 \leqslant \int (\mathbf{1}_{\mathbf{A}'_{<g_0}} \cdot |\mathbf{b}'_{g_0}| \cdot G) \, d\boldsymbol{\mu}'_{g_0} = \int (\mathbf{1}_{\mathbf{A}'_{<g_0}} \cdot |\mathbf{b}'_{g_0}|) \cdot (\mathbf{1}_{\mathbf{A}'_{<g_0}} \cdot G) \, d\boldsymbol{\mu}'_{g_0}.$$

Therefore, by the Cauchy–Schwarz inequality, we obtain that[1]

$$R_1 \leqslant \Big(\int \mathbf{1}_{\mathbf{A}'_{<g_0}} \cdot |\mathbf{b}'_{g_0}|^2 \, d\boldsymbol{\mu}'_{g_0} \Big)^{1/2} \cdot \Big(\int \mathbf{1}_{\mathbf{A}'_{<g_0}} \cdot G^2 \, d\boldsymbol{\mu}'_{g_0} \Big)^{1/2}. \tag{7.52}$$

On the other hand, by (7.42) and property (P7), we have

$$\int \mathbf{1}_{\mathbf{A}'_{<g_0}} \cdot |\mathbf{b}'_{g_0}|^2 \, d\boldsymbol{\mu}'_{g_0} = \mathbb{E}\big(|\mathrm{b}'_{g_0}|^2 \,|\, A'_{<g_0}\big) \cdot \boldsymbol{\mu}'(A'_{<g_0}) \leqslant \frac{\boldsymbol{\mu}'(A'_{<g_0})}{F(M_{r'})}.$$

By the choice of G and the previous estimates, we conclude that (7.50) is satisfied and the proof of Claim 7.12 is completed. □

The second step of this part of the proof is based on an application of assumption (A1). To this end, we will represent the double integral in the right-hand side of (7.50) as the probability of the intersection of a family of events which correspond to a hypergraph bundle over $\mathcal{H}$. Specifically, for every $g \subseteq [k]$ let

$$i(g) = (g \cap g_0) \cup \{j + k : j \in g \setminus g_0\} \tag{7.53}$$

and set $\mathcal{W} = \mathcal{G}_{<g_0} \cup \mathcal{G}' \cup \{i(g) : g \in \mathcal{G}'\}$. Clearly, $\mathcal{W}$ is a nonempty, closed under set inclusion hypergraph on $[2k]$ of order $r' - 1$. Also let $\psi\colon [2k] \to [k]$ be defined by $\psi(j) = j$ if $j \in [k]$ and $\psi(j) = j - k$ if $j \in [2k] \setminus [k]$. Notice that for every nonempty $w \in \mathcal{W}$ the restriction of ψ on w is an injection and, moreover, for every $g \in \mathcal{G}_{<g_0} \cup \mathcal{G}'$ we have

$$\psi(g) = \psi\big(i(g)\big) = g. \tag{7.54}$$

It follows that the triple $(2k, \mathcal{W}, \varphi \circ \psi)$ is a hypergraph bundle over $\mathcal{H}$ and so it defines a new hypergraph system $\mathscr{W} = (2k, \langle (\Omega''_j, \Sigma''_j, \mu''_j) : j \in [2k] \rangle, \mathcal{W})$ where

$$(\Omega''_j, \Sigma''_j, \mu''_j) = (\Omega'_{\psi(j)}, \Sigma'_{\psi(j)}, \mu'_{\psi(j)}) = (\Omega_{\varphi(\psi(j))}, \Sigma_{\varphi(\psi(j))}, \mu_{\varphi(\psi(j))}) \tag{7.55}$$

for every $j \in [2k]$. Recall that associated with $\mathscr{W}$ we have the product $(\boldsymbol{\Omega}'', \boldsymbol{\Sigma}'', \boldsymbol{\mu}'')$ of the spaces $\langle (\Omega''_j, \Sigma''_j, \mu''_j) : j \in [2k] \rangle$ as well as the product $(\boldsymbol{\Omega}''_w, \boldsymbol{\Sigma}''_w, \boldsymbol{\mu}''_w)$ of the spaces $\langle (\Omega''_j, \Sigma''_j, \mu''_j) : j \in w \rangle$ for every nonempty $w \subseteq [2k]$. Note that, by (7.53) and (7.55), for every nonempty $g \subseteq [k]$ the following hold.

(P8) We have $(\boldsymbol{\Omega}''_g, \boldsymbol{\Sigma}''_g, \boldsymbol{\mu}''_g) = (\boldsymbol{\Omega}'_g, \boldsymbol{\Sigma}'_g, \boldsymbol{\mu}'_g)$; in particular, $\boldsymbol{\Omega}''_{[k]} = \boldsymbol{\Omega}'$ and $\boldsymbol{\Omega}''_{g_0} = \boldsymbol{\Omega}'_{g_0}$. Moreover, the spaces $(\boldsymbol{\Omega}''_{i(g)}, \boldsymbol{\Sigma}''_{i(g)}, \boldsymbol{\mu}''_{i(g)})$ and $(\boldsymbol{\Omega}'_g, \boldsymbol{\Sigma}'_g, \boldsymbol{\mu}'_g)$ are isomorphic via the bijection $\boldsymbol{\Omega}''_{i(g)} \ni (\omega''_l)_{l \in i(g)} \mapsto (\omega'_j)_{j \in g} \in \boldsymbol{\Omega}'_g$ where $\omega'_j = \omega''_l$ if $l = i(j)$.

Next, as in (7.2), for every nonempty $w \subseteq [2k]$ let $\mathcal{B}''_w = \{(\pi''_w)^{-1}(\mathbf{A}'') : \mathbf{A}'' \in \boldsymbol{\Sigma}''_w\}$ where $\pi''_w\colon \boldsymbol{\Omega}'' \to \boldsymbol{\Omega}''_w$ is the natural projection, and let $\langle A''_w : w \in \mathcal{W} \rangle$ be the family of events described in the statement of the lemma for the hypergraph bundle $(2k, \mathcal{W}, \varphi \circ \psi)$. Then observe that for every nonempty $g \in \mathcal{G}_{<g_0} \cup \mathcal{G}'$ we have

$$A''_g = (\pi''_{[k]})^{-1}(A'_g) \in \mathcal{B}''_g \tag{7.56}$$

and, respectively, for every $g \in \mathcal{G}'$

$$A''_{i(g)} = (\pi''_{i([k])})^{-1}(A'_g) \in \mathcal{B}''_{i(g)} \tag{7.57}$$

[1]The doubling of the characteristic function of $\mathbf{A}'_{<g_0}$ in (7.52) is needed in order to prevent some critical losses in (7.65) below. Similar technical maneuvers are present and in other proofs of the hypergraph removal lemma (see, e.g., [**Go5**]).

where, by (P8), we view $\pi''_{[k]}$ and $\pi''_{i([k])}$ as projections from $\mathbf{\Omega}''$ onto $\mathbf{\Omega}'$. We have the following claim.

CLAIM 7.13. *We have*

$$\boldsymbol{\mu}''\Big(\bigcap_{w\in\mathcal{W}} A''_w\Big) = \int \mathbf{1}_{\mathbf{A}'_{<g_0}} \cdot \Big(\int \big(\prod_{g\in\mathcal{G}'} \mathbf{1}_{A'_g}\big)_{\mathbf{y}}\, d\boldsymbol{\mu}'_{[k]\setminus g_0}\Big)^2\, d\boldsymbol{\mu}'_{g_0}. \tag{7.58}$$

PROOF OF CLAIM 7.13. We set $\mathcal{W}'=\mathcal{W}\setminus\mathcal{G}_{<g_0}$. Moreover, by (P8), write

$$\mathbf{\Omega}'' = \mathbf{\Omega}''_{[2k]\setminus g_0}\times\mathbf{\Omega}''_{g_0} = \mathbf{\Omega}''_{[2k]\setminus g_0}\times\mathbf{\Omega}'_{g_0}$$

and for every $\boldsymbol{\omega}''\in\mathbf{\Omega}''$ let $\boldsymbol{\omega}''=(\mathbf{x},\mathbf{y})$ where $\mathbf{x}=\pi''_{[2k]\setminus g_0}(\boldsymbol{\omega}'')$ and $\mathbf{y}=\pi''_{g_0}(\boldsymbol{\omega}'')$. By Fubini's theorem, we have

$$\begin{aligned}\boldsymbol{\mu}\Big(\bigcap_{w\in\mathcal{W}} A''_w\Big) &= \int\Big(\int\big(\prod_{w\in\mathcal{W}}\mathbf{1}_{A''_w}\big)_{\mathbf{y}}\, d\boldsymbol{\mu}''_{[2k]\setminus g_0}\Big)\, d\boldsymbol{\mu}'_{g_0}\\ &= \int\Big(\int\big(\prod_{g\in\mathcal{G}_{<g_0}}\mathbf{1}_{A''_g}\big)_{\mathbf{y}}\cdot\big(\prod_{w\in\mathcal{W}'}\mathbf{1}_{A''_w}\big)_{\mathbf{y}}\, d\boldsymbol{\mu}''_{[2k]\setminus g_0}\Big)\, d\boldsymbol{\mu}'_{g_0}. \end{aligned}\tag{7.59}$$

Since $\bigcap_{g\in\mathcal{G}_{<g_0}} A'_g = (\pi'_{g_0})^{-1}(\mathbf{A}'_{<g_0})$, by (7.56), we have $\bigcap_{g\in\mathcal{G}_{<g_0}} A''_g = (\pi''_{g_0})^{-1}(\mathbf{A}'_{<g_0})$. Thus, we may rewrite (7.59) as

$$\boldsymbol{\mu}\Big(\bigcap_{w\in\mathcal{W}} A''_w\Big) = \int \mathbf{1}_{\mathbf{A}'_{<g_0}}\cdot\Big(\int\big(\prod_{g\in\mathcal{G}'}\mathbf{1}_{A''_g}\big)_{\mathbf{y}}\cdot\big(\prod_{g\in\mathcal{G}'}\mathbf{1}_{A''_{i(g)}}\big)_{\mathbf{y}}\, d\boldsymbol{\mu}''_{[2k]\setminus g_0}\Big)\, d\boldsymbol{\mu}'_{g_0}. \tag{7.60}$$

Now fix $\mathbf{y}\in\mathbf{\Omega}''_{g_0}$. Set $I=[k]\setminus g_0$ and recall that, by (P8), we may view π''_I and $\pi''_{i(I)}$ as projections from $\mathbf{\Omega}''$ onto $\mathbf{\Omega}'_I$. Also let $g\in\mathcal{G}'$ and observe that, by (7.56) and (7.57), we have $(A''_g)_{\mathbf{y}} = (\pi''_I)^{-1}\big((A'_g)_{\mathbf{y}}\big)$ and $(A''_{i(g)})_{\mathbf{y}} = (\pi''_{i(I)})^{-1}\big((A'_g)_{\mathbf{y}}\big)$. Since I and $i(I)$ are disjoint, it follows that the sets $(\bigcap_{g\in\mathcal{G}'} A''_g)_{\mathbf{y}}$ and $(\bigcap_{g\in\mathcal{G}'} A''_{i(g)})_{\mathbf{y}}$ are independent in $(\mathbf{\Omega}''_{[2k]\setminus g_0}, \mathbf{\Sigma}''_{[2k]\setminus g_0}, \boldsymbol{\mu}''_{[2k]\setminus g_0})$ and both have measure equal to

$$\boldsymbol{\mu}'_{[k]\setminus g_0}\Big(\big(\bigcap_{g\in\mathcal{G}'} A'_g\big)_{\mathbf{y}}\Big) = \int\big(\prod_{g\in\mathcal{G}'}\mathbf{1}_{A'_g}\big)_{\mathbf{y}}\, d\boldsymbol{\mu}'_{[k]\setminus g_0}.$$

Hence, (7.58) is satisfied and the proof of Claim 7.13 is completed. □

We are ready to estimate R_1. First, we set

$$c = c(M_r, k, r', 0)\quad\text{and}\quad C = C(M_0, k, r', 0). \tag{7.61}$$

Next, observe that the triple $(k, \mathcal{G}_{<g_0}, \varphi)$ is a hypergraph bundle over $\mathcal{H}$, the order of $\mathcal{G}_{<g_0}$ is $r'-1$ and $|\{g\in\mathcal{G}_{<g_0} : |g|=r'-1\}| = |\partial g_0| = |g_0| = r'$. Hence, by assumption (A1), (7.32), (7.33) and property (P2), we have

$$\boldsymbol{\mu}'\Big(\bigcap_{g\in\mathcal{G}_{<g_0}} A'_g\Big) \leqslant (1+c)\cdot\prod_{g\in\mathcal{G}_{<g_0}} p_{\phi(g)} + \frac{C}{F(M_0)}. \tag{7.62}$$

On the other hand, as we have already mentioned, the triple $(2k, \mathcal{W}, \varphi\circ\psi)$ is a hypergraph bundle over $\mathcal{H}$ and the order of $\mathcal{W}$ is $r'-1$. Note that, by the choice of $\mathcal{W}$, we have $|\{w\in\mathcal{W} : |w|=r'-1\}| \leqslant 2\binom{k}{r'-1} \leqslant 2k^{r'-1}$. Moreover, by (7.54),

$$\prod_{w\in\mathcal{W}} p_{\varphi(\psi(w))} = \prod_{g\in\mathcal{G}_{<g_0}} p_{\phi(g)}\cdot\prod_{g\in\mathcal{G}'} p^2_{\phi(g)}.$$

Thus, using assumption (A1), (7.32), (7.33) and (P2) once again, we see that

$$\boldsymbol{\mu}''\Big(\bigcap_{w\in\mathcal{W}} A''_w\Big) \leqslant (1+c)\cdot \prod_{g\in\mathcal{G}_{<g_0}} p_{\phi(g)} \cdot \prod_{g\in\mathcal{G}'} p^2_{\phi(g)} + \frac{C}{F(M_0)}. \tag{7.63}$$

Now since $|\varphi(g)| = |g| \geqslant 1$ for every $g\in\mathcal{G}'$, by (7.17) and (7.30), we have

$$\prod_{g\in\mathcal{G}'} p_{\phi(g)} \geqslant \prod_{g\in\mathcal{G}'} \frac{1}{\log F(M_{|g|})} \geqslant (\log M_0)^{-2^k} \geqslant M_0^{-2^k}. \tag{7.64}$$

Combining (7.62)–(7.64), we obtain that

$$\boldsymbol{\mu}'\Big(\bigcap_{g\in\mathcal{G}_{<g_0}} A'_g\Big)\cdot \boldsymbol{\mu}''\Big(\bigcap_{w\in\mathcal{W}} A''_w\Big) \leqslant \Big((1+c)\cdot \prod_{g\in\mathcal{G}_{<g_0}\cup\mathcal{G}'} p_{\phi(g)} + \frac{C\cdot M_0^{2^k}}{F(M_0)}\Big)^2$$

and so, by Claims 7.12 and 7.13,

$$R_1 \leqslant F(M_{r'})^{-1/2}\cdot \Big((1+c)\cdot \prod_{g\in\mathcal{G}_{<g_0}\cup\mathcal{G}'} p_{\phi(g)} + \frac{C\cdot M_0^{2^k}}{F(M_0)}\Big).$$

This estimate is already strong enough but we need to write it in a form which is suitable for the induction. Specifically, by (7.17) and (7.30) once again, we have

$$\big(\log F(M_{r'})\big)^{\ell}\cdot\Big(\prod_{\{g\in\mathcal{G}:|g|=r'\}} p_{\varphi(g)}\Big) \geqslant 1$$

and, consequently,

$$R_1 \leqslant F(M_{r'})^{-1/2}\cdot \big(\log F(M_{r'})\big)^{\ell}\cdot (1+c)\cdot \prod_{g\in\mathcal{G}} p_{\phi(g)} + \frac{C\cdot M_0^{2^k}}{F(M_0)}. \tag{7.65}$$

Since $\zeta(\ell) = \sup\{x^{-1/4}(\log x)^{\ell} : x\geqslant 1\}$ and $M_r \leqslant M_{r'} \leqslant F(M_{r'})$, we get that

$$F(M_{r'})^{-1/2}\cdot\big(\log F(M_{r'})\big)^{\ell} = F(M_{r'})^{-1/4}\cdot \frac{\big(\log F(M_{r'})\big)^{\ell}}{F(M_{r'})^{1/4}} \leqslant M_r^{-1/4}\cdot\zeta(\ell).$$

Hence, by (7.61) and (7.65), we conclude that

$$R_1 \leqslant \big(1 + c(M_r,k,r',0)\big)\cdot M_r^{-1/4}\cdot\zeta(\ell)\cdot\prod_{g\in\mathcal{G}} p_{\phi(g)} + \frac{C(M_0,k,r',0)\cdot M_0^{2^k}}{F(M_0)}. \tag{7.66}$$

Verification of the inductive assumptions. We are ready for the last step of the argument. Specifically, by (7.45), (7.47) and (7.66), and using the definition of the numbers $c(M_r,k,r',\ell)$ and $C(M_0,k,r',\ell)$ in (7.32) and (7.33) respectively, we obtain that

$$\begin{aligned} \Big|\boldsymbol{\mu}'\Big(\bigcap_{g\in\mathcal{G}} A'_g\Big) - \prod_{g\in\mathcal{G}} p_{\varphi(g)}\Big| &\leqslant D_1 + R_1 + R_2 \\ &\leqslant c(M_r,k,r',\ell)\cdot\prod_{g\in\mathcal{G}} p_{\varphi(g)} + \frac{C(M_0,k,r',\ell)}{F(M_0)}. \end{aligned}$$

That is, the estimate in (7.43) is satisfied for the hypergraph bundle $(k,\mathcal{G},\varphi)$. This completes the proof of the general inductive step and so the entire proof of Lemma 7.9′ is completed. ∎

7.4. Proof of Theorem 7.2

Fix $n,r\in\mathbb{N}$ with $n\geqslant r\geqslant 2$ and $0<\varepsilon\leqslant 1$. For every positive integer m let $c(m,n,r,n^r)$ and $C(m,n,r,n^r)$ be as (7.32) and (7.33) respectively, and set

$$m_0(n,r)=\min\{m\geqslant 1: c(m',n,r,n^r)<1/4 \text{ for every } m'\geqslant m\}. \tag{7.67}$$

(Note that, by property (P1), $m_0(n,r)$ is well-defined.) Moreover, set

$$M_r=\max\{2,m_0(n,r),\varepsilon^{-1}\} \ \text{ and } \ F(m)=4m^{2^n}\cdot C(m,n,r,n^r)+e^{m^2\,2^{n+1}} \tag{7.68}$$

and let $F(0)=1$. Observe that the map $F\colon\mathbb{N}\to\mathbb{R}$ is a growth function. Finally, let $\Phi_{n,r,F}$ be as in (7.16) and define

$$\delta(n,r,\varepsilon)=2^{-1}\cdot F(\Phi_{n,r,F}(M_r))^{-2^n} \ \text{ and } \ K(n,r,\varepsilon)=\Phi_{n,r,F}(M_r). \tag{7.69}$$

We will show that $\delta(n,r,\varepsilon)$ and $K(n,r,\varepsilon)$ are as desired.

Indeed, let $\mathscr{H}=(n,\langle(\Omega_i,\Sigma_i,\mu_i):i\in[n]\rangle,\mathcal{H})$ be an r-uniform hypergraph system and for every $e\in\mathcal{H}$ let $E_e\in\mathcal{B}_e$ such that

$$\boldsymbol{\mu}\Big(\bigcap_{e\in\mathcal{H}}E_e\Big)\leqslant\delta(n,r,\varepsilon). \tag{7.70}$$

Clearly, we may assume that for every $e\in\mathcal{H}$ the sets E_e and $\boldsymbol{\Omega}\setminus E_e$ are nonempty. Write $\widehat{\mathcal{H}}=\mathcal{H}_0\cup\cdots\cup\mathcal{H}_r$ where $\mathcal{H}_0,\dots,\mathcal{H}_r$ are as in (7.13), and set $\mathcal{P}_\emptyset=\{\boldsymbol{\Omega}\}$. Moreover, set $\mathcal{P}_e=\{E_e,\boldsymbol{\Omega}\setminus E_e\}$ for every $e\in\mathcal{H}_r$. We apply Lemma 7.7 to F,M_r and $\langle\mathcal{P}_e:e\in\mathcal{H}_r\rangle$ and we obtain: (i) a finite sequence $(M_j)_{j=1}^{r-1}$ of positive integers, (ii) a collection $\langle\mathcal{P}_{e'}:e'\in\mathcal{H}_1\cup\cdots\cup\mathcal{H}_{r-1}\rangle$ of partitions of $\boldsymbol{\Omega}$, and (iii) for every $j\in\{2,\dots,r\}$, every $e\in\mathcal{H}_j$ and every $A\in\mathcal{P}_e$ a decomposition $\mathbf{1}_A=\mathrm{s}_A+\mathrm{b}_A+\mathrm{u}_A$ as described in Lemma 7.7. As in Section 7.3, we enlarge the collection in (ii) by attaching the initial partitions $\langle\mathcal{P}_e:e\in\mathcal{H}_r\rangle$ and $\mathcal{P}_\emptyset$. Recall that the new collection $\langle\mathcal{P}_e:e\in\widehat{\mathcal{H}}\rangle$ generates a finite partition of $\boldsymbol{\Omega}$ whose atoms are of the form $\bigcap_{e\in\widehat{\mathcal{H}}}A_e$ where $A_e\in\mathcal{P}_e$ for every $e\in\widehat{\mathcal{H}}$. We have the following estimate for the measure of the atoms which correspond to good families in the sense of Definition 7.8.

Claim 7.14. *For every good family $\langle A_e:e\in\widehat{\mathcal{H}}\rangle$ we have*

$$\boldsymbol{\mu}\Big(\bigcap_{e\in\widehat{\mathcal{H}}}A_e\Big)>\delta(n,r,\varepsilon). \tag{7.71}$$

Proof. Set $M_0=F(M_1)$. For every nonempty $e\in\widehat{\mathcal{H}}$ let p_e be as in (7.34) and notice that, by (7.17) and (7.30), we have $p_e\geqslant(\log F(M_1))^{-1}\geqslant M_0^{-1}$. Hence, by Lemma 7.9, we see that

$$\boldsymbol{\mu}\Big(\bigcap_{e\in\widehat{\mathcal{H}}}A_e\Big)\geqslant\big(1-c(M_r,n,r,n^r)\big)\cdot M_0^{-2^n}-\frac{C(M_0,n,r,n^r)}{F(M_0)}$$

which implies, by the choice of M_r and F in (7.68), that

$$\boldsymbol{\mu}\Big(\bigcap_{e\in\widehat{\mathcal{H}}}A_e\Big)>2^{-1}\cdot M_0^{-2^n}.$$

Finally, by (7.17), we have $M_1\leqslant\Phi_{n,r,F}(M_r)$. Therefore, by (7.69), we obtain that $2^{-1}M_0^{-2^n}\geqslant\delta(n,r,\varepsilon)$ and the proof of Claim 7.14 is completed. $\square$

Our next goal is to obtain an upper bound for the measure of the union of all atoms which are not generated by good families. Specifically, for every $j \in [r]$, every $e \in \mathcal{H}_j$ and every $A \in \mathcal{P}_e$ let $\mathcal{B}_{A,e}$ be the set of all families $\langle A_{e'} : e' \subsetneq e\rangle$ with $A_{e'} \in \mathcal{P}_{e'}$ for every $e' \subsetneq e$ and such that

$$\mathbb{E}\Big(\mathbf{1}_A \,\Big|\, \bigcap_{e' \in \partial e} A_{e'}\Big) \leqslant \frac{1}{\log F(M_j)} \ \text{ or } \ \mathbb{E}\Big(\mathrm{b}_A^2 \,\Big|\, \bigcap_{e' \subsetneq e} A_{e'}\Big) \geqslant \frac{1}{F(M_j)}. \tag{7.72}$$

Also set

$$B_{A,e} = \bigcup_{\langle A_{e'} : e' \subsetneq e\rangle \in \mathcal{B}_{A,e}} \ \bigcap_{e' \subsetneq e} A_{e'} \tag{7.73}$$

and notice that if a family $\langle A_e : e \in \widehat{\mathcal{H}}\rangle$ is not good, then there exists a nonempty $e \in \widehat{\mathcal{H}}$ such that the set $\bigcap_{e' \subsetneq e} A_{e'}$ is contained in $B_{A_e,e}$.

Claim 7.15. *Let $j \in [r]$, $e \in \mathcal{H}_j$ and $A \in \mathcal{P}_e$. Then the following hold.*

(a) *The set $B_{A,e}$ belongs to the algebra generated by the family $\bigcup_{e' \subsetneq e} \mathcal{P}_{e'}$.*
(b) *We have $\boldsymbol{\mu}(A \cap B_{A,e}) \leqslant 2/\log F(M_j)$.*

Proof. Part (a) is straightforward and so we only need to show part (b). To this end, let $\mathcal{B}_1$ be the set of all families $\langle A_{e'} : e' \in \partial e\rangle$ with $A_{e'} \in \mathcal{P}_{e'}$ for every $e' \in \partial e$ and such that $\mathbb{E}\big(\mathbf{1}_A \mid \bigcap_{e' \in \partial e} A_{e'}\big) \leqslant 1/\log F(M_j)$. Next, let $\mathcal{B}_2$ be the set of all families $\langle A_{e'} : e' \subsetneq e\rangle$ with $A_{e'} \in \mathcal{P}_{e'}$ for every $e' \subsetneq e$ and such that $\mathbb{E}\big(\mathrm{b}_A^2 \mid \bigcap_{e' \subsetneq e} A_{e'}\big) \geqslant 1/F(M_j)$. Finally, set

$$B_1 = \bigcup_{\langle A_{e'} : e' \in \partial e\rangle \in \mathcal{B}_1} \ \bigcap_{e' \in \partial e} A_{e'} \ \text{ and } \ B_2 = \bigcup_{\langle A_{e'} : e' \subsetneq e\rangle \in \mathcal{B}_2} \ \bigcap_{e' \subsetneq e} A_{e'}$$

and notice that $B_{A,e} \subseteq B_1 \cup B_2$. Therefore, it is enough to estimate the quantities $\boldsymbol{\mu}(A \cap B_1)$ and $\boldsymbol{\mu}(A \cap B_2)$. Indeed, let $\langle A_{e'} : e' \in \partial e\rangle \in \mathcal{B}_1$ and observe that

$$\boldsymbol{\mu}\Big(A \cap \bigcap_{e' \in \partial e} A_{e'}\Big) \leqslant \frac{1}{\log F(M_j)} \cdot \boldsymbol{\mu}\Big(\bigcap_{e' \in \partial e} A_{e'}\Big).$$

Note that if $\langle A_{e'} : e' \in \partial e\rangle$ and $\langle C_{e'} : e' \in \partial e\rangle$ are two distinct families in $\mathcal{B}_1$, then the sets $\bigcap_{e' \in \partial e} A_{e'}$ and $\bigcap_{e' \in \partial e} C_{e'}$ are disjoint. It follows that

$$\boldsymbol{\mu}(A \cap B_1) \leqslant \frac{1}{\log F(M_j)} \cdot \boldsymbol{\mu}(B_1) \leqslant \frac{1}{\log F(M_j)}. \tag{7.74}$$

Moreover, for every $\langle A_{e'} : e' \subsetneq e\rangle \in \mathcal{B}_2$ we have

$$\boldsymbol{\mu}\Big(A \cap \bigcap_{e' \subsetneq e} A_{e'}\Big) \leqslant F(M_j) \cdot \int \Big(\prod_{e' \subsetneq e} \mathbf{1}_{A_{e'}}\Big) \cdot \mathrm{b}_A^2 \, d\boldsymbol{\mu}.$$

Again observe that for any pair $\langle A_{e'} : e' \subsetneq e\rangle$ and $\langle C_{e'} : e' \subsetneq e\rangle$ of distinct families in $\mathcal{B}_2$ the sets $\bigcap_{e' \subsetneq e} A_{e'}$ and $\bigcap_{e' \subsetneq e} C_{e'}$ are disjoint. Hence, by (7.21),

$$\boldsymbol{\mu}(A \cap B_2) \leqslant F(M_j) \cdot \int_{B_2} \mathrm{b}_A^2 \, d\boldsymbol{\mu} \leqslant F(M_j) \cdot \|\mathrm{b}_A\|_{L_2}^2 \leqslant \frac{1}{F(M_j)} \tag{7.75}$$

and the proof of Claim 7.15 is completed. $\square$

Now for every $e \in \mathcal{H}$ we define

$$F_e = \mathbf{\Omega} \setminus \Big(B_{E_e,e} \cup \bigcup_{\emptyset \neq e' \subsetneq e} \ \bigcup_{A \in \mathcal{P}_{e'}} \big(A \cap B_{A,e'} \big) \Big) \tag{7.76}$$

and we claim that the sets $\langle F_e : e \in \mathcal{H} \rangle$ and the partitions $\langle \mathcal{P}_e : e \in \widehat{\mathcal{H}} \rangle$ satisfy the requirements of the theorem. First observe that, by Claim 7.15, for every $e \in \mathcal{H}$ the set F_e belongs to the algebra generated by the family $\bigcup_{e' \subsetneq e} \mathcal{P}_{e'}$. Moreover, by Lemma 7.7 and (7.69), we see that $\mathcal{P}_e \subseteq \mathcal{B}_e$ and $|\mathcal{P}_e| \leqslant K(n,r,\varepsilon)$ for every $e \in \widehat{\mathcal{H}}$. (Recall that, by convention, $\mathcal{B}_\emptyset = \{\emptyset, \mathbf{\Omega}\}$.) On the other hand, by Lemma 7.7 and Claim 7.15, for every $e \in \mathcal{H}$ we have

$$\begin{aligned} \boldsymbol{\mu}(E_e \setminus F_e) \ &\overset{(7.76)}{\leqslant} \ \boldsymbol{\mu}(E_e \cap B_{E_e,e}) + \sum_{j=1}^{r-1} \sum_{e' \in \mathcal{H}_j} \sum_{A \in \mathcal{P}_{e'}} \boldsymbol{\mu}(A \cap B_{A,e'}) \\ &\leqslant \ \frac{2}{\log F(M_r)} + \sum_{j=1}^{r-1} \frac{|\mathcal{H}_j| \cdot M_j \cdot 2}{\log F(M_j)} \overset{(7.68)}{\leqslant} \varepsilon. \end{aligned}$$

Thus, it is enough to show that $\bigcap_{e \in \mathcal{H}} F_e = \emptyset$. Assume, towards a contradiction, that this set is nonempty. Then there exists a family $\langle A_{e'} : e' \in \widehat{\mathcal{H}} \setminus \mathcal{H} \rangle$ with $A_{e'} \in \mathcal{P}_{e'}$ for every $e' \in \widehat{\mathcal{H}} \setminus \mathcal{H}$ and such that

$$\emptyset \neq \bigcap_{e' \in \widehat{\mathcal{H}} \setminus \mathcal{H}} A_{e'} \subseteq \bigcap_{e \in \mathcal{H}} F_e. \tag{7.77}$$

We enlarge the collection $\langle A_{e'} : e' \in \widehat{\mathcal{H}} \setminus \mathcal{H} \rangle$ by setting $A_e = E_e$ for every $e \in \mathcal{H}$. It follows that

$$\bigcap_{e \in \widehat{\mathcal{H}}} A_e \subseteq \bigcap_{e \in \mathcal{H}} E_e \tag{7.78}$$

and, consequently, the family $\langle A_e : e \in \widehat{\mathcal{H}} \rangle$ is not good. (For if not, by Claim 7.14, we would have

$$\delta(n,r,\varepsilon) < \boldsymbol{\mu}\Big(\bigcap_{e \in \widehat{\mathcal{H}}} A_e \Big) \overset{(7.78)}{\leqslant} \boldsymbol{\mu}\Big(\bigcap_{e \in \mathcal{H}} E_e \Big) \overset{(7.70)}{\leqslant} \delta(n,r,\varepsilon)$$

which is clearly impossible.) Hence, there exists a nonempty $e_0 \in \widehat{\mathcal{H}}$ such that

$$\bigcap_{e' \subsetneq e_0} A_{e'} \subseteq B_{A_{e_0}, e_0}. \tag{7.79}$$

Fix $e_1 \in \mathcal{H}$ with $e_0 \subseteq e_1$. If $e_0 = e_1$, then $A_{e_0} = E_{e_1}$ and so, by (7.77) and (7.79),

$$\emptyset \neq \bigcap_{e' \in \widehat{\mathcal{H}} \setminus \mathcal{H}} A_{e'} \subseteq F_{e_1} \cap B_{E_{e_1}, e_1} \overset{(7.76)}{=} \emptyset.$$

On the other hand, if $e_0 \subsetneq e_1$, then, invoking (7.77) and (7.79) once again, we have

$$\emptyset \neq \bigcap_{e' \in \widehat{\mathcal{H}} \setminus \mathcal{H}} A_{e'} \subseteq F_{e_1} \cap \Big(A_{e_0} \cap \bigcap_{e' \subsetneq e_0} A_{e'} \Big) \subseteq F_{e_1} \cap (A_{e_0} \cap B_{A_{e_0}, e_0}) \overset{(7.76)}{=} \emptyset.$$

Therefore, we have $\bigcap_{e \in \mathcal{H}} F_e = \emptyset$ and the proof of Theorem 7.2 is completed.

7.5. Applications

7.5.1. The hypergraph removal lemma. Let X, Y be nonempty finite sets with $|X| \leqslant |Y|$, and let $\mathcal{H}, \mathcal{G}$ be hypergraphs on X and Y respectively. A hypergraph $(Z, \mathcal{Z})$ is said to be a *copy of* $\mathcal{H}$ *in* $\mathcal{G}$ if $Z = \varphi(X)$ and $\mathcal{Z} = \{\varphi(e) : e \in \mathcal{H}\}$ where $\varphi \colon X \to Y$ is an injection satisfying $\varphi(e) \in \mathcal{G}$ for every $e \in \mathcal{H}$. The hypergraph $\mathcal{G}$ is called $\mathcal{H}$*-free* if there is no copy of $\mathcal{H}$ in $\mathcal{G}$.

The following result is known as the *hypergraph removal lemma* and is the main result in this subsection.

Theorem 7.16. *For every $n, r \in \mathbb{N}$ with $n \geqslant r \geqslant 2$ and every $0 < \varepsilon \leqslant 1$ there exist a strictly positive constant $\varrho(n, r, \varepsilon)$ and a positive integer $N(n, r, \varepsilon)$ with the following property. Let $\mathcal{H}$ be an r-uniform hypergraph on $[n]$ and $\mathcal{G}$ an r-uniform hypergraph on $[N]$ with $N \geqslant N(n, r, \varepsilon)$. If $\mathcal{G}$ contains at most $\varrho(n, r, \varepsilon) N^n$ copies of $\mathcal{H}$, then one can delete εN^r edges of $\mathcal{G}$ to make it $\mathcal{H}$-free.*

The first instance of Theorem 7.16 is for $\mathcal{H} = K_3$ (that is, when $\mathcal{H}$ is the complete graph on 3 vertices) and can be traced back to the work of Ruzsa and Szemerédi in [**RS**]. This particular case is already non-trivial and is known as the *triangle removal lemma*. The case of a general graph $\mathcal{H}$ appeared in the literature somewhat later (see [**ADLRY, Fu**]). On the other hand, the first important contribution in the context of hypergraphs was made by Frankl and Rödl [**FR1, FR2**]. Theorem 7.16 was finally proved in full generality by Gowers [**Go5**] and, independently, by Nagle, Rödl, Schacht and Skokan [**NRS, RSk**].

We will give a proof of Theorem 7.16 using Theorem 7.2. To this end, we first observe that Theorem 7.2 immediately yields the following version of Theorem 7.16 which deals with copies of simplices in uniform partite hypergraphs[2]. Although less general, this version is needed in applications in Ramsey theory.

Corollary 7.17. *Let $r \in \mathbb{N}$ with $r \geqslant 2$ and $0 < \varepsilon \leqslant 1$, and let $\delta(r+1, r, \varepsilon)$ be as in Theorem 7.2. Also let $V_1, \dots, V_{r+1}$ be pairwise disjoint nonempty finite sets and $\mathcal{G}$ an $(r+1)$-partite r-uniform hypergraph on $V_1, \dots, V_{r+1}$. If $\mathcal{G}$ contains at most $\delta(r+1, r, \varepsilon) \prod_{i=1}^{r+1} |V_i|$ copies of the r-simplex $K^{(r)}_{r+1} = \binom{[r+1]}{r}$, then for every $e \in \binom{[r+1]}{r}$ one can delete $\varepsilon \prod_{i \in e} |V_i|$ edges of $\mathcal{G} \cap \prod_{i \in e} V_i$ to make it simplex-free.*

We proceed to the proof of Theorem 7.16.

Proof of Theorem 7.16. Let $\delta(n, r, \varepsilon/n^r)$ be as in Theorem 7.2 and set

$$\varrho(n, r, \varepsilon) = \frac{\delta(n, r, \varepsilon/n^r)}{2n!} \quad \text{and} \quad N(n, r, \varepsilon) = \Big\lceil \frac{n^2}{\delta(n, r, \varepsilon/n^r)} \Big\rceil. \tag{7.80}$$

We will show that with these choices the result follows. Indeed, fix $N \geqslant N(n, r, \varepsilon)$ and let $\mathcal{H}, \mathcal{G}$ be r-uniform hypergraphs on $[n]$ and $[N]$ respectively. Also let $\mathrm{cop}(\mathcal{H}, \mathcal{G})$ be the number of copies of $\mathcal{H}$ in $\mathcal{G}$ and assume that

$$\mathrm{cop}(\mathcal{H}, \mathcal{G}) \leqslant \varrho(n, r, \varepsilon) N^n. \tag{7.81}$$

---

[2]Let $n, r \in \mathbb{N}$ with $n \geqslant r \geqslant 2$ and $V_1, \dots, V_n$ pairwise disjoint nonempty sets, and recall that an *n-partite r-uniform hypergraph on the vertex sets $V_1, \dots, V_n$* is a collection of r-element subsets F of $V_1 \cup \dots \cup V_n$ such that $|F \cap V_i| \leqslant 1$ for every $i \in [n]$.

We view the set $[N]$ as a discrete probability space equipped with the uniform probability measure, and we define an r-uniform hypergraph system

$$\mathscr{H} = (n, \langle(\Omega_i, \Sigma_i, \mu_i) : i \in [n]\rangle, \mathcal{H})$$

where $\Omega_i = [N]$ for every $i \in [n]$. Next, for every $e \in \mathcal{H}$ let

$$E_e = \big\{(y_i)_{i=1}^n \in [N]^n : \{y_i : i \in e\} \in \mathcal{G}\big\} \tag{7.82}$$

and set

$$E = \bigcap_{e \in \mathcal{H}} E_e. \tag{7.83}$$

Observe that

$$\mathrm{cop}(\mathcal{H}, \mathcal{G}) \leqslant |E| \leqslant n! \cdot \mathrm{cop}(\mathcal{H}, \mathcal{G}) + \binom{n}{2} N^{n-1}. \tag{7.84}$$

The first inequality is straightforward. On the other hand, denoting by E^* the set of all $(y_i)_{i=1}^n \in E$ such that $y_i \neq y_j$ for every $i, j \in [n]$ with $i \neq j$, we have

$$|E| - \binom{n}{2} N^{n-1} \leqslant |E^*| \leqslant n! \cdot \mathrm{cop}(\mathcal{H}, \mathcal{G})$$

which implies, of course, the second inequality in (7.84).

By (7.81), (7.84) and the choice of $\varrho(n, r, \varepsilon)$ and $N(n, r, \varepsilon)$ in (7.80), we see that $|E| \leqslant \delta(n, r, \varepsilon/n^r) N^n$. Therefore, by Theorem 7.2, for every $e \in \mathcal{H}$ there exists a set $\mathbf{F}_e \subseteq [N]^e$ such that, setting $F_e = \pi_e^{-1}(\mathbf{F}_e)$ (where, as usual, $\pi_e \colon [N]^n \to [N]^e$ is the natural projection), we have

$$\frac{|E_e \setminus F_e|}{N^n} \leqslant \frac{\varepsilon}{n^r} \quad \text{and} \quad \bigcap_{e \in \mathcal{H}} F_e = \emptyset. \tag{7.85}$$

For every $e \in \mathcal{H}$ let $\mathbf{E}_e = \pi_e(E_e) = \big\{(y_i)_{i \in e} \in [N]^e : \{y_i : i \in e\} \in \mathcal{G}\big\}$ and define

$$\mathcal{G}' = \mathcal{G} \setminus \bigcup_{e \in \mathcal{H}} \big\{\{y_i : i \in e\} : (y_i)_{i \in e} \in \mathbf{E}_e \setminus \mathbf{F}_e\big\}.$$

We claim that $\mathcal{G}'$ is as desired. First observe that $\mathcal{G}'$ is contained in $\mathcal{G}$. Moreover,

$$|\mathcal{G} \setminus \mathcal{G}'| \leqslant \big|\bigcup_{e \in \mathcal{H}} \mathbf{E}_e \setminus \mathbf{F}_e\big| \leqslant \sum_{e \in \mathcal{H}} |\mathbf{E}_e \setminus \mathbf{F}_e| = N^r \cdot \sum_{e \in \mathcal{H}} \frac{|E_e \setminus F_e|}{N^n} \overset{(7.85)}{\leqslant} \varepsilon N^r.$$

Finally, notice that $\mathcal{G}'$ is $\mathcal{H}$-free. Indeed, assume, towards a contradiction, that there is a copy $(Z, \mathcal{Z})$ of $\mathcal{H}$ in $\mathcal{G}'$. Recall that $Z = \varphi([n])$ and $\mathcal{Z} = \{\varphi(e) : e \in \mathcal{H}\}$ where $\varphi \colon [n] \to [N]$ is an injection satisfying $\varphi(e) \in \mathcal{G}'$ for every $e \in \mathcal{H}$. If $(y_i)_{i=1}^n \in [N]^n$ is defined by $y_i = \varphi(i)$ for every $i \in [n]$, then we have

$$(y_i)_{i=1}^n \in \bigcap_{e \in \mathcal{H}} (E_e \cap F_e) \subseteq \bigcap_{e \in \mathcal{H}} F_e$$

which contradicts, of course, (7.85). The proof of Theorem 7.16 is completed. $\square$

7.5.2. Ramsey-theoretic consequences. We start with the following "geometric" version of Corollary 7.17. For every integer $r \geqslant 2$ by $\mathbf{e}_1, \dots, \mathbf{e}_r$ we denote the standard basis of $\mathbb{R}^r$.

THEOREM 7.18. *For every integer $r \geqslant 2$ and every $0 < \delta \leqslant 1$ there exists a positive integer $\mathrm{Smp}(r,\delta)$ with the following property. If $n \geqslant \mathrm{Smp}(r,\delta)$, then every $D \subseteq [n]^r$ with $|D| \geqslant \delta n^r$ contains a set of the form $\{\mathbf{e}\} \cup \{\mathbf{e} + \lambda \mathbf{e}_i : 1 \leqslant i \leqslant r\}$ for some $\mathbf{e} \in [n]^r$ and some positive integer λ.*

The case "$r = 2$" in Theorem 7.18 is known as the *corners theorem* and is due to Ajtai and Szemerédi [**ASz**]. The general case is due to Furstenberg and Katznelson [**FK1**].

PROOF OF THEOREM 7.18. We follow the proof from [**So**]. Fix an integer $r \geqslant 2$ and $0 < \delta \leqslant 1$. Set $\varepsilon = 2^{-r-1}(1+r^2)^{-1}\delta^2$ and define

$$\mathrm{Smp}(r,\delta) = \lceil \delta(r+1, r, \varepsilon)^{-1} \rceil \tag{7.86}$$

where $\delta(r+1, r, \varepsilon)$ is as in Theorem 7.2. We will show that with this choice the result follows. Indeed, let $n \geqslant \mathrm{Smp}(r,\delta)$ and $D \subseteq [n]^r$ with $|D| \geqslant \delta n^r$. We have the following claim.

CLAIM 7.19. *There exist $\mathbf{x}_0 \in [2n]^r$ and $D' \subseteq D$ with $|D'| \geqslant 2^{-r}\delta^2 n^r$ and such that $D' = \mathbf{x}_0 - D'$.*

PROOF OF CLAIM 7.19. Set $X = [2n]^r$ and $Y = [n]^r$. Also let $\boldsymbol{\mu}_X$ and $\boldsymbol{\mu}_Y$ be the uniform probability measures on X and Y respectively. Note that for every $\mathbf{y} \in Y$ we have $\mathbf{y} + D \subseteq X$, and so $\boldsymbol{\mu}_X(\mathbf{y} + D) = \boldsymbol{\mu}_X(D) \geqslant 2^{-r}\delta$. Moreover,

$$\begin{aligned} \int \boldsymbol{\mu}_Y\big(D \cap (\mathbf{x} - D)\big)\, d\boldsymbol{\mu}_X &= \int \Big(\int \mathbf{1}_D(\mathbf{y}) \cdot \mathbf{1}_{\mathbf{x}-D}(\mathbf{y})\, d\boldsymbol{\mu}_Y \Big)\, d\boldsymbol{\mu}_X \\ &= \int \Big(\int \mathbf{1}_D(\mathbf{y}) \cdot \mathbf{1}_{\mathbf{y}+D}(\mathbf{x})\, d\boldsymbol{\mu}_Y \Big)\, d\boldsymbol{\mu}_X \\ &= \int \mathbf{1}_D(\mathbf{y}) \cdot \Big(\int \mathbf{1}_{\mathbf{y}+D}(\mathbf{x})\, d\boldsymbol{\mu}_X \Big)\, d\boldsymbol{\mu}_Y \\ &\geqslant 2^{-r}\delta \cdot \int \mathbf{1}_D(\mathbf{y})\, d\boldsymbol{\mu}_Y \geqslant 2^{-r}\delta^2. \end{aligned}$$

We select $\mathbf{x}_0 \in X$ with $\boldsymbol{\mu}_Y\big(D \cap (\mathbf{x}_0 - D)\big) \geqslant 2^{-r}\delta^2$ and we set $D' = D \cap (\mathbf{x}_0 - D)$. Clearly, $\mathbf{x}_0$ and D' are as desired. The proof of Claim 7.19 is completed. $\square$

Let $\mathbf{x}_0$ and D' be as in Claim 7.19. Since $D' = \mathbf{x}_0 - D'$, it is enough to show that D' contains a set of the form $\{\mathbf{e}'\} \cup \{\mathbf{e}' + \lambda' \mathbf{e}_i : 1 \leqslant i \leqslant r\}$ for some $\mathbf{e}' \in [n]^r$ and some nonzero $\lambda' \in \mathbb{Z}$.

Assume, towards a contradiction, that no such configuration is contained in D'. We define an $(r+1)$-partite r-uniform hypergraph $\mathcal{G}$ with vertex sets $V_1, \dots, V_{r+1}$ as follows. First, for every $i \in [r]$ set

$$V_i = \{H^i_m : m \in J_i\}$$

where $J_i = [n]$ and $H^i_m = \{(x_1, \dots, x_r) \in \mathbb{R}^r : x_i = m\}$ for every $m \in J_i$. Also let

$$V_{r+1} = \{H^{r+1}_m : m \in J_{r+1}\}$$

where $J_{r+1} = \{r, \dots, rn\}$ and $H^{r+1}_m = \{(x_1, \dots, x_r) \in \mathbb{R}^r : x_1 + \dots + x_r = m\}$ for every $m \in J_{r+1}$. Notice that for every $1 \leqslant i_1 < \dots < i_r \leqslant r+1$ and every $H^{i_1}_{m_1} \in V_{i_1}, \dots, H^{i_r}_{m_r} \in V_{i_r}$ the hyperplanes $H^{i_1}_{m_1}, \dots, H^{i_r}_{m_r}$ intersect in a unique point of $[n]^r$. We define the edges of $\mathcal{G}$ to be those sets $\{H^{i_1}_{m_1}, \dots, H^{i_r}_{m_r}\}$ for which the unique common point of $H^{i_1}_{m_1}, \dots, H^{i_r}_{m_r}$ belongs to D'.

Now observe that our assumption for the set D' implies that all r-simplices of $\mathcal{G}$ are degenerate in the following sense. Let $S = \{H^1_{m_1}, \dots, H^{r+1}_{m_{r+1}}\}$ where $H^i_{m_i} \in V_i$ for every $i \in [r+1]$, and assume that S is an r-simplex of $\mathcal{G}$. By the definition of $\mathcal{G}$, we see that

$$\{(m_1, \dots, m_r)\} \cup \Big\{(m_1, \dots, m_r) + \big(m_{r+1} - \sum_{l=1}^{r} m_l\big) \cdot \mathbf{e}_i : 1 \leqslant i \leqslant r\Big\} \subseteq D'$$

and so we must have $m_{r+1} - (m_1 + \dots + m_r) = 0$. In other words, all the hyperplanes $H^1_{m_1}, \dots, H^{r+1}_{m_{r+1}}$ must contain the point $(m_1, \dots, m_r) \in D'$.

It follows from the previous discussion that there is a natural bijection between the set of all r-simplices of $\mathcal{G}$ and the set D'. In particular, if s is the number of r-simplices of $\mathcal{G}$, then we have

$$s = |D'| \leqslant n^r \overset{(7.86)}{\leqslant} \delta(r+1, r, \varepsilon) \cdot n^{r+1} \leqslant \delta(r+1, r, \varepsilon) \cdot \prod_{i=1}^{r+1} |V_i|.$$

By Corollary 7.17, Claim 7.19 and the choice of ε, we may remove

$$\varepsilon \cdot n^r + r \cdot (\varepsilon \cdot r \cdot n^r) = 2^{-r-1}\delta^2 \cdot n^r \leqslant |D'|/2$$

edges of $\mathcal{G}$ to make it simplex-free. On the other hand, note that the r-simplices of $\mathcal{G}$ are pairwise edge-disjoint. Indeed, if $S = \{H^1_{m_1}, \dots, H^{r+1}_{m_{r+1}}\}$ is an r-simplex of $\mathcal{G}$, then every edge of S determines the unique common point of $H^1_{m_1}, \dots, H^{r+1}_{m_{r+1}}$. Thus, we have to remove at least $|D'|$ edges of $\mathcal{G}$ in order to make it simplex-free. This is clearly a contradiction and the proof of Theorem 7.18 is completed. $\square$

The following theorem is due to Szemerédi [**Sz1**].

Theorem 7.20. *For every integer $k \geqslant 2$ and every $0 < \delta \leqslant 1$ there exists a positive integer $\mathrm{Sz}(k, \delta)$ with the following property. If $n \geqslant \mathrm{Sz}(k, \delta)$, then every $D \subseteq [n]$ with $|D| \geqslant \delta n$ contains an arithmetic progression of length k.*

Szemerédi's theorem is a deep and remarkably influential result. In particular, there are numerous different proofs of Theorem 7.20 some of which are discussed in [**TV**, Chapter 11]. The best known general upper bounds for the numbers $\mathrm{Sz}(k, \delta)$ are due to Gowers [**Go3**]:

$$\mathrm{Sz}(k, \delta) \leqslant 2^{2^{\delta^{-2^{2^{k+9}}}}}. \tag{7.87}$$

We will present a proof of Szemerédi's theorem using Theorem 7.18. The argument is amenable to generalizations, but has the drawback that it offers very poor upper bounds for the numbers $\mathrm{Sz}(k, \delta)$.

We will need the following lemma. For every positive integer d and every $\mathbf{u} = (u_1, \dots, u_d) \in \mathbb{R}^d$ we set $\|\mathbf{u}\|_\infty = \max\{|u_1|, \dots, |u_d|\}$.

Lemma 7.21. *Let d, m be positive integers and $\mathbf{u}_1, \dots, \mathbf{u}_m$ nonzero vectors in $\mathbb{N}^d$, and set $M = \max\{\|\mathbf{u}_i\|_\infty : 1 \leqslant i \leqslant m\}$. Define $\Phi \colon \mathbb{N}^{d+m} \to \mathbb{N}^d$ by*

$$\Phi\big(\mathbf{x}, (y_1, \dots, y_m)\big) = \mathbf{x} + y_1 \cdot \mathbf{u}_1 + \dots + y_m \cdot \mathbf{u}_m \tag{7.88}$$

for every $\mathbf{x} \in \mathbb{N}^d$ and every $(y_1, \dots, y_m) \in \mathbb{N}^m$. Also let n be a positive integer, $0 < \delta \leqslant 1$ and $D \subseteq [n]^d$ with $|D| \geqslant \delta n^d$. Then there exists $\mathbf{z}_0 \in \mathbb{Z}^{d+m}$ such that

$$|\big(\mathbf{z}_0 + \Phi^{-1}(D)\big) \cap [n]^{d+m}| \geqslant \Big(\frac{\delta}{(1+Mm)^d}\Big) \cdot n^{d+m}. \tag{7.89}$$

Proof. It is similar to the proof of Claim 7.19. Set $X = \{-Mmn+1, \dots, n\}^d$, $Y = [n]^m$ and let $\boldsymbol{\mu}_X$ and $\boldsymbol{\mu}_Y$ be the uniform probability measures on X and Y. Also set $Z = X \times Y$ and let $\boldsymbol{\mu}_Z$ be the uniform probability measure on Z. Moreover, define $\phi \colon \mathbb{N}^m \to \mathbb{N}^d$ by $\phi\big((y_1, \dots, y_m)\big) = y_1 \cdot \mathbf{u}_1 + \dots + y_m \cdot \mathbf{u}_m$. For every $\mathbf{y} \in Y$ we have $D - \phi(\mathbf{y}) \subseteq X$ and so $\boldsymbol{\mu}_X\big(D - \phi(\mathbf{y})\big) = \boldsymbol{\mu}_X(D) \geqslant \delta/(1+Mm)^d$. Hence,

$$\begin{aligned} \boldsymbol{\mu}_Z\big(\Phi^{-1}(D)\big) &= \int \mathbf{1}_D\big(\Phi(\mathbf{x}, \mathbf{y})\big)\, d\boldsymbol{\mu}_Z = \int \mathbf{1}_D\big(\mathbf{x} + \phi(\mathbf{y})\big)\, d\boldsymbol{\mu}_Z \\ &= \int \Big(\int \mathbf{1}_{D-\phi(\mathbf{y})}(\mathbf{x})\, d\boldsymbol{\mu}_X \Big)\, d\boldsymbol{\mu}_Y \\ &= \int \boldsymbol{\mu}_X\big(D - \phi(\mathbf{y})\big)\, d\boldsymbol{\mu}_Y \geqslant \frac{\delta}{(1+Mm)^d}. \end{aligned}$$

Finally, observe that Z can be partitioned into translates of $[n]^{d+m}$. Therefore, there exists $\mathbf{z}_0' \in \mathbb{Z}^{d+m}$ such that $|\Phi^{-1}(D) \cap \big(\mathbf{z}_0' + [n]^{d+m}\big)| \geqslant \delta\,(1+Mm)^{-d}\, n^{d+m}$. The vector $\mathbf{z}_0 = -\mathbf{z}_0'$ is as desired. The proof of Lemma 7.21 is completed. □

We proceed to the proof of Theorem 7.20.

Proof of Theorem 7.20. We will show that

$$\mathrm{Sz}(k, \delta) \leqslant \mathrm{Smp}(k-1, \delta/k^2) \tag{7.90}$$

for every integer $k \geqslant 3$ and every $0 < \delta \leqslant 1$. Indeed, fix $k \geqslant 3$ and $0 < \delta \leqslant 1$, and let $n \geqslant \mathrm{Smp}(k-1, \delta/k^2)$ and $D \subseteq [n]$ with $|D| \geqslant \delta n$. We define $\Phi \colon \mathbb{N}^{k-1} \to \mathbb{N}$ by

$$\Phi(x_1, x_2, \dots, x_{k-1}) = x_1 + 2x_2 + \dots + (k-1)x_{k-1}.$$

By Lemma 7.21 applied for "$d = 1$" and "$m = k-2$", there exists $\mathbf{z}_0 \in \mathbb{Z}^{k-1}$ such that $\mathrm{dens}_{[n]^{k-1}}\big(\mathbf{z}_0 + \Phi^{-1}(D)\big) \geqslant \delta/k^2$. Next, we apply Theorem 7.18 and we select $\mathbf{e} \in [n]^{k-1}$ and $\lambda > 0$ such that $\{\mathbf{e}\} \cup \{\mathbf{e} + \lambda \mathbf{e}_i : 1 \leqslant i \leqslant k-1\} \subseteq \mathbf{z}_0 + \Phi^{-1}(D)$. Observe that $\Phi(\mathbf{e}_i) = i$ for every $i \in [k-1]$. Therefore, setting $c = \Phi(\mathbf{e} - \mathbf{z}_0)$, we see that the arithmetic progression $\{c + i \cdot \lambda : 0 \leqslant i \leqslant k-1\}$ is contained D. The proof of Theorem 7.20 is completed. □

Our last application is known as the *multidimensional Szemerédi theorem* and is due to Furstenberg and Katznelson [**FK1**].

THEOREM 7.22. *For every pair k, d of positive integers with $k \geqslant 2$ and every $0 < \delta \leqslant 1$ there exists a positive integer* $\mathrm{MSz}(k,d,\delta)$ *with the following property. If $n \geqslant \mathrm{MSz}(k,d,\delta)$, then every $D \subseteq [n]^d$ with $|D| \geqslant \delta n^d$ contains a set of the form $\big\{\mathbf{c} + \lambda\mathbf{x} : \mathbf{x} \in \{0,\dots,k-1\}^d\big\}$ for some $\mathbf{c} \in \mathbb{N}^d$ and some positive integer λ.*

The first quantitative information for the numbers $\mathrm{MSz}(k,d,\delta)$ became available as a consequence of the hypergraph removal lemma, but there are now several different effective proofs of Theorem 7.22 (we will present an alternative approach in Section 8.4). Despite this progress, the best known upper bounds for the numbers $\mathrm{MSz}(k,d,\delta)$ have an Ackermann-type dependence with respect to k for each fixed $d \geqslant 2$ and $0 < \delta \leqslant 1$. We proceed to the proof.

PROOF OF THEOREM 7.22. It is similar to the proof of Theorem 7.20. The case "$d = 1$" is the content of Theorem 7.20, and so we may assume that $d \geqslant 2$. We claim that

$$\mathrm{MSz}(k,d,\delta) \leqslant \mathrm{Smp}(k^d - 1, \delta/k^{d+1}). \tag{7.91}$$

To see this, fix $n \geqslant \mathrm{Smp}(k^d - 1, \delta/k^{d+1})$ and let $D \subseteq [n]^d$ with $|D| \geqslant \delta n^d$. Set $m = k^d - 1 - d$ and enumerate the set $\{0,\dots,k-1\}^d \setminus \{\mathbf{0}, \mathbf{e}_1, \dots, \mathbf{e}_d\}$ as $\mathbf{u}_1, \dots, \mathbf{u}_m$. Finally, define $\Phi\colon \mathbb{N}^{d+m} \to \mathbb{N}^d$ by

$$\Phi\big(\mathbf{x}, (y_1,\dots,y_m)\big) = \mathbf{x} + y_1\mathbf{u}_1 + \cdots + y_m\mathbf{u}_m.$$

By Lemma 7.21, we have

$$\mathrm{dens}_{[n]^{k^d-1}}\big(\mathbf{z}_0 + \Phi^{-1}(D)\big) \geqslant \delta/k^{d+1}$$

for some $\mathbf{z}_0 \in \mathbb{Z}^{k^d-1}$. Therefore, by Theorem 7.18 and the choice of n, there exist $\mathbf{e} \in [n]^{k^d-1}$ and $\lambda > 0$ such that $\{\mathbf{e}\} \cup \{\mathbf{e} + \lambda\mathbf{e}_i : 1 \leqslant i \leqslant k^d - 1\} \subseteq \mathbf{z}_0 + \Phi^{-1}(D)$. Note that $\Phi(\mathbf{e}_i) = \mathbf{e}_i$ if $i \in [d]$, while $\Phi(\mathbf{e}_{d+i}) = \mathbf{u}_i$ if $i \in [k^d - 1]$. Hence, setting $\mathbf{c} = \Phi(\mathbf{e} - \mathbf{z}_0)$, we see that $\big\{\mathbf{c} + \lambda\mathbf{x} : \mathbf{x} \in \{0,\dots,k-1\}^d\big\} \subseteq D$. The proof of Theorem 7.22 is completed. $\square$

7.6. Notes and remarks

7.6.1. Theorem 7.2 and its proof are due to Tao [**Tao1**]. Actually, Tao considered only those probability spaces which are relevant in the context of graphs and hypergraphs (namely, nonempty finite sets equipped with their uniform probability measures), but his approach works in full generality. Theorem 7.3 is new.

7.6.2. Theorem 7.16 was conjectured by Erdős, Frankl and Rödl [**EFR**] in the mid-1980s. There is now a variety of different proofs and extensions of this important result; see, e.g., [**AT, ES, RSc1, RSc2, RSc3, Tao3**]. Nevertheless, all effective proofs of Theorem 7.16 follow the same strategy as the proof of Theorem 7.2 and proceed by establishing a version of the regularity lemma for uniform hypergraphs and a corresponding counting lemma. In particular, the best known lower bounds for the constant $\varrho(n,r,\varepsilon)$ in Theorem 7.16 have an Ackermann-type dependence with respect to r. The problem of improving upon these estimates is of fundamental importance and has been asked by several authors (see, e.g., [**Tao1**]).

CHAPTER 8

The density Hales–Jewett theorem

The following result is known as the *density Hales–Jewett theorem* and is due to Furstenberg and Katznelson [**FK4**].

THEOREM 8.1. *For every integer $k \geqslant 2$ and every $0 < \delta \leqslant 1$ there exists a positive integer N with the following property. If $n \geqslant N$ and A is an alphabet with $|A| = k$, then every $D \subseteq A^n$ with $|D| \geqslant \delta |A^n|$ contains a combinatorial line of A^n. The least positive integer with this property will be denoted by* $\mathrm{DHJ}(k, \delta)$.

The density Hales–Jewett theorem is a fundamental result of Ramsey theory. It has several strong results as consequences, including Szemerédi's theorem on arithmetic progressions [**Sz1**], its multidimensional version [**FK1**] and the IP_r-Szemerédi theorem [**FK2**]. We present these applications, among others, in Section 8.4.

The rest of this chapter is devoted to the proof of Theorem 8.1. The case "$k = 2$" follows from a classical result in extremal combinatorics due to Sperner [**Sp**]. Sperner's theorem and its relation with the density Hales–Jewett theorem are discussed in Section 8.1. In Section 8.2 we present some preliminary tools which are needed for the proof of Theorem 8.1 but are not directly related to the main argument. The proof of Theorem 8.1 is completed in Section 8.3.

8.1. Sperner's theorem

A family $\mathcal{A}$ of subsets of a nonempty set X is called an *antichain* if none of the sets is contained in any other, that is, if $A \not\subseteq B$ for every $A, B \in \mathcal{A}$ with $A \neq B$. Notice that for every positive integer $k \leqslant |X|$ the family $\binom{X}{k}$ is an antichain of subsets of X. It follows, in particular, that there exists an antichain of subsets of $[n]$ of cardinality $\binom{n}{\lfloor n/2 \rfloor}$ for every integer $n \geqslant 1$. The following theorem due to Sperner [**Sp**] asserts that this is the largest antichain of subsets of $[n]$.

THEOREM 8.2. *Let n be a positive integer and $\mathcal{A}$ an antichain of subsets of $[n]$. Then we have $|\mathcal{A}| \leqslant \binom{n}{\lfloor n/2 \rfloor}$.*

PROOF. We follow the proof from [**Lu**]. The case "$n = 1$" is straightforward, and so we may assume that $n \geqslant 2$. Let $\mathcal{C}$ be the set of all finite sequences $(S_i)_{i=1}^n$ of subsets of $[n]$ such that $|S_i| = i$ for every $i \in [n]$ and $S_i \subseteq S_{i+1}$ for every $i \in [n-1]$. Note that $|\mathcal{C}| = n!$. For every nonempty subset S of $[n]$ let $\mathcal{C}(S)$ be the set of all sequences from $\mathcal{C}$ which contain S. Observe that if $|S| = r$, then

$$|\mathcal{C}(S)| = r!\,(n-r)! \geqslant \lfloor n/2 \rfloor!\,(n - \lfloor n/2 \rfloor)!.$$

Finally, notice that if $\mathcal{A}$ is an antichain of subsets of $[n]$, then $\mathcal{C}(A)\cap\mathcal{C}(B)=\emptyset$ for every $A,B\in\mathcal{A}$ with $A\neq B$. Therefore,

$$n! = |\mathcal{C}| \geqslant |\bigcup_{A\in\mathcal{A}}\mathcal{C}(A)| = \sum_{A\in\mathcal{A}}|\mathcal{C}(A)| \geqslant |\mathcal{A}|\big(\lfloor n/2\rfloor!\,\big(n-\lfloor n/2\rfloor\big)!\big)$$

which yields that $|\mathcal{A}|\leqslant\binom{n}{[n/2]}$. The proof of Theorem 8.2 is completed. □

Now let n be a positive integer and observe that we may identify every subset of $[n]$ with a word over $\{0,1\}$ of length n via its characteristic function. Note that this particular bijection

$$\mathcal{P}([n])\ni F\mapsto \mathbf{1}_F\in\{0,1\}^n$$

has the following property. It maps pairs of subsets of $[n]$ which are comparable under inclusion, to combinatorial lines of $\{0,1\}^n$. Taking into account these observations and using Sperner's theorem we obtain the following corollary.

Corollary 8.3. *For every* $0<\delta\leqslant 1$ *we have* $\mathrm{DHJ}(2,\delta)\leqslant 4/\delta^2$.

Proof. By a standard approximation related to Stirling's formula, we have

$$e^{7/8} < \frac{n!}{(n/e)^n\sqrt{n}} < e$$

for every $n\geqslant 2$ (see, e.g., [**Ru**, page 200]). Hence,

$$\binom{n}{\lfloor n/2\rfloor} < 2\frac{1}{\sqrt{n}}2^n$$

for every positive integer n. Using this estimate and Theorem 8.2, we see that $|\mathcal{A}|/2^n < 2/\sqrt{n}$ for every antichain $\mathcal{A}$ of subsets of $[n]$. Therefore, if $\delta\geqslant 2/\sqrt{n}$, then every collection of subsets of $[n]$ of cardinality at least $\delta 2^n$ contains two subsets S and T with $S\neq T$ and $S\subseteq T$. The proof of Corollary 8.3 is completed. □

8.2. Preliminary tools

The first result in this section asserts that the density Hales–Jewett theorem implies its multidimensional version.

Proposition 8.4. *Let* $k\in\mathbb{N}$ *with* $k\geqslant 2$ *and assume that for every* $0<\varrho\leqslant 1$ *the number* $\mathrm{DHJ}(k,\varrho)$ *has been estimated. Then for every integer* $m\geqslant 1$ *and every* $0<\delta\leqslant 1$ *there exists a positive integer* $\mathrm{MDHJ}(k,m,\delta)$ *with the following property. If* $n\geqslant\mathrm{MDHJ}(k,m,\delta)$ *and* A *is an alphabet with* $|A|=k$*, then every subset of* A^n *with density at least* δ *contains an* m*-dimensional combinatorial subspace of* A^n.

Proof. By induction on m. The case "$m=1$" follows, of course, from our assumptions. Let $m\in\mathbb{N}$ with $m\geqslant 1$ and assume that the result has been proved up to m. For every $0<\delta\leqslant 1$ let $M=\mathrm{DHJ}(k,\delta/2)$ and set

$$\mathrm{MDHJ}(k,m+1,\delta)=M+\mathrm{MDHJ}(k,m,\delta 2^{-1}(k+1)^{-M}). \tag{8.1}$$

We will show that the positive integer $\mathrm{MDHJ}(k,m+1,\delta)$ is as desired. To this end, let $n\geqslant\mathrm{MDHJ}(k,m+1,\delta)$. Also let A be an alphabet with $|A|=k$ and fix a subset

D of A^n with $\mathrm{dens}(D) \geqslant \delta$. For every $w \in A^{n-M}$ set $D_w = \{y \in A^M : w^\frown y \in D\}$ and observe that

$$\mathbb{E}_{w \in A^{n-M}} \mathrm{dens}(D_w) = \mathrm{dens}(D) \geqslant \delta.$$

Therefore, there exists a subset E of A^{n-M} with $\mathrm{dens}(E) \geqslant \delta/2$ such that for every $w \in E$ we have $\mathrm{dens}(D_w) \geqslant \delta/2$. By the choice of M, for every $w \in E$ there exists a combinatorial line L_w of A^M such that $L_w \subseteq D_w$. The number of combinatorial lines of A^M is $(k+1)^M - k^M$, and so, less than $(k+1)^M$. Hence, by the classical pigeonhole principle, there exist a combinatorial line L of A^M and a subset F of E with $\mathrm{dens}(F) \geqslant \delta 2^{-1}(k+1)^{-M}$ and such that $L \subseteq D_w$ for every $w \in F$. Since $n - M \geqslant \mathrm{MDHJ}(k, m, \delta 2^{-1}(k+1)^{-M})$ there exists an m-dimensional combinatorial subspace W of A^{n-M} with $W \subseteq F$. We set $V = W^\frown L$. Then V is an $(m+1)$-dimensional combinatorial subspace of A^n and clearly $V \subseteq D$. The proof of Proposition 8.4 is completed. $\square$

The next result is a simpler version of Lemma 6.22 and asserts that every dense subset of a hypercube of sufficiently large dimension, becomes extremely uniformly distributed when restricted on a suitable combinatorial space.

Lemma 8.5. *Let A be a finite alphabet with $|A| \geqslant 2$, m a positive integer and $0 < \varepsilon < 1$. Also let $n \geqslant \varepsilon^{-1}|A|^m m$ and let D be a subset of A^n with $\mathrm{dens}(D) > \varepsilon$. Then there exist an integer l with $m \leqslant l < n$ and an m-dimensional combinatorial subspace W of A^l such that for every $w \in W$ we have $\mathrm{dens}(D_w) \geqslant \mathrm{dens}(D) - \varepsilon$ where $D_w = \{y \in A^{n-l} : w^\frown y \in D\}$ is the section of D at w.*

Proof. We set $\varrho = \varepsilon(|A|^m - 1)^{-1}$. Also let $W_1 = A^m$ and observe that $\mathbb{E}_{w \in W_1} \mathrm{dens}(D_w) = \mathrm{dens}(D)$. Note that if W_1 does not satisfy the requirements of the lemma, then there exists $w_1 \in W_1$ such that $\mathrm{dens}(D_{w_1}) \geqslant \mathrm{dens}(D) + \varrho$. Next we set $W_2 = w_1^\frown A^m$ and we observe that $\mathbb{E}_{w \in W_2} \mathrm{dens}(D_w) = \mathrm{dens}(D_{w_1}) \geqslant \mathrm{dens}(D) + \varrho$. Again we note that if W_2 is not the desired combinatorial subspace, then there exists $w_2 \in W_2$ such that $\mathrm{dens}(D_{w_2}) \geqslant \mathrm{dens}(D) + 2\varrho$. This process must terminate, of course, after at most $\lfloor \varrho^{-1} \rfloor$ iterations. Noticing that $(\lfloor \varrho^{-1} \rfloor + 1)m < n$ the proof of Lemma 8.5 is completed. $\square$

By Proposition 8.4 and Lemma 8.5, we obtain the following corollary.

Corollary 8.6. *Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every $0 < \varrho \leqslant 1$ the number $\mathrm{DHJ}(k, \varrho)$ has been estimated. Then for every integer $m \geqslant 1$ and every $0 < \delta \leqslant 1$ there exists a positive integer $\mathrm{MDHJ}^*(k, m, \delta)$ with the following property. If $n \geqslant \mathrm{MDHJ}^*(k, m, \delta)$ and A is an alphabet with $|A| = k+1$, then for every $D \subseteq A^n$ with density at least δ and every $B \subseteq A$ with $|B| = k$ there exists an m-dimensional combinatorial subspace V of A^n such that $V \upharpoonright B$ is contained in D, where $V \upharpoonright B$ is as in (1.21).*

Proof. Let $M = \mathrm{MDHJ}(k, m, \delta/2)$ and set

$$\mathrm{MDHJ}^*(k, m, \delta) = (\delta/2)^{-1}(k+1)^M M. \tag{8.2}$$

We claim that with this choice the result follows. Indeed, let $n \geqslant \mathrm{MDHJ}^*(k, m, \delta)$ and let A be an alphabet with $|A| = k+1$. Also let D be a subset of A^n with

$\mathrm{dens}(D) \geqslant \delta$ and fix $B \subseteq A$ with $|B| = k$. By Lemma 8.5 and (8.2), there exist some $l \in \mathbb{N}$ with $M \leqslant l < n$ and an M-dimensional combinatorial subspace W of A^l such that $\mathrm{dens}(D_w) \geqslant \delta/2$ for every $w \in W$. We set $Z = W \upharpoonright B$. On the one hand, we have $|D \cap (Z^\frown A^{n-l})| \geqslant (\delta/2)|Z^\frown A^{n-l}|$ since $\mathrm{dens}(D_z) \geqslant \delta/2$ for every $z \in Z$. On the other hand, the family $\{Z^\frown y : y \in A^{n-l}\}$ forms a partition of $Z^\frown A^{n-l}$ into sets of equal size. Hence, there exists $y_0 \in A^{n-l}$ such that $|D \cap (Z^\frown y_0)| \geqslant (\delta/2)|Z^\frown y_0|$. Let I_W be the canonical isomorphism associated with the combinatorial space W (see Definition 1.2) and define $\Phi\colon B^M \to Z^\frown y_0$ by the rule $\Phi(w) = \mathrm{I}_W(w)^\frown y_0$. Notice that Φ is a bijection. By the choice of M, there exists an m-dimensional combinatorial subspace U of B^M with $U \subseteq \Phi^{-1}(D)$. If V is the unique m-dimensional combinatorial subspace of A^n with $V \upharpoonright B = \Phi(U)$, then V is as desired. The proof of Corollary 8.6 is completed. □

We close this section with the following measure-theoretic consequence of the density Hales–Jewett theorem.

PROPOSITION 8.7. *Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every $0 < \varrho \leqslant 1$ the number $\mathrm{DHJ}(k, \varrho)$ has been estimated. Let $0 < \delta \leqslant 1$ and set*

$$n_0 = n_0(k, \delta) = \mathrm{DHJ}(k, \delta/2) \quad \textit{and} \quad \zeta(k, \delta) = \frac{\delta/2}{(k+1)^{n_0} - k^{n_0}}. \tag{8.3}$$

If A is an alphabet with $|A| = k$, then for every combinatorial space W of $A^{<\mathbb{N}}$ of dimension at least n_0 and every family $\{D_w : w \in W\}$ of measurable events in a probability space (Ω, Σ, μ) satisfying $\mu(D_w) \geqslant \delta$ for every $w \in W$, there exists a combinatorial line L of W such that

$$\mu\Big(\bigcap_{w \in L} D_w\Big) \geqslant \zeta(k, \delta). \tag{8.4}$$

Proposition 8.7 appears as Proposition 2.1 in [**FK4**], though the argument in its proof can be traced in an old paper of Erdős and Hajnal [**EH**]. In Subsection 8.4.3 we will present an extension of this result.

PROOF OF PROPOSITION 8.7. Clearly, we may assume that W is of the form A^n for some $n \geqslant n_0$. Let $\{D_w : w \in A^n\}$ be as in the statement of the lemma. We select $y_0 \in A^{n-n_0}$ and we set

$$X = \big\{\omega \in \Omega : \mathrm{dens}\big(\{v \in A^{n_0} : \omega \in D_{v^\frown y_0}\}\big) \geqslant \delta/2\big\}.$$

Notice that $\mu(X) \geqslant \delta/2$ since $\mu\big(D_{v^\frown y_0}\big) \geqslant \delta$ for every $v \in A^{n_0}$. Let $\omega \in X$ be arbitrary. By the choice of n_0 in (8.3), there exists a combinatorial line L_ω of A^{n_0} such that $L_\omega \subseteq \{v \in A^{n_0} : \omega \in D_{v^\frown y_0}\}$. In other words, for every $\omega \in X$ there exists a combinatorial line L_ω of A^{n_0} with

$$\omega \in \bigcap_{v \in L_\omega} D_{v^\frown y_0}.$$

The number of combinatorial lines of A^{n_0} is equal to $(k+1)^{n_0} - k^{n_0}$ and so there exist a combinatorial line L_0 of A^{n_0} and a measurable subset X_0 of X with

$$\mu(X_0) \geqslant \frac{\delta/2}{(k+1)^{n_0} - k^{n_0}} \tag{8.5}$$

and such that $L_\omega = L_0$ for every $\omega \in X_0$. We set $L = \{w^\frown y_0 : w \in L_0\}$ and we observe that L is a combinatorial line of A^n. Moreover,

$$\mu\Big(\bigcap_{w\in L} D_w\Big) = \mu\Big(\bigcap_{v\in L_0} D_{v^\frown y_0}\Big) \geqslant \mu(X_0) \overset{(8.5)}{\geqslant} \frac{\delta/2}{(k+1)^{n_0}-k^{n_0}} \overset{(8.3)}{=} \zeta(k,\delta).$$

The proof of Proposition 8.7 is completed. □

8.3. Proof of Theorem 8.1

The proof proceeds by induction on k and is based on a *density increment strategy*, a method invented by Roth [**Ro**]. The case "$k=2$" is, of course, the content of Corollary 8.3. Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every $0 < \varrho \leqslant 1$ the number $\mathrm{DHJ}(k,\varrho)$ has been estimated. This assumption permits us to introduce some numerical invariants. Specifically, for every $0 < \delta \leqslant 1$ we set

$$m_0 = \mathrm{DHJ}(k,\delta/4), \quad \theta = \frac{\delta/4}{(k+1)^{m_0}-k^{m_0}}, \quad \eta = \frac{\delta\theta}{48} \ \text{ and } \ \gamma = \frac{\eta^2}{2k}. \tag{8.6}$$

The main step of the proof of Theorem 8.1 is the following dichotomy.

PROPOSITION 8.8. *Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every $0 < \varrho \leqslant 1$ the number* $\mathrm{DHJ}(k,\varrho)$ *has been estimated. Then for every $0 < \delta \leqslant 1$ and every integer $d \geqslant 1$ there exists a positive integer $N(k,d,\delta)$ with the following property. If $n \geqslant N(k,d,\delta)$ and A is an alphabet with $|A| = k+1$, then for every subset D of A^n with* $\mathrm{dens}(D) \geqslant \delta$ *we have that either:* (i) *D contains a combinatorial line of A^n, or* (ii) *there exists a d-dimensional combinatorial subspace V of A^n such that* $\mathrm{dens}_V(D) \geqslant \mathrm{dens}(D) + \gamma$ *where γ is as in* (8.6).

Using Proposition 8.8 the numbers $\mathrm{DHJ}(k+1,\delta)$ can be estimated easily with a standard iteration. Indeed, fix $0 < \delta \leqslant 1$ and define a sequence (n_i) in $\mathbb{N}$ recursively by the rule

$$\begin{cases} n_0 = 1, \\ n_{i+1} = N(k, n_i, \delta). \end{cases} \tag{8.7}$$

Then, by Proposition 8.8, we have

$$\mathrm{DHJ}(k+1,\delta) \leqslant n_{\lceil \gamma^{-1} \rceil}.$$

It remains to prove Proposition 8.8. This is our goal in the next subsection.

8.3.1. Proof of Proposition 8.8. We follow the proof from [**DKT2**]. First we introduce some pieces of notation. Specifically, for every integer $m \geqslant 1$ and every $0 < \varepsilon \leqslant 1$ we set

$$n(m,\varepsilon) = \varepsilon^{-1}(k+1)^m m. \tag{8.8}$$

Notice that the number $n(m,\varepsilon)$ is the threshold appearing in Lemma 8.5 for an alphabet of cardinality $k+1$. *We also fix an alphabet A with $k+1$ letters and, in what follows, we will assume that for every $0 < \varrho \leqslant 1$ the number* $\mathrm{DHJ}(k,\varrho)$ *has been estimated.*

Our objective in the first part of the proof is to obtain a "probabilistic" strengthening of our assumptions. This "probabilistic" strengthening refers to the natural question whether a dense subset of Γ^n, where Γ is an alphabet with k letters and n is sufficiently large, not only will contain a combinatorial line but actually a non-trivial portion of them. Unfortunately this is not true, as is shown in the following example.

EXAMPLE 8.1. Let $0<\varepsilon<1$ be arbitrary. Also let Γ be a finite alphabet with $|\Gamma|\geqslant 2$. We will show that for every sufficiently large integer n there exists $D\subseteq\Gamma^n$ with $\mathrm{dens}(D)\geqslant 1-\varepsilon$ and such that $|\{L\in \mathrm{Subsp}_1(\Gamma^n): L\subseteq D\}|\leqslant \varepsilon|\mathrm{Subsp}_1(\Gamma^n)|$.

For every $g\in\Gamma$, every integer $n\geqslant 1$ and every $w=(w_0,\dots,w_{n-1})\in\Gamma^n$ we set $N_g(w)=|\{i\in\{0,\dots,n-1\}: w_i=g\}|$. Moreover, let

$$E(\Gamma,g,n)=\Big\{w\in\Gamma^n: |N_g(w)-\frac{n}{|\Gamma|}|\geqslant n^{2/3}\Big\}$$

and define

$$D(\Gamma,n)=\Gamma^n\setminus\Big(\bigcup_{g\in\Gamma}E(\Gamma,g,n)\Big).$$

We have the following properties.

(P1) For every $g\in\Gamma$ and every $n\geqslant 1$ we have $\mathrm{dens}\big(E(\Gamma,g,n)\big)\leqslant n^{-1/3}$.
(P2) For every $n\geqslant 1$ we have $\mathrm{dens}\big(D(\Gamma,n)\big)\geqslant 1-|\Gamma|n^{-1/3}$.
(P3) For every $n\geqslant 27(|\Gamma|+1)^3$ we have

$$|\{L\in\mathrm{Subsp}_1(\Gamma^n): L\subseteq D(\Gamma,n)\}|\leqslant (n/2)^{-1/3}|\mathrm{Subsp}_1(\Gamma^n)|.$$

By (P2) and (P3), it is clear that the set $D(\Gamma,n)$ is as desired as long as n is sufficiently large depending on $|\Gamma|$ and ε.

To see that the above properties are satisfied, fix $g\in\Gamma$ and $n\geqslant 1$. For every $i\in\{0,\dots,n-1\}$ define $X_{g,i}\colon\Gamma^n\to\{0,1\}$ by $X_{g,i}\big((w_0,\dots,w_{n-1})\big)=1$ if $w_i=g$ and $X_{g,i}\big((w_0,\dots,w_{n-1})\big)=0$ otherwise. Also let $X_g=\sum_{i=0}^{n-1}X_{g,i}$. Note that the sequence $(X_{g,i})_{i=0}^{n-1}$ is an independent sequence of random variables. (Here, we view Γ^n as a discrete probability space equipped with the uniform probability measure.) Moreover, $\mathbb{E}(X_{g,i})=1/|\Gamma|$ and $\mathrm{Var}(X_{g,i})=\mathbb{E}(X_{g,i}^2)-\mathbb{E}(X_{g,i})^2=1/|\Gamma|-1/|\Gamma|^2<1$ for every $i\in\{0,\dots,n-1\}$. Therefore, $\mathbb{E}(X_g)=\sum_{i=0}^{n-1}\mathbb{E}(X_{g,i})=n/|\Gamma|$ and, because of independence, $\mathrm{Var}(X_g)=\sum_{i=0}^{n-1}\mathrm{Var}(X_{g,i})<n$. Finally, observe that $X_g(w)=N_g(w)$ for every $w\in\Gamma^n$ and $n^{2/3}=n^{1/6}n^{1/2}>n^{1/6}\sqrt{\mathrm{Var}(X_g)}$. Hence,

$$E(\Gamma,g,n)\subseteq\Big\{w\in\Gamma^n: |X_g(w)-\mathbb{E}(X_g)|\geqslant n^{1/6}\sqrt{\mathrm{Var}(X_g)}\Big\}.$$

By Chebyshev's inequality, this inclusion implies property (P1). Property (P2) follows immediately by (P1) and the definition of $D(\Gamma,n)$. The last property is also an easy consequence of (P1). Indeed, fix an integer $n\geqslant 27(|\Gamma|+1)^3$. Also fix a letter x not belonging to Γ which we view as a variable, and identify every combinatorial line of Γ^n with a word over $\Gamma\cup\{x\}$ of length n. Note that the cardinality of the wildcard set of every combinatorial line contained in $D(\Gamma,n)$ is less than $2n^{2/3}$. By the choice of n, we see that $2n^{2/3}\leqslant\frac{n}{|\Gamma|+1}-n^{1/3}$ and so the set $\{L\in\mathrm{Subsp}_1(\Gamma^n): L\subseteq D(\Gamma,n)\}$ is contained in the set $E(\Gamma\cup\{x\},x,n)$. Thus, by property (P1) applied for the alphabet "$\Gamma\cup\{x\}$" and "$g=x$", we obtain that $|\{L\in\mathrm{Subsp}_1(\Gamma^n): L\subseteq D(\Gamma,n)\}|\leqslant n^{-1/3}(|\Gamma|+1)^n$. Taking into account the fact that $n \geqslant 27(|\Gamma|+1)^3$ we see that $|\mathrm{Subsp}_1(\Gamma^n)| = (|\Gamma|+1)^n - |\Gamma|^n \geqslant 2^{1/3}(|\Gamma|+1)^n$. Combining the previous estimates we conclude that property (P3) is satisfied.

In spite of the above example, we will show that dense subsets of hypercubes indeed contain plenty of combinatorial lines, but when restricted on appropriately chosen combinatorial spaces. The main tools for locating these combinatorial spaces are Proposition 2.25, Lemma 8.5 and Proposition 8.7.

We start with the following lemma.

Lemma 8.9. *Let $0 < \delta \leqslant 1$ and $m \geqslant m_0$. If $n \geqslant n(\mathrm{GR}(k,m,1,2), \eta^2/2)$, then for every $D \subseteq A^n$ with $\mathrm{dens}(D) \geqslant \delta$ and every $B \subseteq A$ with $|B| = k$ there exist some $l \in \mathbb{N}$ with $m \leqslant l < n$ and an m-dimensional subspace U of A^l such that*

(a) *for every $u \in U$ we have $\mathrm{dens}(D_u) \geqslant \mathrm{dens}(D) - \eta^2/2$, and*

(b) *for every combinatorial line L of $U \upharpoonright B$ we have $\mathrm{dens}\big(\bigcap_{u\in L} D_u\big) \geqslant \theta$*

where $U \upharpoonright B$ is as in (1.21) and $D_u = \{y \in A^{n-l} : u^\frown y \in D\}$ for every $u \in U$.

Proof. By Lemma 8.5, there exist an integer l with $\mathrm{GR}(k,m,1,2) \leqslant l < n$ and a combinatorial subspace W of A^l with $\dim(W) = \mathrm{GR}(k,m,1,2)$ such that $\mathrm{dens}(D_w) \geqslant \mathrm{dens}(D) - \eta^2/2$ for every $w \in W$. We set

$$\mathcal{L} = \Big\{L \in \mathrm{Subsp}_1(W \upharpoonright B) : \mathrm{dens}\Big(\bigcap_{w\in L} D_w\Big) \geqslant \theta\Big\}.$$

As in Subsection 1.3.2, using the canonical isomorphism I_W associated with W, we identify $W \upharpoonright B$ with $B^{\dim(W)}$ and $\mathrm{Subsp}_1(W \upharpoonright B)$ with $\mathrm{Subsp}_1(B^{\dim(W)})$. Hence, by Proposition 2.25, there is an m-dimensional combinatorial subspace V of $W \upharpoonright B$ such that either $\mathrm{Subsp}_1(V) \subseteq \mathcal{L}$ or $\mathrm{Subsp}_1(V) \cap \mathcal{L} = \emptyset$. If $\mathrm{Subsp}_1(V) \subseteq \mathcal{L}$, then let U be the unique combinatorial subspace of A^l with $U \upharpoonright B = V$. Clearly, U satisfies the requirements of the lemma.

Therefore, it is enough to show that $\mathrm{Subsp}_1(V) \cap \mathcal{L} \neq \emptyset$. Indeed, notice that since $V \subseteq W$, we have $\mathrm{dens}(D_v) \geqslant \mathrm{dens}(D) - \eta^2/2 \geqslant \delta/2$ for every $v \in V$. Moreover, $\dim(V) = m \geqslant m_0 = \mathrm{DHJ}(k, \delta/4)$ and so, by Proposition 8.7, there exists $L \in \mathrm{Subsp}_1(V) \cap \mathcal{L}$. The proof of Lemma 8.9 is completed. $\square$

The next lemma completes the first part of the proof.

Lemma 8.10. *Let $0 < \delta \leqslant 1$ and $m \geqslant m_0$. Also let $n \geqslant n(\mathrm{GR}(k,m,1,2), \eta^2/2)$ and $D \subseteq A^n$ with $\mathrm{dens}(D) \geqslant \delta$. Then either:* (i) *there exists an m-dimensional combinatorial subspace X of A^n such that $\mathrm{dens}_X(A) \geqslant \mathrm{dens}(D) + \eta^2/2$, or* (ii) *for every $B \subseteq A$ with $|B| = k$ there exists an m-dimensional combinatorial subspace W of A^n such that $\mathrm{dens}_W(D) \geqslant \mathrm{dens}(D) - 2\eta$ and*

$$|\{L \in \mathrm{Subsp}_1(W \upharpoonright B) : L \subseteq D\}| \geqslant (\theta/2)|\mathrm{Subsp}_1(W \upharpoonright B)|. \tag{8.9}$$

Proof. Assume that part (i) is not satisfied, that is, for every m-dimensional combinatorial subspace X of A^n we have $\mathrm{dens}_X(D) < \mathrm{dens}(D) + \eta^2/2$. Fix $B \subseteq A$ with $|B| = k$. By Lemma 8.5, there exist $l \in \{m, \dots, n-1\}$ and an m-dimensional combinatorial subspace U of A^l such that $\mathrm{dens}(D_u) \geqslant \mathrm{dens}(D) - \eta^2/2$ for every

$u \in U$ and $\mathrm{dens}\big(\bigcap_{u\in L} D_u\big) \geqslant \theta$ for every $L \in \mathrm{Subsp}_1(U \upharpoonright B)$. The first property implies, in particular, that

$$\mathbb{E}_{y\in A^{n-l}} \mathrm{dens}_{U^\frown y}(D) \geqslant \mathrm{dens}(D) - \eta^2/2. \tag{8.10}$$

Observe that for every $y \in A^{n-l}$ the set $U^\frown y$ is an m-dimensional combinatorial subspace of A^n. Hence, by our assumption, we have $\mathrm{dens}_{U^\frown y}(D) < \mathrm{dens}(D) + \eta^2/2$ for every $y \in A^{n-l}$. By Lemma E.3 and (8.10), there exists $H_1 \subseteq A^{n-l}$ with $\mathrm{dens}(H_1) \geqslant 1 - \eta$ and such that $\mathrm{dens}_{U^\frown y}(D) \geqslant \mathrm{dens}(D) - 2\eta$ for every $y \in H_1$.

Now for every $y \in A^{n-l}$ let $\mathcal{L}_y = \{L \in \mathrm{Subsp}_1(U \upharpoonright B) : y \in \bigcap_{u\in L} D_u\}$. Since $\mathrm{dens}\big(\bigcap_{u\in L} D_u\big) \geqslant \theta$ for every $L \in \mathrm{Subsp}_1(U \upharpoonright B)$ we have

$$\mathbb{E}_{y\in A^{n-l}} \frac{|\mathcal{L}_y|}{|\mathrm{Subsp}_1(U \upharpoonright B)|} = \mathbb{E}_{L\in \mathrm{Subsp}_1(U\upharpoonright B)} \mathrm{dens}\Big(\bigcap_{u\in L} D_u\Big) \geqslant \theta. \tag{8.11}$$

Therefore, there exists a subset H_2 of A^{n-l} with $\mathrm{dens}(H_2) \geqslant \theta/2$ and such that $|\mathcal{L}_y| \geqslant (\theta/2)|\mathrm{Subsp}_1(U \upharpoonright B)|$ for every $y \in H_2$. By the choice of θ and η in (8.6), we have $\eta < \theta/2$. It follows that the set $H_1 \cap H_2$ is nonempty. We select $y_0 \in H_1 \cap H_2$ and we set $W = U^\frown y_0$. It is clear that W is as desired. The proof of Lemma 8.10 is completed. $\square$

In the second part of the proof we will show that if a dense subset of A^n contains no combinatorial line, then it must correlate more than expected with a "simple" subset of A^n. The proper concept of "simplicity" in this context is related to the notion of an insensitive set introduced by Shelah in [**Sh1**]. In particular, the reader is advised to review the material in Subsection 2.1.1.

We also need to introduce some more numerical invariants. Specifically, let

$$\lambda = \frac{k+1}{k} \ \text{ and } \ M_0 = \max\Big\{m_0, \frac{\log \eta^{-1}}{\log \lambda}\Big\}. \tag{8.12}$$

where m_0 and η are as in (8.6). We have the following lemma.

LEMMA 8.11. *Let $0 < \delta \leqslant 1$ and $m \geqslant M_0$. Also let $n \geqslant n(\mathrm{GR}(k,m,1,2),\eta^2/2)$ and $D \subseteq A^n$ with $\mathrm{dens}(D) \geqslant \delta$. Finally, let $a \in A$ and set $B = A \setminus \{a\}$. Assume that D contains no combinatorial line of A^n and $\mathrm{dens}_X(A) < \mathrm{dens}(D) + \eta^2/2$ for every m-dimensional combinatorial subspace X of A^n. Then there exist an m-dimensional combinatorial subspace W of A^n and a subset C of W satisfying the following properties.*

(a) *We have $\mathrm{dens}_W(C) \geqslant \theta/4$ and $C = \bigcap_{b\in B} C_b$ where C_b is (a,b)-insensitive in W for every $b \in B$.*
(b) *We have $\mathrm{dens}_W\big(D \cap (W \setminus C)\big) \geqslant \big(\mathrm{dens}(D) + 6\eta\big)\, \mathrm{dens}_W(W \setminus C)$ and, moreover, $\mathrm{dens}_W\big(D \cap (W \setminus C)\big) \geqslant \mathrm{dens}(D) - 3\eta$.*

PROOF. By Lemma 8.10 and our assumptions, there exists an m-dimensional combinatorial subspace W of A^n with $\mathrm{dens}_W(D) \geqslant \mathrm{dens}(D) - 2\eta$ and satisfying (8.9). For every $L \in \mathrm{Subsp}_1(W \upharpoonright B)$ let v_L be the unique variable word over A such that $L = \{u_L(b) : b \in B\}$ and define

$$C = \big(D \cap (W \upharpoonright B)\big) \cup \{u_L(a) : L \in \mathrm{Subsp}_1(W \upharpoonright B) \text{ and } L \subseteq D\}. \tag{8.13}$$

We will show that W and C are as desired.

First observe that the map $\mathrm{Subsp}_1(W \upharpoonright B) \ni L \mapsto v_L(a) \in W$ is one-to-one. Therefore, by (8.9), we have

$$|\{v_L(a) : L \in \mathrm{Subsp}_1(W \upharpoonright B) \text{ and } L \subseteq D\}| \geqslant (\theta/2)|\mathrm{Subsp}_1(W \upharpoonright B)|. \tag{8.14}$$

Hence,

$$\begin{aligned} |C| &\geqslant |\{v_L(a) : L \in \mathrm{Subsp}_1(W \upharpoonright B) \text{ and } L \subseteq D\}| \overset{(8.14)}{\geqslant} (\theta/2)|\mathrm{Subsp}_1(W \upharpoonright B)| \\ &= (\theta/2)\big((k+1)^m - k^m\big) \overset{(8.12)}{\geqslant} (\theta(1-\eta)/2)(k+1)^m \overset{(8.6)}{\geqslant} (\theta/4)|W| \end{aligned}$$

which is equivalent to saying that $\mathrm{dens}_W(C) \geqslant \theta/4$. Next let $\mathrm{I}_W : A^m \to W$ be the canonical isomorphism associated with W. For every $b \in B$ and every $w \in A^m$ let $w^{a\to b}$ be the unique element of B^m obtained by replacing all appearances of the letter a in w by b. (Notice that $w^{a\to b} = w$ if $w \in B^m$.) We set

$$C_b = \big\{\mathrm{I}_W(w) : w \in A^m \text{ and } w^{a\to b} \in \mathrm{I}_W^{-1}(D)\big\}.$$

Then observe that C_b is (a,b)-insensitive in W for every $b \in B$, and $C = \bigcap_{b\in B} C_b$.

We proceed to the proof of the second part of the lemma. Our assumption that D contains no combinatorial line of A^n implies that $D \cap C \subseteq W \upharpoonright B$. In particular, we have $\mathrm{dens}_W(D \cap C) \leqslant \lambda^{-m} \leqslant \lambda^{-M_0} \leqslant \eta$ by the choice of λ and M_0 in (8.12). Since $\mathrm{dens}_W(D) \geqslant \mathrm{dens}(D) - 2\eta$, we see that $\mathrm{dens}_W\big(D \cap (W \setminus C)\big) \geqslant \mathrm{dens}(D) - 3\eta$. Therefore,

$$\begin{aligned} \frac{\mathrm{dens}_W\big(D \cap (W \setminus C)\big)}{\mathrm{dens}_W(W \setminus C)} &\geqslant \frac{\mathrm{dens}(D) - 3\eta}{1 - \theta/4} \geqslant \big(\mathrm{dens}(D) - 3\eta\big)(1 + \theta/4) \\ &\overset{(8.6)}{\geqslant} \mathrm{dens}(D) + 6\eta. \end{aligned}$$

The proof of Lemma 8.11 is completed. □

The following corollary completes the second part of the proof.

Corollary 8.12. *Let $0 < \delta \leqslant 1$ and $m \geqslant M_0$. Let $n \geqslant n(\mathrm{GR}(k,m,1,2), \eta^2/2)$ and $D \subseteq A^n$ with $\mathrm{dens}(D) \geqslant \delta$. Finally, let $a \in A$ and set $B = A \setminus \{a\}$. Assume that D contains no combinatorial line of A^n. Then there exist an m-dimensional combinatorial subspace W of A^n and a family $\{S_b : b \in B\}$ of subsets of W such that S_b is (a,b)-insensitive in W for every $b \in B$ and, moreover, setting $S = \bigcap_{b\in B} S_b$ we have $\mathrm{dens}_W(S) \geqslant \gamma$ and $\mathrm{dens}_W(D \cap S) \geqslant \big(\mathrm{dens}(D) + 2\gamma\big)\,\mathrm{dens}_W(S)$.*

Proof. Assume that there exists an m-dimensional combinatorial subspace X of A^n such that $\mathrm{dens}_X(D) \geqslant \delta + \eta^2/2$. Then we set $W = X$ and $S_b = X$ for every $b \in B$. Since $\eta^2/2 \geqslant 2\gamma$, it is clear that with these choices the result follows. Otherwise, by Lemma 8.11, there exist an m-dimensional combinatorial subspace W of A^n and a set $C = \bigcap_{b\in B} C_b$, where C_b is (a,b)-insensitive in W for every $b \in B$, such that $\mathrm{dens}_W\big(D \cap (W \setminus C)\big) \geqslant (\mathrm{dens}(D) + 6\eta)\,\mathrm{dens}_W(W \setminus C)$ and $\mathrm{dens}_W\big(D \cap (W \setminus C)\big) \geqslant \delta - 3\eta$. Let $\{b_1, \dots, b_k\}$ be an enumeration of B. We set $P_1 = W \setminus C_{b_1}$ and $P_i = C_{b_1} \cap \cdots \cap C_{b_{i-1}} \cap (W \setminus C_{b_i})$ if $i \in \{2, \dots, k\}$. Notice that the family $\{P_1, \dots, P_k\}$ is a partition of $W \setminus C$. Therefore, setting

$\lambda_i = \mathrm{dens}_W(P_i)/\mathrm{dens}_W(W \setminus C)$ and $\delta_i = \mathrm{dens}_W(D \cap P_i)/\mathrm{dens}_W(P_i)$ for every $i \in [k]$ (with the convention that $\delta_i = 0$ if P_i happens to be empty), we see that

$$\sum_{i=1}^{k} \lambda_i \delta_i = \frac{\mathrm{dens}_W\big(D \cap (W \setminus C)\big)}{\mathrm{dens}_W(W \setminus C)} \geqslant \mathrm{dens}(D) + 6\eta.$$

Hence, there exists $i_0 \in [k]$ such that $\delta_{i_0} \geqslant \mathrm{dens}(D) + 3\eta \geqslant \mathrm{dens}(D) + 2\gamma$ and $\lambda_{i_0} \geqslant 3\eta/k$. We define $S_{b_i} = C_{b_i}$ if $i < i_0$, $S_{b_{i_0}} = W \setminus C_{b_{i_0}}$ and $S_{b_i} = W$ if $i > i_0$. Clearly S_b is (a,b)-insensitive in W for every $b \in B$. Moreover, setting $S = \bigcap_{b\in B} S_b$ we see that $S = P_{i_0}$, and so, $\mathrm{dens}_W(S) = \lambda_{i_0}\mathrm{dens}_W(W \setminus C) \geqslant (3\eta/k)(\delta - 3\eta) \geqslant \gamma$ and $\mathrm{dens}_W(D \cap S) = \delta_{i_0}\mathrm{dens}_W(S) \geqslant \big(\mathrm{dens}(D) + 2\gamma\big)\,\mathrm{dens}_W(S)$. The proof of Corollary 8.12 is completed. $\square$

In the third, and last, part of the proof our goal is to almost entirely partition the set S obtained by Corollary 8.12 into combinatorial spaces of sufficiently large dimension. This is achieved by appropriately modifying an argument of Ajtai and Szemerédi [**ASz**].

First we deal with the case of an arbitrary insensitive set. For every $0 < \beta \leqslant 1$ and every integer $m \geqslant 1$ we set

$$M_1 = \mathrm{MDHJ}^*(k, m, \beta) \ \text{ and } \ F(m,\beta) = \lceil \beta^{-1}(k{+}1{+}m)^{M_1}(k{+}1)^{M_1 - m} M_1 \rceil \quad (8.15)$$

where $\mathrm{MDHJ}^*(k, m, \beta)$ is as in Corollary 8.6. We have the following lemma.

Lemma 8.13. *Let $0 < \beta \leqslant 1/2$ and $m \geqslant 1$. If $n \geqslant F(m,\beta)$, then for every $a, b \in A$ with $a \neq b$ and every (a,b)-insensitive subset S of A^n with* $\mathrm{dens}(S) \geqslant 2\beta$ *there exists a family $\mathcal{V}$ of pairwise disjoint m-dimensional combinatorial subspaces of A^n which are all contained in S and are such that* $\mathrm{dens}(S \setminus \cup\mathcal{V}) < 2\beta$.

Proof. We set $\Theta = \beta(k+1+m)^{-M_1}(k+1)^{m-M_1}$. Notice that $\Theta < \beta$ and

$$n \geqslant F(m,\beta) = \lceil \Theta^{-1} M_1 \rceil \geqslant \Theta^{-1} M_1. \quad (8.16)$$

Fix $a, b \in A$ with $a \neq b$ and set $B = A \setminus \{a\}$. We will determine a positive integer $r_0 \leqslant \lfloor \Theta^{-1} \rfloor$ and we will select, recursively, a strictly decreasing sequence $S = S_0 \supsetneq S_1 \supsetneq \cdots \supsetneq S_{r_0}$ of subsets of S and a sequence $\mathcal{V}_1, \dots, \mathcal{V}_{r_0}$ of families of m-dimensional combinatorial subspaces of A^n subject to the following conditions.

(C1) For every $r \in [r_0]$ the family $\mathcal{V}_r$ consists of pairwise disjoint m-dimensional combinatorial subspaces of A^n which are contained in $S_{r-1} \setminus S_r$. Moreover, we have $\mathrm{dens}(\cup\mathcal{V}_r) \geqslant \Theta$.

(C2) For every $r \in \{0, \dots, r_0\}$ and every $z \in A^{rM_1}$ the set

$$S_r^z = \{t \in A^{n - rM_1} : t^\frown z \in S_r\}$$

is (a,b)-insensitive.

(C3) If $r \in \{0, \dots, r_0 - 1\}$, then we have $2\beta \leqslant \mathrm{dens}(S_r) \leqslant \mathrm{dens}(S) - r\Theta$. On the other hand, we have $\mathrm{dens}(S_{r_0}) < 2\beta$.

The first step is identical to the general one, and so let r be a positive integer with $r \leqslant \lfloor \Theta^{-1} \rfloor$ and assume that the sequences $(S_j)_{j=0}^{r}$ and $(\mathcal{V}_j)_{j=1}^{r}$ have been selected. If $\mathrm{dens}(S_r) < 2\beta$, then we set "$r_0 = r$" and we terminate the recursive selection.

Otherwise, we have $\mathrm{dens}(S_r) \geqslant 2\beta$. Note that this estimate and condition (C1) yield that $r\Theta + 2\beta \leqslant 1$. Since $\Theta < \beta$ we see that $\lfloor \Theta^{-1} \rfloor \geqslant (r+1)$ and so

$$n - (r+1)M_1 \overset{(8.16)}{\geqslant} \Theta^{-1}M_1 - \lfloor \Theta^{-1} \rfloor M_1 \geqslant 0. \tag{8.17}$$

It follows, in particular, that we may write A^n as $A^{n-(r+1)M_1} \times A^{M_1} \times A^{rM_1}$. For every $(t,z) \in A^{n-(r+1)M_1} \times A^{rM_1}$ let $S_r^{(t,z)} = \{y \in A^{M_1} : t^\frown y^\frown z \in S_r\}$. By our inductive assumptions, the set S_r^z is (a,b)-insensitive for every $z \in A^{rM_1}$. Noticing that $S_r^{(t,z)}$ is section of S_r^z at t, we see that $S_r^{(t,z)}$ is (a,b)-insensitive for every $(t,z) \in A^{n-(r+1)M_1} \times A^{rM_1}$. Also observe that

$$\mathbb{E}_{(t,z)\in A^{n-(r+1)M_1}\times A^{rM_1}} \mathrm{dens}\big(S_r^{(t,z)}\big) = \mathrm{dens}(S_r) \geqslant 2\beta.$$

Hence, there exists a subset Γ_0 of $A^{n-(r+1)M_1} \times A^{rM_1}$ of density at least β such that $\mathrm{dens}(S_r^{(t,z)}) \geqslant \beta$ for every $(t,z) \in \Gamma_0$. Let $(t,z) \in \Gamma_0$ be arbitrary. By Corollary 8.6 and the choice of M_1 in (8.15), there exists an m-dimensional combinatorial subspace $V_{(t,z)}$ of A^{M_1} such that $V_{(t,z)} \upharpoonright B \subseteq S_r^{(t,z)}$, and so, $V_{(t,z)} \subseteq S_r^{(t,z)}$ since $S_r^{(t,z)}$ is (a,b)-insensitive. The number of m-dimensional combinatorial subspaces of A^{M_1} is less than $(k+1+m)^{M_1}$. Therefore, by the classical pigeonhole principle, there exists an m-dimensional combinatorial subspace V of A^{M_1} such that, setting

$$\Gamma = \big\{(t,z) \in A^{n-(r+1)M_1} \times A^{rM_1} : t^\frown V^\frown z \subseteq S_r\big\}, \tag{8.18}$$

we have

$$\mathrm{dens}(\Gamma) \geqslant \frac{\mathrm{dens}(\Gamma_0)}{(k+1+m)^{M_1}} \geqslant \beta(k+1+m)^{-M_1}. \tag{8.19}$$

We define

$$\mathcal{V}_{r+1} = \{t^\frown V^\frown z : (t,z) \in \Gamma\} \ \text{ and } \ S_{r+1} = S_r \setminus \cup\mathcal{V}_{r+1}. \tag{8.20}$$

We will show that $\mathcal{V}_{r+1}$ and S_{r+1} are as desired. Indeed, notice first that

$$\begin{aligned} |\cup\mathcal{V}_{r+1}| &= \mathrm{dens}(\Gamma)(k+1)^{n-M_1}(k+1)^m \\ &\overset{(8.19)}{\geqslant} \beta(k+1+m)^{-M_1}(k+1)^{m-M_1}(k+1)^n = \Theta(k+1)^n. \end{aligned}$$

Using this estimate and invoking the definition of $\mathcal{V}_{r+1}$ and S_{r+1} we see that condition (C1) is satisfied. To see that condition (C2) is also satisfied, fix $z_1 \in A^{(r+1)M_1}$. We need to prove that the set $S_{r+1}^{z_1} = \{t \in A^{n-(r+1)M_1} : t^\frown z_1 \in S_{r+1}\}$ is (a,b)-insensitive. By (8.20), it is enough to show that

$$t_0{}^\frown z_1 \in S_r \Leftrightarrow t_1{}^\frown z_1 \in S_r \tag{8.21}$$

and

$$t_0{}^\frown z_1 \in \cup\mathcal{V}_{r+1} \Leftrightarrow t_1{}^\frown z_1 \in \cup\mathcal{V}_{r+1} \tag{8.22}$$

for every $t_0, t_1 \in A^{n-(r+1)M_1}$ which are (a,b)-equivalent. Fix such a pair t_0, t_1 and write $z_1 = y_0{}^\frown z_0$ where $(y_0, z_0) \in A^{M_1} \times A^{rM_1}$. By our inductive assumptions, the set $S_r^{z_0}$ is (a,b)-insensitive. Since $t_0{}^\frown y_0$ and $t_1{}^\frown y_0$ are (a,b)-equivalent, we have

$$t_0{}^\frown z_1 \in S_r \Leftrightarrow t_0{}^\frown y_0 \in S_r^{z_0} \Leftrightarrow t_1{}^\frown y_0 \in S_r^{z_0} \Leftrightarrow t_1{}^\frown z_1 \in S_r$$

and so (8.21) is satisfied. Next observe that, by the definition of the set Γ in (8.18), we have $\{t \in A^{n-(r+1)M_1} : (t, z_0) \in \Gamma\} = \{t \in A^{n-(r+1)M_1} : t^\frown V \subseteq S_r^{z_0}\}$. Invoking the (a,b)-insensitivity of $S_r^{z_0}$, we obtain that the set $\{t \in A^{n-(r+1)M_1} : (t, z_0) \in \Gamma\}$ is also (a,b)-insensitive. Therefore,

$$\begin{aligned} t_0{}^\frown z_1 \in \cup \mathcal{V}_{r+1} \;&\Leftrightarrow\; t_0{}^\frown y_0{}^\frown z_0 \in \cup\{t^\frown V^\frown z : (t,z) \in \Gamma\} \\ &\Leftrightarrow\; (t_0, z_0) \in \Gamma \text{ and } y_0 \in V \Leftrightarrow (t_1, z_0) \in \Gamma \text{ and } y_0 \in V \\ &\Leftrightarrow\; t_1{}^\frown y_0{}^\frown z_0 \in \cup\{t^\frown V^\frown z : (t,z) \in \Gamma\} \Leftrightarrow t_1{}^\frown z_1 \in \cup \mathcal{V}_{r+1}. \end{aligned}$$

It follows that (8.22) is also satisfied, and thus condition (C2) is fulfilled. Since condition (C3) for the set S_{r+1} will be checked in the next iteration, the recursive selection is completed.

Now, by (8.16) and condition (C1), we see that the above algorithm will eventually terminate after at most $\lfloor \Theta^{-1} \rfloor$ iterations. We set $\mathcal{V} = \mathcal{V}_1 \cup \dots \cup \mathcal{V}_{r_0}$. By conditions (C1) and (C3), the family $\mathcal{V}$ is as desired. The proof of Lemma 8.13 is completed. $\square$

By recursion on $r \in [k]$, for every $0 < \beta \leqslant 1$ and every integer $m \geqslant 1$ we define $F^{(r)}(m,\beta)$ by the rule

$$F^{(1)}(m,\beta) = F(m,\beta) \quad \text{and} \quad F^{(r+1)}(m,\beta) = F^{(r)}\big(F(m,\beta),\beta\big). \tag{8.23}$$

We have the following corollary.

Corollary 8.14. *Let $r \in [k]$, $0 < \beta \leqslant 1/2r$ and $m \geqslant 1$. Let $n \geqslant F^{(r)}(m,\beta)$, $a \in A$ and $b_1, \dots, b_r$ distinct elements of $A \setminus \{a\}$. For every $i \in [r]$ let S_i be an (a, b_i)-insensitive subset of A^n. We set $S = S_1 \cap \dots \cap S_r$. If $\mathrm{dens}(S) \geqslant 2r\beta$, then there exists a family $\mathcal{V}$ of pairwise disjoint m-dimensional combinatorial subspaces of A^n which are all contained in S and are such that* $\mathrm{dens}(S \setminus \cup\mathcal{V}) < 2r\beta$.

Proof. By induction on r. The case "$r = 1$" follows from Lemma 8.13. Let $r \in [k-1]$ and assume that the result has been proved up to r. Fix $n \geqslant F^{(r+1)}(m,\beta)$, $a \in A$ and $b_1, \dots, b_{r+1}$ distinct elements of $A \setminus \{a\}$. Also let $S_1, \dots, S_{r+1}$ be as in the statement of the corollary. By our inductive assumptions, there exists a family $\mathcal{V}'$ of pairwise disjoint $F(m,\beta)$-dimensional combinatorial subspaces of A^n which are all contained in $S' := S_1 \cap \dots \cap S_r$ and are such that $\mathrm{dens}(S' \setminus \cup\mathcal{V}') < 2r\beta$. Let $\mathcal{V}'' = \{V \in \mathcal{V}' : \mathrm{dens}_V(S_{r+1}) \geqslant 2\beta\}$. Notice that for every $V \in \mathcal{V}''$ the set $V \cap S_{r+1}$ is (a, b_{r+1})-insensitive in V and $\dim(V) = F(m,\beta)$. Hence, by identifying each $V \in \mathcal{V}''$ with $A^{F(m,\beta)}$ via the canonical isomorphism I_V and applying Lemma 8.13, we obtain for every $V \in \mathcal{V}''$ a collection $\mathcal{V}_V$ of pairwise disjoint m-dimensional combinatorial subspaces of V which are all contained in $V \cap S_{r+1}$ and are such that $\mathrm{dens}_V(S_{r+1} \setminus \cup\mathcal{V}_V) < 2\beta$. We set $\mathcal{V} = \{W : V \in \mathcal{V}'' \text{ and } W \in \mathcal{V}_V\}$. Clearly, $\mathcal{V}$ is as desired. The proof of Corollary 8.14 is completed. $\square$

We are now ready to give the proof of Proposition 8.8.

Proof of Proposition 8.8. For every $d \in \mathbb{N}$ with $d \geqslant 1$ and every $0 < \delta \leqslant 1$ let $\beta = \gamma^2/2k$ and $m(d) = \max\{M_0, F^{(k)}(d,\beta)\}$. We define

$$N(k,d,\delta) = n\big(\mathrm{GR}\big(k, m(d), 1, 2\big), \eta^2/2\big). \tag{8.24}$$

Let $n \geqslant N(k,d,\delta)$, an alphabet A with $|A| = k+1$ and a subset D of A^n with $\mathrm{dens}(D) \geqslant \delta$. Assume that D contains no combinatorial line of A^n. Fix $a \in A$ and set $B = A \setminus \{a\}$. By Corollary 8.12, there exist a combinatorial subspace W of A^n of dimension $m(d)$ and a family $\{S_b : b \in B\}$ of subsets of W such that S_b is (a,b)-insensitive in W for every $b \in B$ and, setting $S = \bigcap_{b\in B} S_b$, we have $\mathrm{dens}_W(S) \geqslant \gamma$ and $\mathrm{dens}_W(D \cap S) \geqslant \big(\mathrm{dens}(D) + 2\gamma\big)\,\mathrm{dens}_W(S)$. By Corollary 8.14, there exists a family $\mathcal{V}$ of pairwise disjoint d-dimensional combinatorial subspaces of W such that $\cup\mathcal{V} \subseteq S$ and $\mathrm{dens}_W(S \setminus \cup\mathcal{V}) < 2k\beta = \gamma^2$. Therefore,

$$\begin{aligned}
\mathrm{dens}_W(D \cap \cup\mathcal{V}) &\geqslant \mathrm{dens}_W(D\cap S) - \mathrm{dens}_W(S \setminus \cup\mathcal{V})\\
&\geqslant \big(\mathrm{dens}(D) + 2\gamma\big)\,\mathrm{dens}_W(S) - \gamma^2\\
&\geqslant \big(\mathrm{dens}(D) + \gamma\big)\,\mathrm{dens}_W(S) \geqslant \big(\mathrm{dens}(D) + \gamma\big)\,\mathrm{dens}_W(\cup\mathcal{V}).
\end{aligned}$$

Hence, there exists $V \in \mathcal{V}$ such that $\mathrm{dens}_W(D\cap V) \geqslant \big(\mathrm{dens}(D)+\gamma\big)\,\mathrm{dens}_W(V)$ which is equivalent to saying that $\mathrm{dens}_V(D) \geqslant \mathrm{dens}(D)+\gamma$. The proof of Proposition 8.8 is thus completed. □

8.4. Applications

8.4.1. Main applications. We start by presenting a proof of Szemerédi's theorem using the density Hales–Jewett theorem. The argument can be traced in [**HJ**] and can be easily generalized, but gives very weak upper bounds for the numbers $\mathrm{Sz}(k,\delta)$.

Second proof of Theorem 7.20. We fix $k \geqslant 2$ and $0 < \delta \leqslant 1$, and we set $r = \mathrm{DHJ}(k, \delta/2)$. We will show that $\mathrm{Sz}(k,\delta) \leqslant k^r$. Let $n \geqslant k^r$ and $D \subseteq [n]$ with $|D| \geqslant \delta n$. Note that there exists an interval I of $[n]$ with $|I| = k^r$ and such that $|D \cap I| \geqslant (\delta/2)|I|$. By translating, simultaneously, the interval I and the set D, we may assume that $I = \{0,\dots,k^r-1\}$. We set $A = \{0,\dots,k-1\}$ and we define $\phi\colon A^r \to \{0,\dots,k^r-1\}$ by the rule $\phi\big((w_0,\dots,w_{r-1})\big) = \sum_{j=0}^{r-1} w_j k^j$. Observe that ϕ is a bijection and has the following property. It maps combinatorial lines of A^r to arithmetic progressions of $\{0,\dots,k^r-1\}$ of length k. By the choice of r and taking into account these remarks, the second proof of Theorem 7.20 is completed. □

The above reasoning also applies in higher dimensions. In particular, we have the following proof of the multidimensional Szemerédi theorem.

Second proof of Theorem 7.22. Let k, d be positive integers with $k \geqslant 2$ and $0 < \delta \leqslant 1$. We set $r = \mathrm{DHJ}(k^d, \delta/2^d)$ and we claim that $\mathrm{MSz}(k,d,\delta) \leqslant k^r$. Indeed, fix $n \geqslant k^r$ and let $D \subseteq [n]^d$ with $|D| \geqslant \delta n^d$. It is easy to see that there exist subintervals $I_1,\dots,I_d$ of $[n]$ with $|I_1| = \dots = |I_d| = k^r$ and such that $|D \cap (I_1 \times \dots \times I_d)| \geqslant (\delta/2^d)|I_1|\cdots|I_d|$. We may assume, of course, that $I_i = \{0,\dots,k^r-1\}$ for every $i \in [d]$. Set $A = \{0,\dots,k-1\}^d$ and observe that every $w \in A^r$ can be written as $(w_{i,j})_{i=1,j=0}^{d\ \ r-1}$ where $w_{i,j} \in \{0,\dots,k-1\}$ for every $i \in [d]$ and every $j \in \{0,\dots,r-1\}$. We define $\Phi\colon A^r \to \{0,\dots,k^r-1\}^d$ by

$$\Phi(w) = \Big(\sum_{j=0}^{r-1} w_{1,j}k^j, \dots, \sum_{j=0}^{r-1} w_{d,j}k^j\Big).$$

The map Φ is a bijection and so, by the density Hales–Jewett theorem and the choice of r, there exists a combinatorial line L of A^r such that $\Phi(L) \subseteq D$. Let F be the wildcard set of L, S the set of its fixed coordinates and $f \in A^S$ its constant part. Write $f = (f_{i,j})_{i=1,j\in S}^d$ with $f_{i,j} \in \{0,\dots,k-1\}$ for every $i \in [d]$ and every $j \in S$, and set $\mathbf{c} = (c_1,\dots,c_d) \in \mathbb{N}^d$ where $c_i = \sum_{j\in S} f_{i,j}k^j$ for every $i \in [d]$. Also let $\lambda = \sum_{j\in F} k^j$ and observe that $\Phi(L) = \big\{\mathbf{c} + \lambda\mathbf{x} : \mathbf{x} \in \{0,\dots,k-1\}^d\big\}$. The second proof of Theorem 7.22 is completed. □

The last result in this subsection is due to Furstenberg and Katznelson [**FK2**]. To state it we need to introduce some terminology. Let $r \in \mathbb{N}$ with $r \geqslant 1$ and denote by $\mathcal{F}_r$ the set of all nonempty subsets of $\{0,\dots,r-1\}$. An IP$_r$*-system* is a family $(T_\alpha)_{\alpha\in\mathcal{F}_r}$ of transformations on a nonempty set X (that is, $T_\alpha\colon X \to X$ for every $\alpha \in \mathcal{F}_r$) such that $T_{\{0\}},\dots,T_{\{r-1\}}$ are pairwise commuting, and

$$T_{\{i_0,\dots,i_m\}} = T_{\{i_0\}} \circ \dots \circ T_{\{i_m\}} \tag{8.25}$$

for every $0 \leqslant m \leqslant r-1$ and every $0 \leqslant i_0 < \dots < i_m \leqslant r-1$. Notice that if $\alpha,\beta \in \mathcal{F}_r$ with $\alpha \cap \beta = \emptyset$, then $T_\alpha \circ T_\beta = T_{\alpha\cup\beta}$. Two IP$_r$-systems $(T_\alpha)_{\alpha\in\mathcal{F}_r}$ and $(S_\alpha)_{\alpha\in\mathcal{F}_r}$ of transformations on the same set X are called *commuting* if $S_\beta \circ T_\alpha = T_\alpha \circ S_\beta$ for every $\alpha,\beta \in \mathcal{F}_r$. Also recall that a *measure preserving transformation* on a probability space (X,Σ,μ) is a measurable map $T\colon X \to X$ with the property that $\mu\big(T^{-1}(A)\big) = \mu(A)$ for every $A \in \Sigma$.

Theorem 8.15. *For every positive integer k and every $0 < \delta \leqslant 1$ there exist a positive integer* IP-Sz(k,δ) *and a strictly positive constant $\eta(k,\delta)$ with the following property. Let $r \geqslant$ IP-Sz(k,δ) and let $(T_\alpha^{(1)})_{\alpha\in\mathcal{F}_r},\dots,(T_\alpha^{(k)})_{\alpha\in\mathcal{F}_r}$ be commuting* IP$_r$*-systems of measure preserving transformations on a probability space (X,Σ,μ). If $D \in \Sigma$ with $\mu(D) \geqslant \delta$, then there exists $\alpha \in \mathcal{F}_r$ such that*

$$\mu\big(D \cap {T_\alpha^{(1)}}^{-1}(D) \cap \dots \cap {T_\alpha^{(k)}}^{-1}(D)\big) \geqslant \eta(k,\delta). \tag{8.26}$$

Theorem 8.15 is known as the IP$_r$*-Szemerédi theorem* and is a far-reaching extension of the multidimensional Szemerédi theorem. The first effective proof of Theorem 8.15 became available as a consequence of the quantitative information on the density Hales–Jewett numbers obtained in [**P**]. We proceed to the proof.

Proof of Theorem 8.15. We fix a positive integer k and $0 < \delta \leqslant 1$. Let $n_0(k+1,\delta)$ and $\zeta(k+1,\delta)$ be as in (8.3) and set IP-Sz$(k,\delta) = n_0(k+1,\delta)$ and $\eta(k,\delta) = \zeta(k+1,\delta)$. We will show that with these choices the result follows. Indeed, let $r \geqslant$ IP-Sz(k,δ) and let $(T_\alpha^{(1)})_{\alpha\in\mathcal{F}_r},\dots,(T_\alpha^{(k)})_{\alpha\in\mathcal{F}_r}$ be commuting IP$_r$-systems of measure preserving transformations on a probability space (X,Σ,μ). We enlarge this family of commuting IP$_r$-systems by adding the IP$_r$-system $(T_\alpha^{(0)})_{\alpha\in\mathcal{F}_r}$ where $T_\alpha^{(0)}$ is the identity on X for every $\alpha \in \mathcal{F}_r$. We set $A = \{0,\dots,k\}$ and for every $w = (w_0,\dots,w_{r-1}) \in A^r$ we define $R_w = T_{\{0\}}^{(w_0)} \circ T_{\{1\}}^{(w_1)} \circ \dots \circ T_{\{r-1\}}^{(w_{r-1})}$. Notice that R_w is a measure preserving transformation on (X,Σ,μ).

Now let $D \in \Sigma$ with $\mu(D) \geqslant \delta$. The fact that R_w is measure preserving yields that $\mu\big(R_w^{-1}(D)\big) \geqslant \delta$ for every $w \in A^r$. Hence, by Proposition 8.7 and the choice of r and $\eta(k,\delta)$, there exists a combinatorial line L of A^r such that

$$\mu\Big(\bigcap_{w\in L} R_w^{-1}(D)\Big) \geqslant \eta(k,\delta). \tag{8.27}$$

Let α be the wildcard set of the combinatorial line L, S the set of its fixed coordinates and $(f_j)_{j\in S}\in A^S$ its constant part. We set $Q=\prod_{j\in S} T_{\{j\}}^{(f_j)}$. Since the IP$_r$-systems $(T_\alpha^{(0)})_{\alpha\in\mathcal{F}_r}, (T_\alpha^{(1)})_{\alpha\in\mathcal{F}_r},\dots,(T_\alpha^{(k)})_{\alpha\in\mathcal{F}_r}$ are commuting, we have

$$\bigcap_{w\in L} R_w^{-1}(D) = \bigcap_{i=0}^{k} \big(T_\alpha^{(i)}\circ Q\big)^{-1}(D) = Q^{-1}\Big(\bigcap_{i=0}^{k} {T_\alpha^{(i)}}^{-1}(D)\Big). \tag{8.28}$$

Observe that Q is measure preserving. Therefore,

$$\begin{aligned}\mu\big(D\cap {T_\alpha^{(1)}}^{-1}(D)\cap\dots\cap {T_\alpha^{(k)}}^{-1}(D)\big) \quad &= \quad \mu\Big(\bigcap_{i=0}^{k} {T_\alpha^{(i)}}^{-1}(D)\Big)\\ &= \quad \mu\Big(Q^{-1}\Big(\bigcap_{i=0}^{k} {T_\alpha^{(i)}}^{-1}(D)\Big)\Big)\\ &\overset{(8.28)}{=} \quad \mu\Big(\bigcap_{w\in L} R_w^{-1}(D)\Big) \overset{(8.27)}{\geqslant} \eta(k,\delta)\end{aligned}$$

and the proof of Theorem 8.15 is completed. □

8.4.2. Combinatorial consequences for finite fields and abelian groups. In this subsection we will present some combinatorial results of geometric and algebraic nature which are due to Furstenberg and Katznelson [**FK2**]. They were originally obtained as consequences of the IP$_r$-Szemerédi theorem, but can also been proved using the hypergraph removal lemma (see, e.g., [**RSTT**]). In our exposition we will rely on the density Hales–Jewett theorem. We start with the following density version of the affine Ramsey theorem [**GLR**].

THEOREM 8.16. *For every pair q,d of positive integers and every $0<\delta\leqslant 1$ there exists a positive integer $N(q,d,\delta)$ with the following property. If $\mathbb{F}_q$ is a finite field with q elements and V is a vector space over $\mathbb{F}_q$ of dimension at least $N(q,d,\delta)$, then every $D\subseteq V$ with $|D|\geqslant\delta|V|$ contains an affine d-dimensional subspace.*

PROOF. Let q,d be a pair of positive integers and $0<\delta\leqslant 1$, and notice that we may assume that $q=p^k$ for some prime p and some positive integer k. We will show that the positive integer MDHJ(q,d,δ) is as desired. Let V be a vector space over $\mathbb{F}_q$ of dimension at least MDHJ(q,d,δ). Also let $\mathbf{v}_0,\dots,\mathbf{v}_{n-1}$ be a basis of V and $D\subseteq V$ with $|D|\geqslant\delta|V|$. We set $A=\mathbb{F}_q$ and we define $T\colon A^n\to V$ by $T(w)=\sum_{j=0}^{n-1} w_j\mathbf{v}_j$ for every $w=(w_0,\dots,w_{n-1})\in A^n$. Clearly, T is a bijection. Hence, by Proposition 8.4, there exists a d-dimensional combinatorial subspace W of A^n which contained in $T^{-1}(D)$. Let $F_0,\dots,F_{d-1}$ be the wildcard sets of W, S the set of its fixed coordinates and $(f_j)_{j\in S}\in A^S$ its constant part. Then observe that $T(W)=\mathbf{c}+U$ where $\mathbf{c}=\sum_{j\in S} f_j\,\mathbf{v}_j$ and U is the d-dimensional subspace of V generated by the vectors $\sum_{j\in F_0}\mathbf{v}_j,\dots,\sum_{j\in F_{d-1}}\mathbf{v}_j$. The proof of Theorem 8.16 is completed. □

Now let G be an abelian group (written additively) and r a positive integer. As is Subsection 8.4.1, we shall denote by $\mathcal{F}_r$ the set of all nonempty subsets of $\{0,\dots,r-1\}$. An IP_r*-set in* G is a family $(g_\alpha)_{\alpha\in\mathcal{F}_r}$ of elements of G such that $g_{\alpha\cup\beta}=g_\alpha+g_\beta$ whenever $\alpha\cap\beta=\emptyset$. Notice that $(g_\alpha)_{\alpha\in\mathcal{F}_r}$ is an IP_r-set in G if and only if $g_\alpha=\sum_{m\in\alpha}g_{\{m\}}$ for every $\alpha\in\mathcal{F}_r$. We have the following theorem.

Theorem 8.17. *For every positive integer k and every $0<\delta\leqslant 1$ there exist a positive integer $\mathrm{G}(k,\delta)$ and a strictly positive constant $\varepsilon(k,\delta)$ with the following property. Let G be an abelian group, $r\geqslant \mathrm{G}(k,\delta)$ and $(g^{(0)}_\alpha)_{\alpha\in\mathcal{F}_r},\dots,(g^{(k-1)}_\alpha)_{\alpha\in\mathcal{F}_r}$ IP_r-sets in G. Also let J be a nonempty finite subset of G such that*

$$\max\big\{|(g^{(i)}_{\{m\}}+J)\triangle J| : 0\leqslant i\leqslant k-1 \text{ and } 0\leqslant m\leqslant r-1\big\}\leqslant \varepsilon(k,\delta)|J|. \tag{8.29}$$

If $D\subseteq J$ with $|D|\geqslant\delta|J|$, then D contains a set of the form $\{g+g^{(i)}_\alpha : 0\leqslant i\leqslant k-1\}$ for some $g\in G$ and some $\alpha\in\mathcal{F}_r$.

Of course, if G is a finite abelian group, then we may set "$J=G$" and apply Theorem 8.17 directly to dense subsets of G. On the other hand, we note that if G is countable, then for every finite subset X of G and every $\varepsilon>0$ there exists a nonempty finite subset J of G such that $|(x+J)\triangle J|\leqslant\varepsilon|J|$ for every $x\in X$. This property follows from—and is in fact equivalent to—the amenability of countable abelian groups (see, e.g., [**Pat**]). Thus, Theorem 8.17 is also applicable to all countable abelian groups.

Proof of Theorem 8.17. We may assume, of course, that $k\geqslant 2$. We fix $0<\delta\leqslant 1$ and we set

$$\mathrm{G}(k,\delta)=\mathrm{DHJ}(k,\delta/2) \ \text{ and } \ \varepsilon(k,\delta)=\frac{\delta}{2\mathrm{G}(k,\delta)}. \tag{8.30}$$

We will show that with these choices the result follows. Indeed, let $r\geqslant\mathrm{G}(k,\delta)$ and let $(g^{(0)}_\alpha)_{\alpha\in\mathcal{F}_r},\dots,(g^{(k-1)}_\alpha)_{\alpha\in\mathcal{F}_r}$ be IP_r-sets in G. Also let J satisfying (8.29) and fix $D\subseteq J$ with $|D|\geqslant\delta|J|$. We need to find $g\in G$ and $\alpha\in\mathcal{F}_r$ such that the set $\{g+g^{(i)}_\alpha : 0\leqslant i\leqslant k-1\}$ is contained in D. Clearly, we may assume that

$$r=\mathrm{G}(k,\delta).$$

Let $A=\{0,\dots,k-1\}$ and for every $w=(w_0,\dots,w_{r-1})\in A^r$ set

$$s_w=\sum_{m=0}^{r-1}g^{(w_m)}_{\{m\}} \ \text{ and } \ D_w=(D-s_w)\cap J.$$

We need the following simple fact in order to estimate the size of each D_w.

Fact 8.18. *Let n be a positive integer. If $h_0,\dots,h_{n-1}\in G$ and F is a nonempty finite subset of G, then we have*

$$\Big|\Big(\sum_{m=0}^{n-1}h_m+F\Big)\triangle F\Big|\leqslant\sum_{m=0}^{n-1}|(h_m+F)\triangle F|. \tag{8.31}$$

Now let $w = (w_0, \dots, w_{r-1}) \in A^r$. By Fact 8.18, we have

$$\begin{aligned} |(s_w + J) \triangle J| &= \Big|\Big(\sum_{m=0}^{r-1} g_{\{m\}}^{(w_m)} + J\Big) \triangle J\Big| \\ &\overset{(8.31)}{\leqslant} \sum_{m=0}^{r-1} |(g_{\{m\}}^{(w_m)} + J) \triangle J| \overset{(8.29),(8.30)}{\leqslant} (\delta/2)|J|. \end{aligned}$$

Therefore,

$$\begin{aligned} |D_w| &= |(D - s_w) \cap J| = |D \cap (J + s_w)| = |D \setminus \big((J + s_w) \triangle J\big)| \\ &\geqslant |D| - |(J + s_w) \triangle J| \geqslant (\delta/2)|J|. \end{aligned} \tag{8.32}$$

By the choice of r, there exists a combinatorial line L of A^r such that $\bigcap_{w\in L} D_w \neq \emptyset$. We fix $g' \in \bigcap_{w\in L} D_w$. Also let α be the wildcard set of L, S the set of its fixed coordinates and $(f_m)_{m\in S} \in A^S$ its constant part. If $g'' = \sum_{m\in S} g_{\{m\}}^{(f_m)}$, then

$$\bigcap_{w\in L} D_w = \Big(\bigcap_{i=0}^{k-1} (D - g_\alpha^{(i)})\Big) - g''.$$

Therefore, setting $g = g' + g''$, we see that $g + g_\alpha^{(i)} \in D$ for every $i \in \{0, \dots, k-1\}$. The proof of Theorem 8.17 is completed. $\square$

Theorem 8.17 yields the following beautiful refinement of Szemerédi's theorem.

Corollary 8.19. *For every integer $k \geqslant 2$ and every $0 < \delta \leqslant 1$ there exist two positive integers $r = r(k, \delta)$ and $N_0 = N_0(k, \delta)$ with the following property. If $(\lambda_m)_{m=0}^{r-1}$ is a finite sequence in $\mathbb{Z}$ and $n \geqslant N_0 \cdot \max\{|\lambda_m| : 0 \leqslant m < r\}$, then every $D \subseteq [n]$ with $|D| \geqslant \delta n$ contains an arithmetic progression of length k whose common difference is of the form $\sum_{m\in\alpha} \lambda_m$ for some nonempty $\alpha \subseteq \{0, \dots, r-1\}$.*

Proof. Fix an integer $k \geqslant 2$ and $0 < \delta \leqslant 1$, and let $\mathrm{G}(k, \delta)$ and $\varepsilon(k, \delta)$ be as in Theorem 8.17. We set $r = \mathrm{G}(k, \delta)$ and $N_0 = \lceil 2k/\varepsilon(k, \delta)\rceil$. We claim that with these choices the result follows. Indeed, let $(\lambda_m)_{m=0}^{r-1}$ be a finite sequence in $\mathbb{Z}$. For every $i \in \{0, \dots, k-1\}$ and every $\alpha \in \mathcal{F}_r$ let $\lambda_\alpha^{(i)} = i \cdot \sum_{m\in\alpha} \lambda_m$ and notice that $(\lambda_\alpha^{(i)})_{\alpha\in\mathcal{F}_r}$ is an IP_r-set in $\mathbb{Z}$. Also let $n \geqslant N_0 \cdot \max\{|\lambda_m| : 0 \leqslant m < r\}$ and observe that, by the choice of N_0, the set $(\lambda_{\{m\}}^{(i)} + [n]) \triangle [n]$ has cardinality at most $\varepsilon(k, \delta) \cdot n$ for every $i \in \{0, \dots, k-1\}$ and every $m \in \{0, \dots, r-1\}$. Therefore, by Theorem 8.17, every subset D of $[n]$ with $|D| \geqslant \delta n$ contains a set of the form $\{a + \lambda_\alpha^{(i)} : 0 \leqslant i \leqslant k-1\} = \big\{a + i \cdot \sum_{m\in\alpha} \lambda_m : 0 \leqslant i \leqslant k-1\big\}$ for some $a \in \mathbb{Z}$ and some $\alpha \in \mathcal{F}_r$. The proof of Corollary 8.19 is completed. $\square$

Recall that for every positive integer d and every $\mathbf{u} = (u_1, \dots, u_d) \in \mathbb{R}^d$ we set

$$\|\mathbf{u}\|_\infty = \max\big\{|u_1|, \dots, |u_d|\big\}.$$

We have the following multidimensional version of Corollary 8.19. It follows from Theorem 8.17 arguing precisely as above.

COROLLARY 8.20. *For every pair k,d of positive integers with $k \geqslant 2$ and every $0<\delta\leqslant 1$ there exist two positive integers $r=r(k,d,\delta)$ and $N_0=N_0(k,d,\delta)$ with the following property. Let $\{\mathbf{u}_0,\dots,\mathbf{u}_{k-1}\}\subseteq\mathbb{Z}^d$ and $(\lambda_\alpha)_{\alpha\in\mathcal{F}_r}$ an IP_r-set in $\mathbb{Z}$. Also let $n\geqslant N_0\cdot\max\{\|\mathbf{u}_0\|_\infty,\dots,\|\mathbf{u}_{k-1}\|_\infty\}\cdot\max\{|\lambda_{\{j\}}|:0\leqslant j<r\}$. Then every $D\subseteq[n]^d$ with $|D|\geqslant\delta n^d$ contains a set of the form $\{\mathbf{c}+\lambda_\alpha\mathbf{u}_i:0\leqslant i\leqslant k-1\}$ for some $\mathbf{c}\in\mathbb{Z}^d$ and some $\alpha\in\mathcal{F}_r$.*

8.4.3. Measurable events indexed by combinatorial spaces. Our last application is an extension of Proposition 8.7. Specifically, for every $0<\delta\leqslant 1$ set $\zeta(1,\delta)=\delta$ and let $\zeta(p,\delta)$ be as in (8.3) if p is an integer with $p\geqslant 2$. We have the following theorem due to Dodos, Kanellopoulos and Tyros [**DKT4**].

THEOREM 8.21. *For every pair k,m of positive integers with $k\geqslant 2$ and every $0<\delta\leqslant 1$ there exists a positive integer $\mathrm{CorSp}(k,m,\delta)$ with the following property. If A is an alphabet with $|A|=k$, then for every combinatorial space W of $A^{<\mathbb{N}}$ with $\dim(W)\geqslant\mathrm{CorSp}(k,m,\delta)$ and every family $\{D_w:w\in W\}$ of measurable events in a probability space (Ω,Σ,μ) satisfying $\mu(D_w)\geqslant\delta$ for every $w\in W$, there exists an m-dimensional combinatorial subspace V of W such that for every nonempty $F\subseteq V$ we have*

$$\mu\Big(\bigcap_{w\in F}D_w\Big)\geqslant\zeta(|F|,\delta). \tag{8.33}$$

The proof of Theorem 8.21 is based on the notion of the type of a nonempty subset of a combinatorial space, introduced in Subsection 5.1.1. Recall that the definition of this invariant requires the existence of a linear order on the finite alphabet we are working with. However, in what follows, we will follow the convention in Subsection 5.1.2 and we will not refer explicitly to the linear order which is used to define the type.

We start with the following lemma.

LEMMA 8.22. *Let A be a finite alphabet with $|A|\geqslant 2$, d a positive integer and $G\subseteq A^d$ with $|G|\geqslant 2$. Let $\tau(G)$ be the type of G, and set $p=|G|$ and $m=|\tau(G)|$. Then there exists an alphabet $B\subseteq A^m$ with $|B|=p$ and a map $T\colon B^{<\mathbb{N}}\to A^{<\mathbb{N}}$ with the following properties.*

(a) *For every $n\in\mathbb{N}$ we have $T(B^n)\subseteq A^{m\cdot n}$.*
(b) *For every positive integer n and every combinatorial line L of B^n the image $T(L)=\{T(w):w\in L\}$ of L is a subset of $A^{m\cdot n}$ of type $\tau(G)$.*

PROOF. By the definition of the type, we have $\tau(G)=(\tau_i)_{i=0}^{m-1}$ where $m\in[d]$ and $\tau_i\in A^p\setminus\Delta(A^p)$ for every $i\in\{0,\dots,m-1\}$. (Here, $\Delta(A^p)$ is as in (5.1).) Fix $j\in\{0,\dots,p-1\}$ and for every $i\in\{0,\dots,m-1\}$ let $\tau_{i,j}\in A$ be the j-th coordinate of τ_i. We set $\beta_j=(\tau_{i,j})_{i=0}^{m-1}\in A^m$ and we define

$$B=\{\beta_0,\dots,\beta_{p-1}\}. \tag{8.34}$$

We proceed to define the map T which is a variant of the map T in Definition 2.10. Specifically, let $t\in B^{<\mathbb{N}}$ be arbitrary. If $t=\emptyset$, then we set $T(\emptyset)=\emptyset$. Otherwise, write $t=(t_\ell)_{\ell=0}^{n-1}\in B^n\subseteq (A^m)^n$. For every $i\in\{0,\dots,m-1\}$ let $t_{i,\ell}\in A$ be the i-th coordinate of t_ℓ and set

$$T(t)=(t_{0,0},\dots,t_{0,n-1})^\frown\dots^\frown(t_{m-1,0},\dots,t_{m-1,n-1}). \tag{8.35}$$

We claim that B and T are as desired. Indeed, notice first that $T(B^n)\subseteq A^{m\cdot n}$ for every $n\in\mathbb{N}$. To see that the second part of the lemma is also satisfied, fix a positive integer n and let L be a combinatorial line of B^n. Let v be the variable word over B of length n such that $L=\{v(\beta):\beta\in B\}$. Also let X be the wildcard set of L, S the set of its fixed coordinates and $(f_s)_{s\in S}\in B^S$ its constant part. For every $i\in\{0,\dots,m-1\}$ and every $s\in S$ let $f_{i,s}\in A$ be the i-th coordinate of f_s. We set $X_i=\{x+i\cdot n: x\in X\}$ and $S_i=\{s+i\cdot n: s\in S\}$, and we observe that

$$X_i\cap S_i=\emptyset \ \text{ and } \ X_i\cup S_i=\{i\cdot n,\dots,(i+1)\cdot n-1\}. \tag{8.36}$$

It follows, in particular, that for every $r\in\{0,\dots,m\cdot n-1\}$ there exists a unique $i(r)\in\{0,\dots,m-1\}$ such that $r\in\{i(r)\cdot n,\dots,(i(r)+1)\cdot n-1\}=X_{i(r)}\cup S_{i(r)}$. If, in addition, $r\in S_{i(r)}$, then there exists a unique $s(r)\in S$ such that $r=s(r)+i(r)\cdot n$. Now, fix $j\in\{0,\dots,p-1\}$ and write $T\big(v(\beta_j)\big)=(a_r^j)_{r=0}^{m\cdot n-1}\in A^{m\cdot n}$. By the definition of T, for every $r\in\{0,\dots,m\cdot n-1\}$ we have

$$a_r^j=\begin{cases}\tau_{i(r),j} & \text{if } r\in X_{i(r)}\\ f_{i(r),s(r)} & \text{if } r\in S_{i(r)}\end{cases}$$

and, therefore,

$$(a_r^0,\dots,a_r^{p-1})=\begin{cases}(\tau_{i(r),0},\dots,\tau_{i(r),p-1})=\tau_{i(r)} & \text{if } r\in X_{i(r)}\\ f^p_{i(r),s(r)}\in\Delta(A^p) & \text{if } r\in S_{i(r)}.\end{cases} \tag{8.37}$$

Finally, by (8.36), we see that $\max(X_i)<\min(X_{i+1})$ for every $i\in\{0,\dots,m-2\}$ provided, of course, that $m\geqslant 2$. Using this fact and (8.37), we conclude that $T(L)$ and G have the same type. The proof of Lemma 8.22 is completed. $\square$

We proceed with the following lemma.

LEMMA 8.23. *Let $0<\delta\leqslant 1$ and A a finite alphabet with $|A|\geqslant 2$. Also let U be a combinatorial space of $A^{<\mathbb{N}}$ and $\{D_u: u\in U\}$ a family of measurable events in a probability space (Ω,Σ,μ) satisfying $\mu(D_u)\geqslant\delta$ for every $u\in U$. Finally, let d be a positive integer and $G\subseteq A^d$ with $|G|\geqslant 2$. Assume that*

$$\dim(U)\geqslant|\tau(G)|\cdot \mathrm{DHJ}(|G|,\delta/2) \tag{8.38}$$

where $\tau(G)$ is the type of G. Then there exists $H\subseteq U$ with $\tau(H)=\tau(G)$ such that

$$\mu\Big(\bigcap_{u\in H} D_u\Big)\geqslant\zeta(|G|,\delta)$$

where $\zeta(|G|,\delta)$ is as in (8.3).

PROOF. We set $p=|G|$, $m=|\tau(G)|$, $n_0=\mathrm{DHJ}(p,\delta/2)$ and $N=\dim(U)$. Also fix $\alpha\in A$ and let $B\subseteq A^m$ and $T\colon B^{<\mathbb{N}}\to A^{<\mathbb{N}}$ be as in Lemma 8.22 when applied to the set G. We define $\Phi\colon B^{n_0}\to U$ by the rule

$$\Phi(t)=\mathrm{I}_U\big(T(t)^\frown\alpha^{N-m\cdot n_0}\big)$$

where I_U is the canonical isomorphism associated with U (see Definition 1.2) and $\alpha^{N-m\cdot n_0}$ is as in (2.1). (By Lemma 8.22, we have $T(t)\in A^{m\cdot n_0}$ and so the map Φ is well-defined.) Next, we set $D'_t = D_{\Phi(t)}$ for every $t\in B^{n_0}$ and we observe that $\mu(D'_t)\geqslant\delta$. Since $|B|=p$, by the choice of n_0 and Proposition 8.7, there exists a combinatorial line L of B^{n_0} such that

$$\mu\Big(\bigcap_{t\in L} D'_t\Big)\geqslant\zeta(p,\delta). \tag{8.39}$$

We will show that the set $H=\Phi(L)$ is as desired. Indeed, by the definition of Φ, we have $H=\big\{\mathrm{I}_U\big(T(t)^\frown\alpha^{N-m\cdot n_0}\big): t\in L\big\}$ and so, by Lemma 5.1, we obtain that

$$\tau(H)=\tau\big(\{T(t)^\frown\alpha^{N-m\cdot n_0}: t\in L\}\big). \tag{8.40}$$

Next observe that

$$\tau\big(\{T(t)^\frown\alpha^{N-m\cdot n_0}: t\in L\}\big)=\tau\big(T(L)\big). \tag{8.41}$$

On the other hand, by Lemma 8.22, we have

$$\tau\big(T(L)\big)=\tau(G). \tag{8.42}$$

By (8.40)–(8.42), we see that $\tau(H)=\tau(G)$. Finally, notice that

$$\mu\Big(\bigcap_{u\in H} D_u\Big)=\mu\Big(\bigcap_{t\in L} D'_t\Big)\overset{(8.39)}{\geqslant}\zeta(p,\delta)$$

and the proof of Lemma 8.23 is completed. $\square$

The last ingredient of the proof of Theorem 8.21 is the following estimate for the "density Hales–Jewett numbers" which is of independent interest.

Lemma 8.24. *For every $k\in\mathbb{N}$ with $k\geqslant 2$ and every $0<\delta\leqslant 1$ we have*

$$\mathrm{DHJ}(k,\delta)\leqslant\mathrm{DHJ}(k+1,\delta). \tag{8.43}$$

Proof. Let $k\in\mathbb{N}$ with $k\geqslant 2$ and $0<\delta\leqslant 1$, and let A be an alphabet with $|A|=k+1$. We select an element $\alpha\in A$ and we set $B=A\setminus\{\alpha\}$. For every $b\in B$ we define $\pi_b\colon A\to B$ by the rule $\pi_b(\alpha)=b$ and $\pi_b(\beta)=\beta$ if $\beta\in B$. More generally, for every positive integer n and every $w=(w_0,\dots,w_{n-1})\in B^n$ we define a map $\pi_w\colon A^n\to B^n$ by setting

$$\pi_w\big((a_0,\dots,a_{n-1})\big)=\big(\pi_{w_0}(a_0),\dots,\pi_{w_{n-1}}(a_{n-1})\big). \tag{8.44}$$

Notice that π_w is a surjection, and so it induces a probability measure μ_w on B^n defined by

$$\mu_w(X)=\mathrm{dens}_{A^n}\big(\pi_w^{-1}(X)\big) \tag{8.45}$$

for every $X\subseteq B^n$. We have the following claim.

Claim 8.25. *The following hold.*

(a) *For every $w\in B^n$ and every combinatorial line L of A^n the set $\pi_w(L\upharpoonright B)$ is a combinatorial line of B^n where $L\upharpoonright B$ is as in (1.21).*
(b) *For every $X\subseteq B^n$ we have $\mathrm{dens}_{B^n}(X)=\mathbb{E}_{w\in B^n}\mu_w(X)$.*

Proof of Claim 8.25. Part (a) is an immediate consequence of the relevant definitions. For part (b) it is enough to show that $\mathbb{E}_{w\in B^n}\mu_w(\{y\}) = k^{-n}$ for every $y = (y_0,\dots,y_{n-1}) \in B^n$. To this end, fix $y = (y_0,\dots,y_{n-1}) \in B^n$ and for every $w = (w_0,\dots,w_{n-1}) \in B^n$ let $\Delta(w,y) = \big\{i \in \{0,\dots,n-1\} : w_i = y_i\big\}$. Note that

$$\mu_w(\{y\}) = \frac{2^{|\Delta(w,y)|}}{(k+1)^n}. \tag{8.46}$$

Also observe that for every $F \subseteq \{0,\dots,n-1\}$ we have

$$|\{w \in B^n : \Delta(w,y) = F\}| = (k-1)^{n-|F|}. \tag{8.47}$$

Therefore,

$$\begin{aligned}\mathbb{E}_{w\in B^n}\mu_w(\{y\}) &\overset{(8.46)}{=} k^{-n}(k+1)^{-n}\sum_{w\in B^n} 2^{|\Delta(w,y)|}\\ &\overset{(8.47)}{=} k^{-n}(k+1)^{-n}\sum_{i=0}^{n}\binom{n}{i}(k-1)^{n-i}2^i = k^{-n}\end{aligned}$$

and the proof of Claim 8.25 is completed. □

Now let $n \geqslant \mathrm{DHJ}(k+1,\delta)$ and fix a subset D of B^n with $\mathrm{dens}_{B^n}(D) \geqslant \delta$. By Claim 8.25, there exists $w \in B^n$ such that $\mu_w(D) \geqslant \mathrm{dens}_{B^n}(D)$. This implies that $\mathrm{dens}_{A^n}\big(\pi_w^{-1}(D)\big) \geqslant \delta$ and so, by the choice of n, the set $\pi_w^{-1}(D)$ contains a combinatorial line L of A^n. Invoking Claim 8.25 once again, we conclude that the set $\pi_w(L \upharpoonright B)$ is a combinatorial line of B^n which is contained in D. This shows that the estimate in (8.43) is satisfied and the proof of Lemma 8.24 is completed. □

We are now ready to give the proof of Theorem 8.21.

Proof of Theorem 8.21. Fix a pair k,m of positive integers with $k \geqslant 2$ and $0 < \delta \leqslant 1$, and set $d = m \cdot \mathrm{DHJ}(k^m,\delta/2)$. We will show that

$$\mathrm{CorSp}(k,m,\delta) \leqslant \mathrm{RamSp}(k,d,2) \tag{8.48}$$

where $\mathrm{RamSp}(k,d,2)$ is as in Theorem 5.5.

Let $n \geqslant \mathrm{RamSp}(k,d,2)$ and let A be an alphabet with $|A| = k$. Also let W be an n-dimensional combinatorial space of $A^{<\mathbb{N}}$ and $\{D_w : w \in W\}$ a family of measurable events in a probability space (Ω,Σ,μ) satisfying $\mu(D_w) \geqslant \delta$ for every $w \in W$. We define a coloring $c\colon \mathcal{P}(W) \to [2]$ by setting $c(F) = 1$ if F is nonempty and $\mu\big(\bigcap_{w\in F} D_w\big) \geqslant \zeta(|F|,\delta)$. By the choice of n and Theorem 5.5, there exists a d-dimensional combinatorial subspace U of W such that every pair of nonempty subsets of U with the same type is monochromatic. We select $V \in \mathrm{Subsp}_m(U)$ and we claim that V is as desired. Indeed, let F be a nonempty subset of V and observe that, by the definition of the coloring c, it is enough to show that $c(F) = 1$. Set $p = |F|$ and let $\tau(F)$ be the type of F. Clearly, we may assume that $p \geqslant 2$. Also notice that $|\tau(F)| \leqslant m$ and $p \leqslant k^m$. We set $G = \mathrm{I}_U^{-1}(F)$ where I_U is the canonical isomorphism associated with U. By Lemma 5.1, we see that $G \subseteq A^d$ and $\tau(G) = \tau(F)$. Therefore,

$$\dim(U) = d = m \cdot \mathrm{DHJ}(k^m,\delta/2) \geqslant |\tau(G)| \cdot \mathrm{DHJ}(p,\delta/2)$$

where the last inequality follows from Lemma 8.24. By Lemma 8.23, there exists $H \subseteq U$ with $\tau(H) = \tau(G)$ and such that $\mu\big(\bigcap_{u\in H} D_u\big) \geqslant \zeta(p,\delta)$. It follows, in particular, that $c(H) = 1$ and $\tau(H) = \tau(F)$. Since every pair of nonempty subsets of U with the same type is monochromatic, we conclude that $c(F) = 1$ and the proof of Theorem 8.21 is completed. $\square$

We close this section with the following extension[1] of Corollary 8.6.

Corollary 8.26. *Let $0 < \delta \leqslant 1$, A a finite alphabet with $|A| \geqslant 2$, d a positive integer and G a subset of A^d with $|G| \geqslant 2$. Let $\tau(G)$ be the type of G and set $M = d \cdot \mathrm{DHJ}(|G|, \delta/4)$. If W is a combinatorial space of $A^{<\mathbb{N}}$ with*

$$\dim(W) \geqslant 2\delta^{-1}|A|^M M, \tag{8.49}$$

then every $D \subseteq W$ with $\mathrm{dens}_W(D) \geqslant \delta$ contains a set F with $\tau(F) = \tau(G)$.

Proof. We may assume that W is of the form A^n for some $n \geqslant \lceil 2\delta^{-1}|A|^M M\rceil$. By Lemma 8.5, there exist an integer l with $M \leqslant l < N$ and an M-dimensional combinatorial subspace U of A^l such that for every $u \in U$ we have $\mathrm{dens}(D_u) \geqslant \delta/2$ where $D_u = \{y \in A^{n-l} : u^\frown y \in D\}$ is the section of D at u. Since $|\tau(G)| \leqslant d$, by the choice of M and Lemma 8.23, there exists $H \subseteq U$ with $\tau(H) = \tau(G)$ and such that $\mu\big(\bigcap_{u\in H} D_u\big) \geqslant \zeta(|H|, \delta/2)$. In particular, the set $\bigcap_{u\in H} D_u$ is nonempty. We select $y \in \bigcap_{u\in H} D_u$ and we set $F = \{w^\frown y : w \in H\}$. Clearly, F is as desired. The proof of Corollary 8.26 is completed. $\square$

8.5. Notes and remarks

The density Hales–Jewett theorem was, undoubtedly, the culmination of the ergodic-theoretic methods gradually developed by Furstenberg and Katznelson in [**F, FK1, FK2**]. The original proof was also based on several results from coloring Ramsey theory, including the Carlson–Simpson theorem and Carlson's theorem. A different ergodic proof was given in [**A**].

The first combinatorial proof of the density Hales–Jewett theorem was discovered by Polymath [**P**]. This proof also yields the best known upper bounds for the numbers $\mathrm{DHJ}(k,\delta)$. Another combinatorial proof was outlined by Tao in [**Tao4**]. Tao's approach was motivated by the graph-theoretic proofs of Szemerédi's theorem and was based on a variant of Corollary 6.13. The proof we presented is due to Dodos, Kanellopoulos and Tyros [**DKT2**]. It was found in the course of obtaining a density version of the Carlson–Simpson theorem—a result that we will discuss in detail in Chapter 9—and gives essentially the same upper bounds for the numbers $\mathrm{DHJ}(k,\delta)$ as in Polymath's proof. However, these upper bounds are admittedly weak and have an Ackermann-type dependence with respect to k. It is one of the central open problems of Ramsey theory to decide whether the numbers $\mathrm{DHJ}(k,\delta)$ are upper bounded by a primitive recursive function.

[1] The main point in Corollary 8.26 is, of course, the estimate in (8.49). Notice, in particular, that if the cardinality of the set G is fixed, then the lower bound in (8.49) is controlled by a primitive recursive function of the parameters δ, $|A|$ and d.

CHAPTER 9

The density Carlson–Simpson theorem

The following result is the density version of the Carlson–Simpson theorem and is due to Dodos, Kanellopoulos and Tyros [**DKT3**].

THEOREM 9.1. *Let A be a finite alphabet with $|A| \geqslant 2$. Then for every set D of words over A satisfying*

$$\limsup_{n\to\infty} \frac{|D\cap A^n|}{|A^n|} > 0$$

there exist a word w over A and a sequence (u_n) of left variable words over A such that the set

$$\{w\} \cup \big\{w^\frown u_0(a_0)^\frown \dots ^\frown u_n(a_n) : n\in\mathbb{N} \text{ and } a_0,\dots,a_n\in A\big\}$$

is contained in D.

Although Theorem 9.1 is genuinely infinite-dimensional, it will be reduced to an appropriate finite version. This finite version is the content of the following theorem whose proof will occupy the bulk of this chapter. General facts about Carlson–Simpson spaces can be found in Section 1.5.

THEOREM 9.2. *For every pair k,m of positive integers with $k\geqslant 2$ and every $0<\delta\leqslant 1$ there exists a positive integer N with the following property. If A is an alphabet with $|A|=k$, L is a finite subset of $\mathbb{N}$ of cardinality at least N and D is a set of words over A satisfying $|D\cap A^n|\geqslant \delta|A^n|$ for every $n\in L$, then D contains an m-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$. The least positive integer with this property will be denoted by $\mathrm{DCS}(k,m,\delta)$.*

The main point in Theorem 9.2 is that the result is independent of the position of the finite set L. Note, in particular, that this structural property does not follow from Theorem 9.1 with standard arguments based on compactness.

We now briefly describe the contents of this chapter. The first six sections are devoted to the proof of Theorem 9.2. Sections 9.1 and 9.2 contain, mostly, supporting material and the main part of the argument is given in Sections 9.3, 9.4 and 9.5. The proof of Theorem 9.2 is completed in Section 9.6. It proceeds by induction on the cardinality of the finite alphabet A and is based on a density increment strategy. The argument is effective and yields explicit, albeit weak, upper bounds for the numbers $\mathrm{DCS}(k,m,\delta)$. Specific features of the proof are discussed in Subsections 9.4.1 and 9.5.1. Finally, in Section 9.7 we derive Theorem 9.1 from Theorem 9.2 while in Section 9.8 we present some applications.

9.1. The convolution operation

9.1.1. Definitions. We are about to introduce a method of "gluing" a pair of words over a nonempty finite alphabet A. This method can be thought of as a natural extension of the familiar operation of concatenation, and is particularly easy to grasp for pairs of words of given length. Specifically, let n, m be two positive integers and fix a subset L of $\{0,\dots,n+m-1\}$ of cardinality n. Given an element x of A^n and an element y of A^m, the outcome of the "gluing" method for the pair x, y is the unique element z of A^{n+m} which is "equal" to x on L and to y on the rest of the coordinates. This simple process can be extended to arbitrary pairs of words over A.

DEFINITION 9.3. *Let A be a finite alphabet with $|A| \geqslant 2$ and let L be a nonempty finite subset of $\mathbb{N}$. We set*

$$n_L = \max(L) - |L| + 1 \quad \textit{and} \quad \Omega_L = A^{n_L}. \tag{9.1}$$

Also let $l_0 < \dots < l_{|L|-1}$ be the increasing enumeration of the set L and for every $i \in \{0,\dots,|L|-1\}$ set

$$L_i = \{l \in L : l < l_i\}, \quad K_i = \{n \in \mathbb{N} : n < l_i \textit{ and } n \notin L_i\} \quad \textit{and} \quad \kappa_i = |K_i|. \tag{9.2}$$

We define the convolution operation $\mathrm{c}_L \colon A^{<|L|} \times \Omega_L \to A^{<\mathbb{N}}$ associated with L *as follows. For every $i \in \{0,\dots,|L|-1\}$, every $t \in A^i$ and every $\omega \in \Omega_L$ we set*

$$\mathrm{c}_L(t,\omega) = \big(\mathrm{I}_{L_i}(t), \mathrm{I}_{K_i}(\omega \upharpoonright \kappa_i)\big) \in A^{l_i} \tag{9.3}$$

where I_{L_i} and I_{K_i} are the canonical isomorphisms associated with the sets L_i and K_i respectively (*see Definition* 1.1)*.*

More generally, let V be a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ such that $L \subseteq \{0,\dots,\dim(V)\}$. The convolution operation $\mathrm{c}_{L,V} \colon A^{<|L|} \times \Omega_L \to V$ associated with (L,V) *is defined by the rule*

$$\mathrm{c}_{L,V}(t,\omega) = \mathrm{I}_V\big(\mathrm{c}_L(t,\omega)\big) \tag{9.4}$$

where I_V is the canonical isomorphism associated with V (*see Definition* 1.10)*.*

For a specific example, consider the alphabet $A = \{a,b,c,d,e\}$ and let L be the set $\{1,3,7,9\}$. Notice that $n_L = 6$ and $\Omega_L = A^6$. In particular, the convolution operation c_L associated with L is defined for pairs in $A^{<4} \times A^6$. Then for the pair $t = (a,b)$ and $\omega = (c,e,d,b,d,a)$ we have

$$\mathrm{c}_L(t,\omega) = (c, \mathbf{a}, e, \mathbf{b}, d, b, d) \tag{9.5}$$

where in (9.5) we indicated with boldface letters the contribution of t.

We also note the asymmetric role of $A^{<|L|}$ and Ω_L in Definition 9.3. Indeed, while the set $A^{<|L|}$ is the structured part of the domain of c_L, the set Ω_L will be regarded merely as a "sample" space. Specifically, we will view Ω_L as a discrete probability space equipped with the uniform probability measure. Because of this asymmetricity, we may consider the convolution operation as a noncommutative analogue of concatenation.

9.1.2. Basic properties. In this subsection we will present some basic properties of convolution operations. We start with the following fact.

FACT 9.4. *Let A be a finite alphabet with $|A| \geqslant 2$, V a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ and $L = \{l_0 < \dots < l_{|L|-1}\}$ a nonempty finite subset of $\{0, \dots, \dim(V)\}$. For every $t \in A^{<|L|}$ we set*

$$\mathcal{C}_t = \{\mathrm{c}_{L,V}(t, \omega) : \omega \in \Omega_L\}. \tag{9.6}$$

Then for every $t, t' \in A^{<|L|}$ with $t \neq t'$ we have $\mathcal{C}_t \cap \mathcal{C}_{t'} = \emptyset$. Moreover, for every $i \in \{0, \dots, |L|-1\}$ the family $\{\mathcal{C}_t : t \in A^i\}$ forms an equipartition of $V(l_i)$.

PROOF. Let $i \in \{0, \dots, |L|-1\}$ and $t \in A^i$. By (9.6) and the definition of the convolution operation, we see that

$$\mathcal{C}_t \stackrel{(9.4)}{=} \mathrm{I}_V\big(\{\mathrm{c}_L(t,\omega) : \omega \in \Omega_L\}\big) \stackrel{(9.3)}{=} \mathrm{I}_V\big(\{z \in A^{l_i} : z \upharpoonright L_i = \mathrm{I}_{L_i}(t)\}\big) \tag{9.7}$$

and the proof of Fact 9.4 is completed. □

The next fact is the "dual" version of Fact 9.4.

FACT 9.5. *Let A, V and L be as in Fact 9.4. For every $t \in A^{<|L|}$ and every $s \in A^{<\mathbb{N}}$ we set*

$$\Omega_t^s = \{\omega \in \Omega_L : \mathrm{c}_{L,V}(t,\omega) = s\}. \tag{9.8}$$

Then for every $t \in A^{<|L|}$ and every $s, s' \in \mathcal{C}_t$ with $s \neq s'$, where $\mathcal{C}_t$ is as in (9.6), the sets Ω_t^s and $\Omega_t^{s'}$ are nonempty disjoint subsets of Ω_L. Moreover, the family $\{\Omega_t^s : s \in \mathcal{C}_t\}$ forms an equipartition of Ω_L.

PROOF. First observe that the set Ω_t^s is a nonempty subset of Ω_L for every $t \in A^{<|L|}$ and every $s \in \mathcal{C}_t$. Also notice that if $s' \in \mathcal{C}_t$ with $s' \neq s$, then $\Omega_t^s \cap \Omega_t^{s'} = \emptyset$.

Next, fix $t \in A^i$ for some $i \in \{0, \dots, |L|-1\}$. By (9.6), it is clear that the family $\{\Omega_t^s : s \in \mathcal{C}_t\}$ is a partition of Ω_L, and so we only have to show that this family is actually an equipartition. To this end, recall that $l_0 < \dots < l_{|L|-1}$ is the increasing enumeration of the set L. Let $s \in \mathcal{C}_t$ and let z be the unique element of A^{l_i} such that $\mathrm{I}_V(z) = s$. Then observe that

$$\Omega_t^s = \{\omega \in \Omega_L : \mathrm{I}_{K_i}(\omega \upharpoonright \kappa_i) = z \upharpoonright K_i\} \tag{9.9}$$

which implies that $|\Omega_t^s| = |A|^{n_L - \kappa_i}$. The proof of Fact 9.5 is completed. □

Now let A, V and L be as in Fact 9.4. Also let $i \in \{0, \dots, |L|-1\}$ and $t, t' \in A^i$. For every $s \in \mathcal{C}_t$ we select $\omega_s \in \Omega_t^s$ (where $\mathcal{C}_t$ and Ω_s^t are as in (9.6) and (9.8) respectively) and we define

$$r_{t,t'}(s) = \mathrm{c}_{L,V}(t', \omega_s). \tag{9.10}$$

Notice, first, that $r_{t,t'}(s)$ is independent of the choice of ω_s. Indeed, if z is the unique element of A^{l_i} such that $\mathrm{I}_V(z) = s$, then we have

$$\begin{aligned} r_{t,t'}(s) &\stackrel{(9.10)}{=} \mathrm{c}_{L,V}(t', \omega_s) \stackrel{(9.4)}{=} \mathrm{I}_V\big(\mathrm{c}_L(t', \omega_s)\big) \\ &\stackrel{(9.3)}{=} \mathrm{I}_V\Big(\big(\mathrm{I}_{L_i}(t'), \mathrm{I}_{K_i}(\omega_s \upharpoonright \kappa_i)\big)\Big) = \mathrm{I}_V\Big(\big(\mathrm{I}_{L_i}(t'), z \upharpoonright K_i\big)\Big). \end{aligned} \tag{9.11}$$

Next observe that, by (9.6) and (9.10), we have $r_{t,t'}(s) \in \mathcal{C}_{t'}$. It follows, in particular, that the assignment $\mathcal{C}_t \ni s \mapsto r_{t,t'}(s) \in \mathcal{C}_{t'}$ is well-defined. We gather below some properties of this map. They are all straightforward consequences of (9.6), (9.8) and (9.11).

FACT 9.6. *Let A, V and L be as in Fact 9.4. Also let $i \in \{0, \dots, |L|-1\}$ and $t, t' \in A^i$, and consider the map $r_{t,t'} \colon \mathcal{C}_t \to \mathcal{C}_{t'}$. Then the following hold.*

(a) *The map $r_{t,t'}$ is a bijection.*
(b) *For every $s \in \mathcal{C}_t$ we have $\Omega_t^s = \Omega_{t'}^{r_{t,t'}(s)}$.*
(c) *Let $a, b \in A$ with $a \neq b$ and assume that t and t' are (a,b)-equivalent (see Subsection 2.1.1). Then for every $s \in \mathcal{C}_t$ the words s and $r_{t,t'}(s)$ are (a,b)-equivalent.*

We close this subsection with the following lemma.

LEMMA 9.7. *Let A, V and L be as in Fact 9.4. Also let $D \subseteq A^{<\mathbb{N}}$ and set $\mathscr{D} = \mathrm{c}_{L,V}^{-1}(D)$. Then the following hold.*

(a) *For every $t \in A^{<|L|}$ we have $\mathrm{dens}_{\mathcal{C}_t}(D) = \mathrm{dens}_{\{t\}\times\Omega_L}(\mathscr{D})$ where $\mathcal{C}_t$ is as in (9.6).*
(b) *For every $i \in \{0, \dots, |L|-1\}$ we have $\mathrm{dens}_{V(l_i)}(D) = \mathrm{dens}_{A^i \times \Omega_L}(\mathscr{D})$.*

PROOF. Fix $t \in A^{<|L|}$ and for every $s \in \mathcal{C}_t$ let Ω_t^s be as in (9.8). By the definition of $\mathscr{D}$ and $\mathcal{C}_t$, we see that

$$\begin{aligned} \mathscr{D} \cap (\{t\} \times \Omega_L) &= \{(t,\omega) : \mathrm{c}_{L,V}(t,\omega) \in D \cap \mathcal{C}_t\} \\ &= \bigcup_{s \in D \cap \mathcal{C}_t} \{(t,\omega) : \mathrm{c}_{L,V}(t,\omega) = s\} \\ &= \bigcup_{s \in D \cap \mathcal{C}_t} \{t\} \times \Omega_t^s. \end{aligned} \tag{9.12}$$

Moreover, by Fact 9.5, for every $s \in \mathcal{C}_t$ we have $|\Omega_t^s|/|\Omega_L| = 1/|\mathcal{C}_t|$. Therefore,

$$\begin{aligned} \mathrm{dens}_{\{t\}\times\Omega_L}(\mathscr{D}) &= \frac{|\mathscr{D} \cap (\{t\} \times \Omega_L)|}{|\{t\} \times \Omega_L|} \overset{(9.12)}{=} \sum_{s \in D \cap \mathcal{C}_t} \frac{|\{t\} \times \Omega_t^s|}{|\{t\} \times \Omega_L|} \\ &= \sum_{s \in D \cap \mathcal{C}_t} \frac{|\Omega_t^s|}{|\Omega_L|} = \frac{|D \cap \mathcal{C}_t|}{|\mathcal{C}_t|} = \mathrm{dens}_{\mathcal{C}_t}(D). \end{aligned} \tag{9.13}$$

To see that the second part of the lemma is satisfied, let $i \in \{0, \dots, |L|-1\}$ be arbitrary and observe that, by Fact 9.4, we have

$$\mathrm{dens}_{V(l_i)}(D) = \mathbb{E}_{t \in A^i}\, \mathrm{dens}_{\mathcal{C}_t}(D). \tag{9.14}$$

Hence, by (9.14) and the first part of the lemma, we conclude that

$$\mathrm{dens}_{V(l_i)}(D) = \mathbb{E}_{t \in A^i}\, \mathrm{dens}_{\{t\}\times\Omega_L}(\mathscr{D}) = \mathrm{dens}_{A^i \times \Omega_L}(\mathscr{D}).$$

The proof of Lemma 9.7 is completed. $\square$

9.1.3. Coherence properties. We continue the analysis of convolution operations with the following lemma which asserts that all convolution operations preserve Carlson–Simpson spaces.

LEMMA 9.8. *Let A be a finite alphabet with $|A| \geqslant 2$, V a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ and L a nonempty finite subset of $\{0, \dots, \dim(V)\}$ with $|L| \geqslant 2$. Also let U be a Carlson–Simpson subspace of $A^{<|L|}$ and $\omega \in \Omega_L$. Then, setting*

$$U_\omega = \{\mathrm{c}_{L,V}(u, \omega) : u \in U\}, \tag{9.15}$$

we have that U_ω is a Carlson–Simpson subspace of V of dimension $\dim(U)$. Moreover, for every $i \in \{0, \dots, \dim(U)\}$ we have

$$U_\omega(i) = \{\mathrm{c}_{L,V}(u, \omega) : u \in U(i)\}. \tag{9.16}$$

PROOF. We set $X = L_{|L|-1}$ and $Y = K_{|L|-1}$ where $L_{|L|-1}$ and $K_{|L|-1}$ are as in (9.2). Notice, in particular, that $X = L \setminus \{\max(L)\}$, $Y = \{0, \dots, \max(L) - 1\} \setminus X$, $X \cup Y = \{0, \dots, \max(L) - 1\}$ and $|Y| = n_L$. Let $x_0 < \dots < x_{|L|-2}$ be the increasing enumeration of X and let W be the $(|L| - 1)$-dimensional combinatorial subspace of $A^{\max(L)}$ with wildcard sets $\{x_0\}, \dots, \{x_{|L|-2}\}$ and constant part $\mathrm{I}_Y(\omega) \in A^Y$. (Here, I_Y is the canonical isomorphism associated with Y; see Definition 1.1.) By Lemma 1.13, there exists a unique Carlson–Simpson subspace S_ω of $A^{<\max(L)+1}$ with $\dim(S_\omega) = |L| - 1$ and whose last level $S_\omega(|L| - 1)$ is W. Then observe that, by Definition 9.3, we have $\mathrm{c}_L(t, \omega) = \mathrm{I}_{S_\omega}(t)$ for every $t \in A^{<|L|}$ where I_{S_ω} stands for the canonical isomorphism associated with S_ω. Thus, we obtain that

$$\mathrm{c}_{L,V}(t, \omega) = \mathrm{I}_V\big(\mathrm{I}_{S_\omega}(t)\big) \tag{9.17}$$

for every $t \in A^{<|L|}$ which implies, of course, that U_ω is as desired. The proof of Lemma 9.8 is completed. $\square$

The next result will enable us to transfer quantitative information from the space $A^{<\mathbb{N}}$ to the space on which the convolution operations are acting.

LEMMA 9.9. *Let A, V and L be as in Lemma 9.8 and U a Carlson–Simpson subspace of $A^{<|L|}$. Also let $\omega \in \Omega_L$ and define U_ω as in (9.15). Finally, let D be a subset of $A^{<\mathbb{N}}$ and set $\mathscr{D} = \mathrm{c}_{L,V}^{-1}(D)$. Then for every $i \in \{0, \dots, \dim(U)\}$ we have*

$$\mathrm{dens}_{U_\omega(i)}(D) = \mathrm{dens}_{U(i) \times \{\omega\}}(\mathscr{D}). \tag{9.18}$$

PROOF. Let $i \in \{0, \dots, \dim(U)\}$ be arbitrary. By Fact 9.4, we have that $\mathrm{c}_{L,V}(t, \omega) \neq \mathrm{c}_{L,V}(t', \omega)$ for every $t, t' \in U_\omega(i)$ with $t \neq t'$. Hence, by (9.16),

$$|U_\omega(i)| = |U(i)| = |U(i) \times \{\omega\}|. \tag{9.19}$$

Next observe that for every $t \in A^{<|L|}$ we have $(t, \omega) \in \mathscr{D} \cap \big(U(i) \times \{\omega\}\big)$ if and only if $\mathrm{c}_{L,V}(t, \omega) \in D \cap U_\omega(i)$. Therefore,

$$|D \cap U_\omega(i)| = |\mathscr{D} \cap \big(U(i) \times \{\omega\}\big)|. \tag{9.20}$$

Combining (9.19) and (9.20), we see that (9.18) is satisfied and the proof of Lemma 9.9 is completed. $\square$

We close this subsection with the following lemma.

LEMMA 9.10. *Let A, V and L be as in Lemma 9.8 and U a Carlson–Simpson subspace of $A^{<|L|}$. Also let $D \subseteq A^{<\mathbb{N}}$. Then for every $i \in \{0, \dots, \dim(U)\}$ we have*

$$\mathrm{dens}_{\mathrm{c}_{L,V}(U(i)\times\Omega_L)}(D) = \mathbb{E}_{\omega\in\Omega_L} \mathrm{dens}_{U_\omega(i)}(D) \tag{9.21}$$

where U_ω is as in (9.15)*. In particular, for every $i \in \{0, \dots, |L|-1\}$ we have*

$$\mathrm{dens}_{V(l_i)}(D) = \mathbb{E}_{\omega\in\Omega_L} \mathrm{dens}_{R_\omega(i)}(D) \tag{9.22}$$

where $R_\omega = \{\mathrm{c}_{L,V}(t,\omega) : t \in A^{<|L|}\}$ for every $\omega \in \Omega_L$.

PROOF. Fix $i \in \{0, \dots, \dim(U)\}$. There exists a unique $l \in \{0, \dots, |L|-1\}$ such that $U(i)$ is contained in A^l. By Fact 9.4, the family $\{\mathcal{C}_t : t \in U(i)\}$ forms an equipartition of $\mathrm{c}_{L,V}(U(i)\times\Omega_L)$. Therefore, setting $\mathscr{D} = \mathrm{c}_{L,V}^{-1}(D)$, by Lemma 9.7, we obtain that

$$\begin{aligned}
\mathrm{dens}_{\mathrm{c}_{L,V}(U(i)\times\Omega_L)}(D) &= \mathbb{E}_{t\in U(i)}\, \mathrm{dens}_{\mathcal{C}_t}(D) \\
&= \mathbb{E}_{t\in U(i)}\, \mathrm{dens}_{\{t\}\times\Omega_L}(\mathscr{D}) \\
&= \mathrm{dens}_{U(i)\times\Omega_L}(\mathscr{D}) \\
&= \mathbb{E}_{\omega\in\Omega_L} \mathrm{dens}_{U(i)\times\{\omega\}}(\mathscr{D}) \\
&\overset{(9.18)}{=} \mathbb{E}_{\omega\in\Omega_L} \mathrm{dens}_{U_\omega(i)}(D).
\end{aligned} \tag{9.23}$$

Finally, notice that $V(l_i) = \mathrm{c}_{L,V}(A^i \times \Omega_L)$ for every $i \in \{0, \dots, |L|-1\}$. Hence, (9.22) follows from (9.21) and the proof of Lemma 9.10 is completed. $\square$

9.1.4. Convolution operations and regularity. The last result of this section relates the concept of (ε, L)-regularity introduced in Definition 6.23 with convolution operations. This result is, to a large extent, the main motivation for the definition of convolution operations and will be used throughout this chapter.

LEMMA 9.11. *Let A be a finite alphabet with $|A| \geqslant 2$ and $\mathcal{F}$ a family of subsets of $A^{<\mathbb{N}}$. Also let $0 < \varepsilon \leqslant 1$ and $L = \{l_0 < \dots < l_{|L|-1}\}$ a nonempty finite subset of $\mathbb{N}$ such that the family $\mathcal{F}$ is (ε, L)-regular. Finally, let $\mathrm{c}_L \colon A^{<|L|} \times \Omega_L \to A^{<\mathbb{N}}$ be the convolution operation associated with L and set $\mathscr{D} = \mathrm{c}_L^{-1}(D)$ for every $D \in \mathcal{F}$. Then for every $i \in \{0, \dots, |L|-1\}$, every $D \in \mathcal{F}$ and every $t \in A^i$ we have*

$$|\mathrm{dens}_{\Omega_L}(\mathscr{D}_t) - \mathrm{dens}_{A^{l_i}}(D)| \leqslant \varepsilon \tag{9.24}$$

where $\mathscr{D}_t = \{\omega \in \Omega_L : (t,\omega) \in \mathscr{D}\}$ is the section of $\mathscr{D}$ at t.

PROOF. We fix $i \in \{0, \dots, |L|-1\}$ and $D \in \mathcal{F}$. Let L_i and K_i be as in (9.2). The family $\mathcal{F}$ is (ε, L)-regular, and so for every $y \in A^{L_i}$ we have

$$|\mathrm{dens}\big(\{w \in A^{K_i} : (y,w) \in D \cap A^{l_i}\}\big) - \mathrm{dens}_{A^{l_i}}(D)| \leqslant \varepsilon. \tag{9.25}$$

Now let $t \in A^i$ be arbitrary. By the definition of c_L, we see that

$$\mathcal{C}_t \overset{(9.6)}{=} \{\mathrm{c}_L(t,\omega) : \omega \in \Omega_L\} = \{z \in A^{l_i} : z \upharpoonright L_i = \mathrm{I}_{L_i}(t)\}$$

where I_{L_i} stands for the canonical isomorphism associated with the set L_i. Thus, by identifying $\mathcal{C}_t$ with $\{\mathrm{I}_{L_i}(t)\} \times A^{K_i}$,

$$\mathrm{dens}_{\mathcal{C}_t}(D) = \mathrm{dens}\big(\{w \in A^{K_i} : (\mathrm{I}_{L_i}(t), w) \in D \cap A^{l_i}\}\big). \tag{9.26}$$

On the other hand, by Lemma 9.7, we have

$$\mathrm{dens}_{\mathcal{C}_t}(D) = \mathrm{dens}_{\{t\}\times\Omega_L}(\mathscr{D}) = \mathrm{dens}_{\Omega_L}(\mathscr{D}_t). \tag{9.27}$$

Combining (9.26) and (9.27), we obtain that

$$\mathrm{dens}\big(\{w \in A^{K_i} : \big(\mathrm{I}_{L_i}(t), w\big) \in D \cap A^{l_i}\}\big) = \mathrm{dens}_{\Omega_L}(\mathscr{D}_t). \tag{9.28}$$

Therefore, by (9.25) applied for "$y = \mathrm{I}_{L_i}(t)$" and (9.28), we conclude that the estimate in (9.24) is satisfied and the proof of Lemma 9.11 is completed. □

9.2. Iterated convolutions

In this section we study iterations of convolution operations. We point out that this material will be used only in Section 9.5. We start with the following definition.

DEFINITION 9.12. *Let A be a finite alphabet with $|A| \geqslant 2$, let $\mathbf{L} = (L_n)_{n=0}^d$ be a nonempty finite sequence of nonempty finite subsets of $\mathbb{N}$ and let $\mathbf{V} = (V_n)_{n=0}^d$ be a finite sequence of finite-dimensional Carlson–Simpson spaces of $A^{<\mathbb{N}}$ with the same length as $\mathbf{L}$. We say that the pair $(\mathbf{L}, \mathbf{V})$ is A-*compatible*, or simply* compatible *if A is understood, provided that for every $n \in \{0, \dots, d\}$ we have $L_n \subseteq \{0, \dots, \dim(V_n)\}$ and, if $n < d$, then $V_{n+1} \subseteq A^{<|L_n|}$.*

Observe that if $(\mathbf{L}, \mathbf{V})$ is a compatible pair and $\mathbf{L}', \mathbf{V}'$ are initial subsequences of $\mathbf{L}, \mathbf{V}$ with a common length, then the pair $(\mathbf{L}', \mathbf{V}')$ is compatible. Also notice that for every compatible pair $(\mathbf{L}, \mathbf{V}) = \big((L_n)_{n=0}^d, (V_n)_{n=0}^d\big)$ and every $n \in \{0, \dots, d\}$ we can define the convolution operation $\mathrm{c}_{L_n,V_n} \colon A^{<|L_n|} \times \Omega_{L_n} \to V_n$ associated with (L_n, V_n). The main point in Definition 9.12 is that for compatible pairs we can iterate these operations. This is the content of the following definition.

DEFINITION 9.13. *Let A be a finite alphabet with $|A| \geqslant 2$. Also let $d \in \mathbb{N}$ and $(\mathbf{L},\mathbf{V}) = \big((L_n)_{n=0}^d, (V_n)_{n=0}^d\big)$ an A-compatible pair. We set*

$$\mathbf{\Omega_L} = \prod_{n=0}^{d} \Omega_{L_n}. \tag{9.29}$$

By recursion on d, we define the iterated convolution operation

$$\mathrm{c}_{\mathbf{L},\mathbf{V}} \colon A^{<|L_d|} \times \mathbf{\Omega_L} \to V_0 \tag{9.30}$$

associated with $(\mathbf{L}, \mathbf{V})$ *as follows.*

If $d = 0$, then this is the c_{L_0,V_0} convolution operation defined in (9.4). Next let $d \geqslant 1$, set $\mathbf{L}' = (L_n)_{n=0}^{d-1}$ and $\mathbf{V}' = (V_n)_{n=0}^{d-1}$ and assume that the operation $\mathrm{c}_{\mathbf{L}',\mathbf{V}'}$ has been defined. Then we set

$$\mathrm{c}_{\mathbf{L},\mathbf{V}}(t, \omega_0, \dots, \omega_d) = \mathrm{c}_{\mathbf{L}',\mathbf{V}'}\big(\mathrm{c}_{L_d,V_d}(t, \omega_d), \omega_0, \dots, \omega_{d-1}\big) \tag{9.31}$$

for every $t \in A^{<|L_d|}$ *and every* $(\omega_0, \dots, \omega_d) \in \mathbf{\Omega_L}$*. Moreover, we define the* quotient map

$$\mathrm{q}_{\mathbf{L},\mathbf{V}} \colon A^{<|L_d|} \times \mathbf{\Omega_L} \to A^{<|L_{d-1}|} \times \mathbf{\Omega_{L'}} \tag{9.32}$$

associated with $(\mathbf{L}, \mathbf{V})$ *by the rule*

$$\mathrm{q}_{\mathbf{L},\mathbf{V}}(t, \boldsymbol{\omega}, \omega) = \big(\mathrm{c}_{L_d, V_d}(t, \omega), \boldsymbol{\omega}\big) \tag{9.33}$$

for every $t \in A^{<|L_d|}$ *and every* $(\boldsymbol{\omega}, \omega) \in \mathbf{\Omega_{L'}} \times \Omega_{L_d}$.

In the rest of this section we will present several properties of iterated convolutions. Most of these properties are based on the results in Section 9.1. We begin with the following elementary fact.

Fact 9.14. *Let A be a finite alphabet with $|A| \geqslant 2$. Also let d be a positive integer and let* $(\mathbf{L}, \mathbf{V}) = \big((L_n)_{n=0}^d, (V_n)_{n=0}^d\big)$ *be an A-compatible pair. Then, setting* $(\mathbf{L'}, \mathbf{V'}) = \big((L_n)_{n=0}^{d-1}, (V_n)_{n=0}^{d-1}\big)$*, we have* $\mathrm{c}_{\mathbf{L},\mathbf{V}} = \mathrm{c}_{\mathbf{L'},\mathbf{V'}} \circ \mathrm{q}_{\mathbf{L},\mathbf{V}}$.

The following lemma is the multidimensional analogue of Lemma 9.8.

Lemma 9.15. *Let A be a finite alphabet with $|A| \geqslant 2$. Consider an A-compatible pair* $(\mathbf{L}, \mathbf{V}) = \big((L_n)_{n=0}^d, (V_n)_{n=0}^d\big)$ *and let U be a Carlson–Simpson subspace of* $A^{<|L_d|}$ *and* $\boldsymbol{\omega} \in \mathbf{\Omega_L}$*. Then the set*

$$U_{\boldsymbol{\omega}} = \{\mathrm{c}_{\mathbf{L},\mathbf{V}}(u, \boldsymbol{\omega}) : u \in U\} \tag{9.34}$$

is a Carlson–Simpson subspace of V_0 with the same dimension as U. Moreover, for every $i \in \{0, \dots, \dim(U)\}$ *we have*

$$U_{\boldsymbol{\omega}}(i) = \{\mathrm{c}_{\mathbf{L},\mathbf{V}}(u, \boldsymbol{\omega}) : u \in U(i)\}. \tag{9.35}$$

Proof. By induction on d. The case "$d = 0$" is the content of Lemma 9.8. Let $d \geqslant 1$ and assume that the result has been proved up to $d-1$. Fix a compatible pair $(\mathbf{L}, \mathbf{V}) = \big((L_n)_{n=0}^d, (V_n)_{n=0}^d\big)$ and let U and $\boldsymbol{\omega}$ be as in the statement of the lemma. Write $\boldsymbol{\omega} = (\omega_0, \dots, \omega_d)$ and set $\boldsymbol{\omega'} = (\omega_0, \dots, \omega_{d-1})$. Also let $\mathbf{L'} = (L_n)_{n=0}^{d-1}$ and $\mathbf{V'} = (V_n)_{n=0}^{d-1}$ and notice that the pair $(\mathbf{L'}, \mathbf{V'})$ is compatible. Therefore, setting $U_{\omega_d} = \{\mathrm{c}_{L_d,V_d}(u, \omega_d) : u \in U\}$, we see that

$$\begin{aligned} U_{\boldsymbol{\omega}} = \{\mathrm{c}_{\mathbf{L},\mathbf{V}}(u, \boldsymbol{\omega}) : u \in U\} &\overset{(9.31)}{=} \big\{\mathrm{c}_{\mathbf{L'},\mathbf{V'}}\big(\mathrm{c}_{L_d,V_d}(u, \omega_d), \boldsymbol{\omega'}\big) : u \in U\big\} \\ &= \{\mathrm{c}_{\mathbf{L'},\mathbf{V'}}(s, \boldsymbol{\omega'}) : s \in U_{\omega_d}\}. \end{aligned} \tag{9.36}$$

By Lemma 9.8, we have that U_{ω_d} is a Carlson–Simpson subspace of V_d with $\dim(U_{\omega_d}) = \dim(U)$. This implies, in particular, that $U_{\omega_d} \subseteq V_d \subseteq A^{<|L_{d-1}|}$. Therefore, by (9.36) and our inductive assumptions applied for the compatible pair $(\mathbf{L'}, \mathbf{V'})$, the Carlson–Simpson space U_{ω_d} and the element $\boldsymbol{\omega'} \in \mathbf{\Omega_{L'}}$, we conclude that $U_{\boldsymbol{\omega}}$ is a Carlson–Simpson subspace of V_0 of dimension $\dim(U)$. The equality in (9.35) is verified similarly. The proof of Lemma 9.15 is completed. □

The next result is an extension of Lemma 9.9.

LEMMA 9.16. *Let A be a finite alphabet with $|A|\geqslant 2$. Consider an A-compatible pair $(\mathbf{L},\mathbf{V})=\big((L_n)_{n=0}^d,(V_n)_{n=0}^d\big)$ and let U be a Carlson–Simpson subspace of $A^{<|L_d|}$ and $\boldsymbol{\omega}\in\boldsymbol{\Omega}_{\mathbf{L}}$. Also let D be a subset of $A^{<\mathbb{N}}$ and set $\mathscr{D}=\mathrm{c}_{\mathbf{L},\mathbf{V}}^{-1}(D)$. Then for every $i\in\{0,\dots,\dim(W)\}$ we have*

$$\mathrm{dens}_{U_{\boldsymbol{\omega}}(i)}(D)=\mathrm{dens}_{U(i)\times\{\boldsymbol{\omega}\}}(\mathscr{D}). \tag{9.37}$$

PROOF. Let $h\colon A^{<|L_d|}\to V_0$ be defined by the rule $h(t)=\mathrm{c}_{\mathbf{L},\mathbf{V}}(t,\boldsymbol{\omega})$. By induction on d, we see that the map h is an injection. Also let $i\in\{0,\dots,\dim(V)\}$ and notice that, by (9.35), we have $U_{\boldsymbol{\omega}}(i)=h\big(U(i)\big)$. Therefore,

$$|U_{\boldsymbol{\omega}}(i)|=|h\big(U(i)\big)|=|U(i)|=|U(i)\times\{\boldsymbol{\omega}\}|. \tag{9.38}$$

Next observe that

$$h^{-1}\big(U_{\boldsymbol{\omega}}(i)\cap D\big)\times\{\boldsymbol{\omega}\}=\big(U(i)\times\{\boldsymbol{\omega}\}\big)\cap \mathrm{c}_{\mathbf{L},\mathbf{V}}^{-1}(D)$$

and so

$$|U_{\boldsymbol{\omega}}(i)\cap D|=|h^{-1}\big(U_{\boldsymbol{\omega}}(i)\cap D\big)\times\{\boldsymbol{\omega}\}|=|\big(U(i)\times\{\boldsymbol{\omega}\}\big)\cap\mathscr{D}|. \tag{9.39}$$

By (9.38) and (9.39), we conclude that

$$\mathrm{dens}_{U_{\boldsymbol{\omega}}(i)}(D)=\frac{|U_{\boldsymbol{\omega}}(i)\cap D|}{|U_{\boldsymbol{\omega}}(i)|}=\frac{|\big(U(i)\times\{\boldsymbol{\omega}\}\big)\cap\mathscr{D}|}{|U(i)\times\{\boldsymbol{\omega}\}|}=\mathrm{dens}_{U(i)\times\{\boldsymbol{\omega}\}}(\mathscr{D})$$

and the proof of Lemma 9.16 is completed. □

We proceed with the following lemma.

LEMMA 9.17. *Let A be a finite alphabet with $|A|\geqslant 2$. Consider an A-compatible pair $(\mathbf{L},\mathbf{V})=\big((L_n)_{n=0}^d,(V_n)_{n=0}^d\big)$ and for every $t\in A^{<|L_d|}$ set*

$$\mathcal{C}_t=\{\mathrm{c}_{L_d,V_d}(t,\omega):\omega\in\Omega_{L_d}\}. \tag{9.40}$$

Assume that $d\geqslant 1$ and set $(\mathbf{L}',\mathbf{V}')=\big((L_n)_{n=0}^{d-1},(V_n)_{n=0}^{d-1}\big)$. Then the following hold.

(a) *For every $t\in A^{<|L_d|}$ we have $\mathrm{q}_{\mathbf{L},\mathbf{V}}^{-1}(\mathcal{C}_t\times\boldsymbol{\Omega}_{\mathbf{L}'})=\{t\}\times\boldsymbol{\Omega}_{\mathbf{L}}$.*

(b) *For every $t\in A^{<|L_d|}$ and every $\mathcal{D}\subseteq A^{<|L_{d-1}|}\times\boldsymbol{\Omega}_{\mathbf{L}'}$ we have*

$$\mathrm{dens}_{\mathcal{C}_t\times\boldsymbol{\Omega}_{\mathbf{L}'}}(\mathcal{D})=\mathrm{dens}_{\{t\}\times\boldsymbol{\Omega}_{\mathbf{L}}}\big(\mathrm{q}_{\mathbf{L},\mathbf{V}}^{-1}(\mathcal{D})\big). \tag{9.41}$$

PROOF. Let $t\in A^{<|L_d|}$ be arbitrary. By the definition of the quotient map $\mathrm{q}_{\mathbf{L},\mathbf{V}}$ in (9.33), we have

$$\mathrm{q}_{\mathbf{L},\mathbf{V}}^{-1}(\mathcal{C}_t\times\boldsymbol{\Omega}_{\mathbf{L}'})=\mathrm{c}_{L_d,V_d}^{-1}(\mathcal{C}_t)\times\boldsymbol{\Omega}_{\mathbf{L}'}.$$

On the other hand, by Fact 9.4, we see that $\mathrm{c}_{L_d,V_d}^{-1}(\mathcal{C}_t)=\{t\}\times\Omega_{L_d}$. Therefore,

$$\mathrm{q}_{\mathbf{L},\mathbf{V}}^{-1}(\mathcal{C}_t\times\boldsymbol{\Omega}_{\mathbf{L}'})=\big(\{t\}\times\Omega_{L_d}\big)\times\boldsymbol{\Omega}_{\mathbf{L}'}=\{t\}\times\boldsymbol{\Omega}_{\mathbf{L}}$$

and the proof of the first part of the lemma is completed.

We proceed to the proof of part (b). We fix a subset $\mathcal{D}$ of $A^{<|L_{d-1}|}\times\boldsymbol{\Omega}_{\mathbf{L}'}$ and we set $\mathscr{D}=\mathrm{q}_{\mathbf{L},\mathbf{V}}^{-1}(\mathcal{D})$. For every $\boldsymbol{\omega}'\in\boldsymbol{\Omega}_{\mathbf{L}'}$ let $\mathcal{D}_{\boldsymbol{\omega}'}=\{t\in A^{<|L_d|}:(t,\boldsymbol{\omega}')\in\mathcal{D}\}$ and $\mathscr{D}_{\boldsymbol{\omega}'}=\{(t,\omega)\in A^{<|L_d|}\times\Omega_{L_d}:(t,\boldsymbol{\omega}',\omega)\in\mathscr{D}\}$ be the sections at $\boldsymbol{\omega}'$ of $\mathcal{D}$ and $\mathscr{D}$ respectively. Observe that $\mathscr{D}_{\boldsymbol{\omega}'}=\mathrm{c}_{L_d,V_d}^{-1}(\mathcal{D}_{\boldsymbol{\omega}'})$. Hence, by Lemma 9.7, we have

$$\mathrm{dens}_{\{t\}\times\Omega_{L_d}}(\mathscr{D}_{\boldsymbol{\omega}'})=\mathrm{dens}_{\mathcal{C}_t}(\mathcal{D}_{\boldsymbol{\omega}'}). \tag{9.42}$$

Taking the average over all $\boldsymbol{\omega}' \in \boldsymbol{\Omega}_{\mathbf{L}'}$ we conclude that

$$\begin{aligned}\mathrm{dens}_{\{t\}\times\boldsymbol{\Omega}_{\mathbf{L}}}\big(\mathrm{q}^{-1}_{\mathbf{L},\mathbf{V}}(\mathcal{D})\big) &= \mathrm{dens}_{\{t\}\times\boldsymbol{\Omega}_{\mathbf{L}}}(\mathscr{D}) = \mathbb{E}_{\boldsymbol{\omega}'\in\boldsymbol{\Omega}_{\mathbf{L}'}}\mathrm{dens}_{\{t\}\times\Omega_{L_d}}(\mathscr{D}_{\boldsymbol{\omega}'})\\ &\overset{(9.42)}{=} \mathbb{E}_{\boldsymbol{\omega}'\in\boldsymbol{\Omega}_{\mathbf{L}'}}\mathrm{dens}_{\mathcal{C}_t}(\mathcal{D}_{\boldsymbol{\omega}'}) = \mathrm{dens}_{\mathcal{C}_t\times\boldsymbol{\Omega}_{\mathbf{L}'}}(\mathcal{D}).\end{aligned}$$

The proof of Lemma 9.17 is completed. □

We close this section with the following consequence of Lemma 9.17.

Corollary 9.18. *Let A be a finite alphabet with $|A| \geqslant 2$ and let d be a positive integer. Also let $(\mathbf{L},\mathbf{V}) = \big((L_n)_{n=0}^{d},(V_n)_{n=0}^{d}\big)$ be an A-compatible pair and set $(\mathbf{L}',\mathbf{V}') = \big((L_n)_{n=0}^{d-1},(V_n)_{n=0}^{d-1}\big)$. Finally, let $D \subseteq A^{<\mathbb{N}}$ and set $\mathscr{D} = \mathrm{c}^{-1}_{\mathbf{L},\mathbf{V}}(D)$ and $\mathcal{D} = \mathrm{c}^{-1}_{\mathbf{L}',\mathbf{V}'}(D)$. Then for every $t \in A^{<|L_d|}$ we have*

$$\mathrm{dens}_{\boldsymbol{\Omega}_{\mathbf{L}}}(\mathscr{D}_t) = \mathbb{E}_{s\in\mathcal{C}_t}\mathrm{dens}_{\boldsymbol{\Omega}_{\mathbf{L}'}}(\mathcal{D}_s) \tag{9.43}$$

where $\mathscr{D}_t$ is the section of $\mathscr{D}$ at t, $\mathcal{C}_t \subseteq V_d \subseteq A^{<|L_{d-1}|}$ is as in (9.40) and $\mathcal{D}_s$ is the section of $\mathcal{D}$ at s.

Proof. By Fact 9.14, we see that

$$\mathscr{D} = \mathrm{c}^{-1}_{\mathbf{L},\mathbf{V}}(D) = \mathrm{q}^{-1}_{\mathbf{L},\mathbf{V}}\big(\mathrm{c}^{-1}_{\mathbf{L}',\mathbf{V}'}(D)\big) = \mathrm{q}^{-1}_{\mathbf{L},\mathbf{V}}(\mathcal{D}). \tag{9.44}$$

Now let $t \in A^{<|L_d|}$ be arbitrary and observe that

$$\begin{aligned}\mathrm{dens}_{\boldsymbol{\Omega}_{\mathbf{L}}}(\mathscr{D}_t) &= \mathrm{dens}_{\{t\}\times\boldsymbol{\Omega}_{\mathbf{L}}}(\mathscr{D}) \overset{(9.44)}{=} \mathrm{dens}_{\{t\}\times\boldsymbol{\Omega}_{\mathbf{L}}}\big(\mathrm{q}^{-1}_{\mathbf{L},\mathbf{V}}(\mathcal{D})\big)\\ &\overset{(9.41)}{=} \mathrm{dens}_{\mathcal{C}_t\times\boldsymbol{\Omega}_{\mathbf{L}'}}(\mathcal{D}) = \mathbb{E}_{s\in\mathcal{C}_t}\mathrm{dens}_{\boldsymbol{\Omega}_{\mathbf{L}'}}(\mathcal{D}_s).\end{aligned}$$

The proof of Corollary 9.18 is completed. □

9.3. Some basic estimates

In this section we will present three results which are needed for the proof of Theorem 9.2 and are independent of the rest of the argument. The first of these results is a measure-theoretic consequence of Theorem 9.2. It is the analogue of Proposition 8.7 in the context of the density Carlson–Simpson theorem and will be presented in Subsection 9.3.1 together with some related material of probabilistic nature. The next two results are part of a general inductive scheme that we will discuss in Subsection 9.4.1. Specifically, in Subsection 9.3.2 we prove the first instance of Theorem 9.2 which can be seen as a variant of Sperner's theorem. Finally, in Subsection 9.3.3 we estimate the numbers $\mathrm{DCS}(k,m+1,\delta)$ assuming that the numbers $\mathrm{DCS}(k,m,\beta)$ have been estimated for every $0<\beta\leqslant 1$.

9.3.1. Probabilistic tools.

We start by introducing the following classes of probability measures.

Definition 9.19. *Let A be a finite alphabet with $|A| \geqslant 2$, V a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ and L a nonempty finite subset of $\mathbb{N}$.*

(a) *The* Furstenberg–Weiss measure $\mathrm{d}^V_{\mathrm{FW}}$ associated with V *is the probability measure on $A^{<\mathbb{N}}$ defined by the rule*

$$\mathrm{d}^V_{\mathrm{FW}}(D) = \mathbb{E}_{n\in\{0,\dots,\dim(V)\}}\,\mathrm{dens}_{V(n)}(D). \tag{9.45}$$

(b) *The* generalized Furstenberg–Weiss measure d_L associated with L *is the probability measure on* $A^{<\mathbb{N}}$ *defined by*

$$\mathrm{d}_L(D) = \mathbb{E}_{n\in L}\,\mathrm{dens}_{A^n}(D). \tag{9.46}$$

We point out that the class of generalized Furstenberg–Weiss measures is closed under averages. Specifically, if L is a nonempty finite subset of $\mathbb{N}$, then for every $n \in \{1,\dots,|L|\}$ and every set D of words over A we have

$$\mathrm{d}_L(D) = \mathbb{E}_{M\in\binom{L}{n}}\,\mathrm{d}_M(D). \tag{9.47}$$

However, the Furstenberg–Weiss measures associated with Carlson–Simpson spaces do not have this important property. Also observe that if ℓ is an integer with $\ell \geqslant 2$ and $L = \{0,\dots,\ell-1\}$ is the initial interval of $\mathbb{N}$ of size ℓ, then the probability measure d_L coincides with the Furstenberg–Weiss measure associated with the Carlson–Simpson space $A^{<\ell}$. More generally, we have the following lemma which relates these two classes of measures.

Lemma 9.20. *Let A be a finite alphabet with $|A| \geqslant 2$ and L a finite subset of $\mathbb{N}$ with $|L| \geqslant 2$. Then for every subset D of $A^{<\mathbb{N}}$ there exists a Carlson–Simpson space V of $A^{<\mathbb{N}}$ of dimension $|L|-1$ such that $\mathrm{d}^V_{\mathrm{FW}}(D) \geqslant \mathrm{d}_L(D)$ and with $L(V) = L$ where, as in* (1.33)*, $L(V)$ is the level set V.*

Proof. We fix a subset D of $A^{<\mathbb{N}}$. Let $l_0 < \dots < l_{|L|-1}$ be the increasing enumeration of the set L and let $\mathrm{c}_L \colon A^{<|L|} \times \Omega_L \to A^{<\mathbb{N}}$ be the convolution operation associated with L. Moreover, for every $\omega \in \Omega_L$ set

$$R_\omega = \{\mathrm{c}_L(t,\omega) : t \in A^{<|L|}\}$$

and recall that, by Lemma 9.8, the set R_ω is a Carlson–Simpson space of $A^{<\mathbb{N}}$ of dimension $|L|-1$. Also notice that $L(R_\omega) = L$. On the other hand, by Lemma 9.10, for every $i \in \{0,\dots,|L|-1\}$ we have

$$\mathrm{dens}_{A^{l_i}}(D) = \mathbb{E}_{\omega\in\Omega_L}\mathrm{dens}_{R_\omega(i)}(D)$$

and so, by averaging over all $i \in \{0,\dots,|L|-1\}$, we obtain that

$$\mathrm{d}_L(D) = \mathbb{E}_{\omega\in\Omega_L}\,\mathrm{d}^{R_\omega}_{\mathrm{FW}}(D).$$

Therefore, there exists $\omega_0 \in \Omega_L$ with $\mathrm{d}^{R_{\omega_0}}_{\mathrm{FW}}(D) \geqslant \mathrm{d}_L(D)$ and the proof of Lemma 9.20 is completed. □

Now let k and m be positive integers with $k \geqslant 2$ and assume that for every $0 < \beta \leqslant 1$ the number $\mathrm{DCS}(k,m,\beta)$ has been estimated. This assumption permits us to introduce some numerical invariants. Specifically, for every $0 < \eta \leqslant 1$ we set

$$\Lambda(k,m,\eta) = \lceil \eta^{-1}\mathrm{DCS}(k,m,\eta)\rceil \tag{9.48}$$

and

$$\Theta(k,m,\eta) = \frac{2\eta}{\big|\mathrm{SubCS}_m\big([k]^{<\Lambda(k,m,\eta)}\big)\big|}. \tag{9.49}$$

The following proposition is the main result of this subsection.

PROPOSITION 9.21. *Let k, m be positive integers with $k \geqslant 2$ and assume that for every $0 < \beta \leqslant 1$ the number* $\mathrm{DCS}(k, m, \beta)$ *has been estimated.*

Let A be an alphabet with $|A| = k$ and $0 < \delta, \varepsilon \leqslant 1$. Also let L be a finite subset of $\mathbb{N}$ with $|L| \geqslant \Lambda(k, m, \delta\varepsilon/4)$ where $\Lambda(k, m, \delta\varepsilon/4)$ is as in (9.48)*. Finally, let E be a subset of $A^{<\mathbb{N}}$ such that $\mathrm{d}_L(E) \geqslant \varepsilon$. If $\{D_t : t \in E\}$ is a family of measurable events in a probability space (Ω, Σ, μ) satisfying $\mu(D_t) \geqslant \delta$ for every $t \in E$, then there exists an m-dimensional Carlson–Simpson space S of $A^{<\mathbb{N}}$ which is contained in E and such that*

$$\mu\Big(\bigcap_{t \in S} D_t\Big) \geqslant \Theta(k, m, \delta\varepsilon/4)$$

where $\Theta(k, m, \delta\varepsilon/4)$ is as in (9.49)*.*

For the proof of Proposition 9.21 we need the following simple fact.

FACT 9.22. *Let k, m be positive integers with $k \geqslant 2$, $0 < \eta \leqslant 1/2$ and assume that the number* $\mathrm{DCS}(k, m, \eta)$ *has been estimated. Let A be an alphabet with $|A| = k$ and V a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ with* $\dim(V) \geqslant \Lambda(k, m, \eta) - 1$*. Then every $D \subseteq A^{<\mathbb{N}}$ with $\mathrm{d}^V_{\mathrm{FW}}(D) \geqslant 2\eta$ contains an m-dimensional Carlson–Simpson subspace of V.*

PROOF. Let $L = \{n \in \{0, \dots, \dim(V)\} : \mathrm{dens}_{V(n)}(D) \geqslant \eta\}$. By identifying V with $A^{<\dim(V)+1}$ via the canonical isomorphism I_V (see Definition 1.10), it is enough to show that $|L| \geqslant \mathrm{DCS}(k, m, \eta)$. Indeed, by Markov's inequality and the fact that $\mathrm{d}^V_{\mathrm{FW}}(D) \geqslant 2\eta$, we obtain that

$$|L| \geqslant \eta(\dim(V) + 1) \geqslant \eta\Lambda(k, m, \eta) \overset{(9.48)}{\geqslant} \mathrm{DCS}(k, m, \eta)$$

and the proof of Fact 9.22 is completed. □

We proceed to the proof of Proposition 9.21.

PROOF OF PROPOSITION 9.21. We set $\Lambda = \Lambda(k, m, \delta\varepsilon/4)$. By (9.47) and by passing to an appropriate subset of L if necessary, we may assume that $|L| = \Lambda$. By Lemma 9.20, there exists a Carlson–Simpson space V of $A^{<\mathbb{N}}$ with $\dim(V) = \Lambda - 1$, $L(V) = L$ and such that $\mathrm{d}^V_{\mathrm{FW}}(E) \geqslant \mathrm{d}_L(E) \geqslant \varepsilon$. For every $\omega \in \Omega$ let

$$E_\omega = \{t \in E \cap V : \omega \in D_t\}$$

and set

$$Y = \{\omega \in \Omega : \mathrm{d}^V_{\mathrm{FW}}(E_\omega) \geqslant \delta\varepsilon/2\}.$$

Since $\mathrm{d}^V_{\mathrm{FW}}(E) \geqslant \varepsilon$ and $\mu(D_t) \geqslant \delta$ for every $t \in E$, we see that $\mu(Y) \geqslant \delta\varepsilon/2$. On the other hand, by the choice of Λ and Fact 9.22, for every $\omega \in Y$ there exists an m-dimensional Carlson–Simpson subspace S_ω of V with $S_\omega \subseteq E_\omega$. Noticing that $|\mathrm{SubCS}_m(V)| = |\mathrm{SubCS}_m\big([k]^{<\Lambda}\big)|$, we conclude that there exist $S \in \mathrm{SubCS}_m(V)$ and $G \in \Sigma$ with $S_\omega = S$ for every $\omega \in G$ and such that

$$\mu(G) \geqslant \frac{\mu(Y)}{|\mathrm{SubCS}_m(V)|} \geqslant \frac{\delta\varepsilon/2}{|\mathrm{SubCS}_m\big([k]^{<\Lambda}\big)|} \overset{(9.49)}{=} \Theta(k, m, \delta\varepsilon/4).$$

The proof of Proposition 9.21 is completed. □

We isolate, for future use, the following consequence of Proposition 9.21.

COROLLARY 9.23. *Let k, m be positive integers with $k \geqslant 2$ and assume that for every $0 < \beta \leqslant 1$ the number* $\mathrm{DCS}(k, m, \beta)$ *has been estimated.*

Let $0 < \delta \leqslant 1$ and $d \in \mathbb{N}$ with $d \geqslant \Lambda(k, m, \delta/4) - 1$ where $\Lambda(k, m, \delta/4)$ is as in (9.48). *Also let A be an alphabet with $|A| = k + 1$ and V a Carlson–Simpson space of $A^{<\mathbb{N}}$ with $\dim(V) \geqslant \mathrm{CS}(k + 1, d, m, 2)$. If B is a subset of A with $|B| = k$ and $\{D_t : t \in V\}$ is a family of measurable events in a probability space (Ω, Σ, μ) satisfying $\mu(D_t) \geqslant \delta$ for every $t \in V$, then there exists $W \in \mathrm{SubCS}_d(V)$ such that for every $U \in \mathrm{SubCS}_m(W)$ we have*

$$\mu\Big(\bigcap_{t \in U \upharpoonright B} D_t\Big) \geqslant \Theta(k, m, \delta/4)$$

where $\Theta(k, m, \delta/4)$ is as in (9.49) *and $U \upharpoonright B$ is as in* (1.40).

PROOF. We set

$$\mathcal{U} = \Big\{U \in \mathrm{SubCS}_m(V) : \mu\Big(\bigcap_{t \in U \upharpoonright B} D_t\Big) \geqslant \Theta(k, m, \delta/4)\Big\}.$$

By Theorem 4.21, there exists $W \in \mathrm{SubCS}_d(V)$ such that either $\mathrm{SubCS}_m(W) \subseteq \mathcal{U}$ or $\mathrm{SubCS}_m(W) \cap \mathcal{U} = \emptyset$. Therefore, it is enough to show that $\mathrm{SubCS}_m(W) \cap \mathcal{U} \neq \emptyset$. To this end we argue as follows. Let $\mathrm{I}_W : A^{<d+1} \to W$ be the canonical isomorphism associated with W and for every $t \in B^{<d+1}$ set $D'_t = D_{\mathrm{I}_W(t)}$. By Proposition 9.21, there exists $S \in \mathrm{SubCS}_m(B^{<d+1})$ such that

$$\mu\Big(\bigcap_{t \in S} D'_t\Big) \geqslant \Theta(k, m, \delta/4). \tag{9.50}$$

If U is the unique element of $\mathrm{SubCS}_m(W)$ such that $U \upharpoonright B = \mathrm{I}_W(S)$, then, by (9.50), we see that $U \in \mathcal{U}$. The proof of Corollary 9.23 is completed. □

We will also need the following variant of Corollary 9.23.

COROLLARY 9.24. *Let k, m be positive integers with $k \geqslant 2$ and assume that for every $0 < \beta \leqslant 1$ the number* $\mathrm{DCS}(k, m, \beta)$ *has been estimated.*

Let $0 < \delta \leqslant 1$ and $d \in \mathbb{N}$ with $d \geqslant \Lambda(k, m, \delta/4) - 1$ where $\Lambda(k, m, \delta/4)$ is as in (9.48). *Also let A be an alphabet with $|A| = k + 1$ and V a Carlson–Simpson space of $A^{<\mathbb{N}}$ with $\dim(V) \geqslant \mathrm{CS}(k + 1, d, m, 2)$. Finally, let $\{D_t : t \in V\}$ be a family of measurable events in a probability space (Ω, Σ, μ) such that:* (i) *$\mu(D_t) \geqslant \delta$ for every $t \in V$, and* (ii) *there exist $a, b \in A$ with $a \neq b$ such that $D_t = D_{t'}$ for every $t, t' \in V$ which are (a, b)-equivalent (see Subsection* 2.1.1*). Then there exists $W \in \mathrm{SubCS}_d(V)$ such that for every $U \in \mathrm{SubCS}_m(W)$ we have*

$$\mu\Big(\bigcap_{t \in U} D_t\Big) \geqslant \Theta(k, m, \delta/4)$$

where $\Theta(k, m, \delta/4)$ is as in (9.49).

Proof. Set $B = A\setminus\{a\}$. Observe that our assumption that $D_t = D_{t'}$ for every t, t' which are (a,b)-equivalent, implies that for every Carlson–Simpson subspace R of V we have

$$\bigcap_{t\in R} D_t = \bigcap_{t\in R\upharpoonright B} D_t.$$

Using this observation, the result follows from Corollary 9.23. □

9.3.2. Estimation of the numbers $\mathrm{DCS}(2,1,\delta)$**.** This subsection is devoted to the proof of the following proposition which deals with the first non-trivial case of Theorem 9.2.

Proposition 9.25. *Let A be an alphabet with $|A| = 2$ and $0 < \delta \leqslant 1$. Also let D be a subset of $A^{<\mathbb{N}}$ and L_0 a finite subset of $\mathbb{N}$ such that*

$$|L_0| \geqslant \mathrm{RegCS}\big(2, 1, \mathrm{CS}(2, \lceil 17\delta^{-2}\rceil, 1, 2) + 1, \delta/4\big). \tag{9.51}$$

If $|D \cap A^n| \geqslant \delta 2^n$ for every $n \in L_0$, then there exists a Carlson–Simpson line R of $A^{<\mathbb{N}}$ which is contained in D. In particular,

$$\mathrm{DCS}(2,1,\delta) \leqslant \mathrm{RegCS}\big(2, 1, \mathrm{CS}(2, \lceil 17\delta^{-2}\rceil, 1, 2) + 1, \delta/4\big). \tag{9.52}$$

We point out that the estimate for the numbers $\mathrm{DCS}(2,1,\delta)$ obtained in (9.52) is far from being optimal. However, the proof of Proposition 9.25 is conceptually close to the proof of the general case of Theorem 9.2 and can serve as a motivating introduction to the main argument.

Proof of Proposition 9.25. Write the alphabet A as $\{a, b\}$, and let D and L_0 be as in the statement of the proposition. We start with the following claim.

Claim 9.26. *There exists a subset L of L_0 with*

$$|L| = \mathrm{CS}(2, \lceil 17\delta^{-2}\rceil, 1, 2) + 1 \tag{9.53}$$

and satisfying the following property. Let $\mathrm{c}_L \colon A^{<|L|} \times \Omega_L \to A^{<\mathbb{N}}$ be the convolution operation associated with L and set $\mathscr{D} = \mathrm{c}_L^{-1}(D)$. Then for every $t \in A^{<|L|}$ we have $\mathrm{dens}(\mathscr{D}_t) \geqslant 3\delta/4$ where $\mathscr{D}_t = \{\omega \in \Omega_L : (t,\omega) \in \mathscr{D}\}$ is the section of $\mathscr{D}$ at t.

Proof of Claim 9.26. By (9.51) and Lemma 6.24, there exists $L \subseteq L_0$ with $|L| = \mathrm{CS}(2, \lceil 17\delta^{-2}\rceil, 1, 2)+1$ such that the family $\mathcal{F} \coloneqq \{D\}$ is $(\delta/4, L)$-regular. By Lemma 9.11, the set L is as desired. The proof of Claim 9.26 is completed. □

We introduce some terminology. Let W be a Carlson–Simpson space of $A^{<\mathbb{N}}$ of finite dimension. Set $m = \dim(W)$ and let $\langle c, (w_n)_{n=0}^{m-1}\rangle$ be the Carlson–Simpson system over A which generates W. Also let $t, t' \in W$. We say that t' is a *successor* of t in W provided that there exist $i, j \in \{0, \dots, m-1\}$ with $i \leqslant j$ and $a_i, \dots, a_j \in A$ such that $t' = t^\frown w_i(a_i)^\frown \dots ^\frown w_j(a_j)$. If, in addition, we have $a_i = a$, then we say that t' is an *a-successor* of t in W.

Claim 9.27. *Let L and $\{\mathscr{D}_t : t \in A^{<|L|}\}$ be as in Claim 9.26. Then there exists a Carlson–Simpson subspace W of $A^{<|L|}$ with $\dim(W) = \lceil 17\delta^{-2}\rceil$ such that $\mathrm{dens}(\mathscr{D}_t \cap \mathscr{D}_{t'}) \geqslant \delta^2/16$ for every $t, t' \in W$ with t' an a-successor of t in W.*

Proof of Claim 9.27. We define a subset $\mathcal{L}$ of $\mathrm{SubCS}_1\big(A^{<|L|}\big)$ by the rule

$$S \in \mathcal{L} \quad\Leftrightarrow\quad \text{if } \langle s, s_0\rangle \text{ is the Carlson–Simpson system generating } S, \text{ then } \mathrm{dens}(\mathscr{D}_s \cap \mathscr{D}_{s^\frown s_0(a)}) \geqslant \delta^2/16.$$

By (9.53) and Theorem 4.21, there exists a Carlson–Simpson subspace W of $A^{<|L|}$ with $\dim(W) = \lceil 17\delta^{-2}\rceil$ such that either $\mathrm{SubCS}_1(W) \subseteq \mathcal{L}$ or $\mathrm{SubCS}_1(W) \cap \mathcal{L} = \emptyset$. Let $t, t' \in W$ and observe that t' is an a-successor of t in W if and only if there exists a Carlson–Simpson line S of W such that, denoting by $\langle s, s_0\rangle$ the system generating S, we have $t = s$ and $t' = s^\frown s_0(a)$. Therefore, the proof will be completed once we show that $\mathrm{SubCS}_1(W) \cap \mathcal{L} \neq \emptyset$. To this end, set $d = \dim(W) = \lceil 17\delta^{-2}\rceil$ and let $\langle w, (w_n)_{n=0}^{d-1}\rangle$ be the Carlson–Simpson system which generates W. We set $t_0 = w$ and $t_i = w^\frown w_0(a)^\frown \dots ^\frown w_{i-1}(a)$ for every $i \in [d]$. By our assumptions, we have $\mathrm{dens}(\mathscr{D}_{t_i}) \geqslant 3\delta/4$ for every $i \in \{0, \dots, d\}$. Hence, by Lemma E.5 applied for "$\varepsilon = 3\delta/4$" and "$\theta = \delta/4$", there exist $i, j \in \{0, \dots, d\}$ with $i < j$ such that $\mathrm{dens}(\mathscr{D}_{t_i} \cap \mathscr{D}_{t_j}) \geqslant \delta^2/16$. If R is the Carlson–Simpson line of W generated by the system $\langle t_i, w_i{}^\frown \dots ^\frown w_{j-1}\rangle$, then the previous discussion implies that $R \in \mathcal{L}$. The proof of Claim 9.27 is completed. □

The following claim is the last step of the proof of Proposition 9.25.

Claim 9.28. *Let W be the Carlson–Simpson space obtained by Claim 9.27. Then W contains a Carlson–Simpson line S such that $\bigcap_{t\in S} \mathscr{D}_t \neq \emptyset$.*

Proof of Claim 9.28. As in Claim 9.27, we set $d = \dim(W) = \lceil 17\delta^{-2}\rceil$. Also let $\langle w, (w_n)_{n=0}^{d-1}\rangle$ be the Carlson–Simpson system which generates W. For every $i \in \{0, \dots, d-2\}$ set

$$t_i = w^\frown w_0(b)^\frown \dots ^\frown w_i(b) \quad\text{and}\quad s_i = t_i{}^\frown w_{i+1}(a)^\frown \dots ^\frown w_{d-1}(a)$$

and observe that s_i is an a-successor of t_i in W. Therefore, by Claim 9.27, setting $C_i = \mathscr{D}_{t_i} \cap \mathscr{D}_{s_i}$ we have $\mathrm{dens}(C_i) \geqslant \delta^2/16$ for every $i \in \{0, \dots, d-2\}$. Also let $s_{d-1} = w^\frown w_0(b)^\frown \dots ^\frown w_{d-1}(b)$ and $C_{d-1} = \mathscr{D}_{s_{d-1}}$ and notice that, by Claim 9.26, we have $\mathrm{dens}(C_{d-1}) \geqslant 3\delta/4 \geqslant \delta^2/16$. Since $d > 16/\delta^2$, there exist $0 \leqslant i < j \leqslant d-1$ such that $C_i \cap C_j \neq \emptyset$. We define

$$s = t_i \quad\text{and}\quad s_0 = w_{i+1}{}^\frown \dots ^\frown w_j{}^\frown y$$

where $y = w_{j+1}(a)^\frown \dots ^\frown w_{d-1}(a)$ if $j < d-1$ and $y = \emptyset$ otherwise. Let S be the Carlson–Simpson line of W generated by the system $\langle s, s_0\rangle$ and observe that $S = \{t_i\} \cup \{s_i, s_j\}$. Hence,

$$\bigcap_{t\in S} \mathscr{D}_t \supseteq C_i \cap C_j \neq \emptyset$$

and the proof of Claim 9.28 is completed. □

We are now in a position to complete the proof of the proposition. Let S be the Carlson–Simpson line obtained by Claim 9.28. We select $\omega_0 \in \Omega_L$ such that $\omega_0 \in \mathscr{D}_t$ for every $t \in S$, and we set

$$R = \big\{\mathrm{c}_L(t, \omega_0) : t \in S\big\}.$$

By Lemma 9.8, we see that R is a Carlson–Simpson line of $A^{<\mathbb{N}}$. Finally, recall that $\mathscr{D} = \mathrm{c}_L^{-1}(D)$. Since $(t, \omega_0) \in \mathscr{D}$ for every $t \in S$, we conclude that R is contained in D and the proof of Proposition 9.25 is completed. □

9.3.3. Estimation of the numbers $\mathrm{DCS}(k, m+1, \delta)$**.** Let k, m be positive integers with $k \geqslant 2$ and assume that for every $0 < \beta \leqslant 1$ the number $\mathrm{DCS}(k, m, \beta)$ has been estimated. This assumption implies, of course, that for every $\ell \in [m]$ and every $0 < \beta \leqslant 1$ the number $\mathrm{DCS}(k, \ell, \beta)$ has also been estimated. For every $0 < \delta \leqslant 1$ we set

$$\Lambda_0 = \Lambda_0(k, \delta) = \Lambda(k, 1, \delta^2/16) \ \text{ and } \ \Theta_0 = \Theta_0(k, \delta) = \Theta(k, 1, \delta^2/16) \tag{9.54}$$

where $\Lambda(k, 1, \delta^2/16)$ and $\Theta(k, 1, \delta^2/16)$ are as in (9.48) and (9.49) respectively. Also let $h_{k,\delta} \colon \mathbb{N} \to \mathbb{N}$ be defined by the rule

$$h_{k,\delta}(n) = \Lambda_0 + \lceil 2\Theta_0^{-1} n \rceil. \tag{9.55}$$

Notice that for the definition of Λ_0, Θ_0 and $h_{k,\delta}$ we only need to have the number $\mathrm{DCS}(k, 1, \delta^2/16)$ at our disposal.

The following theorem is the main result of this subsection.

Theorem 9.29. *Let k, m be positive integers with $k \geqslant 2$ and assume that for every $0 < \beta \leqslant 1$ the number $\mathrm{DCS}(k, m, \beta)$ has been estimated. Then for every $0 < \delta \leqslant 1$ we have*

$$\mathrm{DCS}(k, m+1, \delta) \leqslant h_{k,\delta}^{(\lceil 8\delta^{-2} \rceil)}\big(\mathrm{DCS}(k, m, \Theta_0/2)\big) \tag{9.56}$$

where Θ_0 and $h_{k,\delta}$ are as in (9.54) and (9.55) respectively.

The proof of Theorem 9.29 is based on the following dichotomy.

Lemma 9.30. *Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every $0 < \beta \leqslant 1$ the number $\mathrm{DCS}(k, 1, \beta)$ has been estimated.*

Let $0 < \delta \leqslant 1$ and let Λ_0 and Θ_0 be as in (9.54). Also let A be an alphabet with $|A| = k$, L a nonempty finite subset of $\mathbb{N}$ and $D \subseteq A^{<\mathbb{N}}$ such that $\mathrm{dens}_{A^\ell}(D) \geqslant \delta$ for every $\ell \in L$. Finally, let n be a positive integer and assume that $|L| \geqslant h_{k,\delta}(n)$ where $h_{k,\delta}$ is as in (9.55). Then, denoting by L_0 the set of the first Λ_0 elements of L, we have that either

(i) *there exist a subset L' of $L \setminus L_0$ with $|L'| \geqslant n$ and a word $t_0 \in A^{\ell_0}$ for some $\ell_0 \in L_0$ such that*

$$\mathrm{dens}_{A^{\ell - \ell_0}}\big(\{s \in A^{<\mathbb{N}} : t_0{}^\frown s \in D\}\big) \geqslant \delta + \delta^2/8 \tag{9.57}$$

for every $\ell \in L'$, or

(ii) *there exist a subset L'' of $L \setminus L_0$ with $|L''| \geqslant n$ and a Carlson–Simpson line S of $A^{<\mathbb{N}}$ with $S \subseteq D$ and $L(S) \subseteq L_0$ (as in (1.33), $L(S)$ is the level set of S) such that if ℓ_1 is the unique integer with $S(1) \subseteq A^{\ell_1}$, then*

$$\mathrm{dens}_{A^{\ell - \ell_1}}\big(\{s \in A^{<\mathbb{N}} : t^\frown s \in D \text{ for every } t \in S(1)\}\big) \geqslant \Theta_0/2 \tag{9.58}$$

for every $\ell \in L''$.

PROOF. Let $M = L \setminus L_0$ and set $M_\ell = \{m - \ell : m \in M\}$ for every $\ell \in L_0$. Moreover, for every $t \in \bigcup_{\ell \in L_0} A^\ell$ let

$$\mathcal{D}_t = \{s \in A^{<\mathbb{N}} : t^\frown s \in D\}. \tag{9.59}$$

Observe that for every $\ell \in L_0$ we have

$$\mathbb{E}_{t \in A^\ell} \mathrm{d}_{M_\ell}(\mathcal{D}_t) = \mathrm{d}_M(D) \tag{9.60}$$

while the fact that $\mathrm{dens}_{A^\ell}(D) \geqslant \delta$ for every $\ell \in L$ implies that

$$\mathrm{d}_M(D) \geqslant \delta \ \text{ and } \ \mathrm{d}_{L_0}(D) \geqslant \delta. \tag{9.61}$$

(Here, d_{M_ℓ}, d_M and d_{L_0} are the generalized Furstenberg–Weiss measures on $A^{<\mathbb{N}}$ associated with the sets M_ℓ, M and L_0 respectively.) On the other hand, since $|L| \geqslant h_{k,\delta}(n) \overset{(9.55)}{=} \Lambda_0 + \lceil 2\Theta_0^{-1} n \rceil$ and $|L_0| = \Lambda_0$, we obtain that

$$|M_\ell| = |M| \geqslant 2\Theta_0^{-1} n \tag{9.62}$$

for every $\ell \in L_0$. We consider the following cases.

CASE 1: *there exist $\ell_0 \in L_0$ and $t_0 \in A^{\ell_0}$ such that* $\mathrm{d}_{M_{\ell_0}}(\mathcal{D}_{t_0}) \geqslant \delta + \delta^2/4$. By (9.49) and (9.54), we see that $\Theta_0 \leqslant \delta^2/8$. Hence, in this case we have

$$|\{m \in M_{\ell_0} : \mathrm{dens}_{A^m}(\mathcal{D}_{t_0}) \geqslant \delta + \delta^2/8\}| \geqslant (\delta^2/8)\,|M_{\ell_0}| \overset{(9.62)}{\geqslant} n.$$

We set $L' = \{m \in M : \mathrm{dens}_{A^{m-\ell_0}}(D_{t_0}) \geqslant \delta + \delta^2/8\}$ and we observe that with this choice the first alternative of the lemma holds true.

CASE 2: *for every $\ell \in L_0$ and every $t \in A^\ell$ we have* $\mathrm{d}_{M_\ell}(\mathcal{D}_t) < \delta + \delta^2/4$. Combining (9.60) and (9.61) we see that $\mathbb{E}_{t \in A^\ell} \mathrm{d}_{M_\ell}(\mathcal{D}_t) \geqslant \delta$ for every $\ell \in L_0$. Thus, by Lemma E.3, in this case we have

$$|\{t \in A^\ell : \mathrm{d}_{M_\ell}(\mathcal{D}_t) \geqslant \delta/2\}| \geqslant (1 - \delta/2)k^\ell$$

for every $\ell \in L_0$. Therefore, setting

$$E = \bigcup_{\ell \in L_0} \{t \in A^\ell : t \in D \text{ and } \mathrm{d}_{M_\ell}(\mathcal{D}_t) \geqslant \delta/2\}, \tag{9.63}$$

we obtain that

$$\mathrm{d}_{L_0}(E) \geqslant \delta/2. \tag{9.64}$$

Now let

$$(\boldsymbol{\Omega}, \boldsymbol{\mu}) = \prod_{\ell \in L_0} \big(A^{<\mathbb{N}}, \mathrm{d}_{M_\ell}\big)$$

be the product of the discrete probability spaces $\big\{(A^{<\mathbb{N}}, \mathrm{d}_{M_\ell}) : \ell \in L_0\big\}$. For every $t \in E$ we define an event $\mathbf{D}_t$ of $\boldsymbol{\Omega}$ as follows. We set

$$\mathbf{D}_t = \prod_{\ell \in L_0} X_\ell^t \tag{9.65}$$

where $X_\ell^t = \mathcal{D}_t$ if $\ell = |t|$ and $X_\ell^t = A^{<\mathbb{N}}$ otherwise. Notice that for every $\ell \in L_0$ and every $t \in E \cap A^\ell$ we have

$$\boldsymbol{\mu}(\mathbf{D}_t) = \mathrm{d}_{M_\ell}(\mathcal{D}_t) \overset{(9.63)}{\geqslant} \delta/2. \tag{9.66}$$

Also recall that $|L_0| = \Lambda_0 \overset{(9.54)}{=} \Lambda(k,1,\delta^2/16)$. Hence, by (9.64), (9.66) and Proposition 9.21, there exists a Carlson–Simpson line S of $A^{<\mathbb{N}}$ which is contained in E and such that

$$\boldsymbol{\mu}\Big(\bigcap_{t\in S}\mathbf{D}_t\Big) \geqslant \Theta(k,1,\delta^2/16) \overset{(9.54)}{=} \Theta_0. \tag{9.67}$$

Notice, in particular, that the level set $L(S)$ of S is contained in L_0, and so if ℓ_1 is the unique integer with $S(1)\subseteq A^{\ell_1}$, then we have $\ell_1\in L_0$. By the definition of the events $\{\mathbf{D}_t : t\in E\}$ in (9.65), we obtain that

$$\mathrm{d}_{M_{\ell_1}}\Big(\bigcap_{t\in S(1)}\mathcal{D}_t\Big) = \boldsymbol{\mu}\Big(\bigcap_{t\in S(1)}\mathbf{D}_t\Big) \overset{(9.67)}{\geqslant} \Theta_0. \tag{9.68}$$

Thus, setting

$$M'_{\ell_1} = \Big\{m\in M_{\ell_1} : \mathrm{dens}_{A^m}\Big(\bigcap_{t\in S(1)}\mathcal{D}_t\Big)\geqslant \Theta_0/2\Big\}, \tag{9.69}$$

by (9.68) and Markov's inequality we have

$$|M'_{\ell_1}| \geqslant (\Theta_0/2)|M_{\ell_1}| \overset{(9.62)}{\geqslant} n. \tag{9.70}$$

Finally, let $L'' = \{\ell_1 + m : m\in M'_{\ell_1}\}$. We will show that L'' and S satisfy the second alternative of the lemma. Indeed, notice first that L'' is contained in $L\setminus L_0$ and $|L''|\geqslant n$. Since $\ell_1\in L_0$ we see, in particular, that $\ell_1<\min(L'')$. Also observe that for every $\ell\in L''$ we have $\ell-\ell_1\in M'_{\ell_1}$. Hence, by (9.59) and (9.69), we conclude that

$$\mathrm{dens}_{A^{\ell-\ell_1}}\big(\{s\in A^{<\mathbb{N}} : t^\frown s\in D \text{ for every } t\in S(1)\}\big)\geqslant \Theta_0/2 \tag{9.71}$$

for every $\ell\in L''$ which implies, of course, that the second alternative is satisfied. The above cases are exhaustive, and so the proof of Lemma 9.30 is completed. □

We proceed to the proof of Theorem 9.29.

Proof of Theorem 9.29. Fix $0<\delta\leqslant 1$ and set $N_0 = \mathrm{DCS}(k,m,\Theta_0/2)$. Let L be an arbitrary finite subset of $\mathbb{N}$ with $|L|\geqslant h_{k,\delta}^{(\lceil 8\delta^{-2}\rceil)}(N_0)$ and let D be a subset of $A^{<\mathbb{N}}$ such that $\mathrm{dens}_{A^\ell}(D)\geqslant\delta$ for every $\ell\in L$. By our assumption for the size of the set L and repeated applications of Lemma 9.30, there exist a subset L'' of L with $|L''|\geqslant N_0$ and a Carlson–Simpson line V of $A^{<\mathbb{N}}$ with $V\subseteq D$, such that if ℓ_1 is the unique integer with $V(1)\subseteq A^{\ell_1}$, then we have $\ell_1<\min(L'')$ and

$$\mathrm{dens}_{A^{\ell-\ell_1}}\big(\{s\in A^{<\mathbb{N}} : t^\frown s\in D \text{ for every } t\in V(1)\}\big)\geqslant \Theta_0/2 \tag{9.72}$$

for every $\ell\in L''$. By the choice of N_0 and (9.72), there exists an m-dimensional Carlson–Simpson space U of $A^{<\mathbb{N}}$ such that

$$U\subseteq \{s\in A^{<\mathbb{N}} : t^\frown s\in D \text{ for every } t\in V(1)\}. \tag{9.73}$$

Therefore, setting

$$S = V(0)\cup\bigcup_{t\in V(1)}\{t^\frown u : u\in U\}, \tag{9.74}$$

we see that S is a Carlson–Simpson space of $A^{<\mathbb{N}}$ of dimension $m+1$ which is contained in D. The proof of Theorem 9.29 is completed. $\square$

9.4. A probabilistic version of Theorem 9.2

9.4.1. Overview. In this subsection we will give an outline of the proof of Theorem 9.2. As we have already mentioned, it proceeds by induction on the cardinality of the finite alphabet A and is based on a density increment strategy. The initial case—that is, the estimation of the numbers $\mathrm{DCS}(2,1,\delta)$—is the content of Proposition 9.25. The next step is given in Theorem 9.29. Indeed, by Theorem 9.29, the proof of Theorem 9.2 reduces to the task of estimating the numbers $\mathrm{DCS}(k,1,\delta)$. To achieve this goal we follow an inductive scheme which can be described as follows:

$$\mathrm{DCS}(k,m,\beta) \text{ for every } m \text{ and every } \beta \ \Rightarrow \ \mathrm{DCS}(k+1,1,\delta). \tag{9.75}$$

Specifically, fix a positive integer k and $0<\delta\leqslant 1$ and assume, as in (9.75), that the numbers $\mathrm{DCS}(k,m,\beta)$ have been estimated for every choice of admissible parameters. Also let A be an alphabet with $|A|=k+1$ and let D be a subset of $A^{<\mathbb{N}}$ not containing a Carlson–Simpson line such that $\mathrm{dens}_{A^n}(D)\geqslant\delta$ for sufficiently many $n\in\mathbb{N}$. Our objective is to find a Carlson–Simpson space W of $A^{<\mathbb{N}}$ such that the density of D has been significantly increased in sufficiently many levels of W. Once this is done the numbers $\mathrm{DCS}(k+1,1,\delta)$ can be estimated with a standard iteration. This is enough, of course, to complete the proof of Theorem 9.2.

To this end we argue as follows. First, we will select a Carlson–Simpson space V of $A^{<\mathbb{N}}$ and a subset S of V which is the intersection of relatively few insensitive sets and correlates with the set D more than expected in many levels of V. (As in Subsection 8.3.1, we view S as a "simple" subset of V.) This is the content of Corollary 9.37 below. Next, we use this information to achieve the density increment. We will comment on this part of the proof in Subsection 9.5.1. At this point we simply mention that the statement of main interest is Corollary 9.55.

Concerning the proof of the first part, we note that it is reduced, essentially, to a "probabilistic" strengthening of our inductive assumptions. A straightforward modification of Example 8.1 shows that a global "probabilistic" version of Theorem 9.2 does not hold true, in the sense that there exist highly dense sets of words containing just a tiny portion of Carlson–Simpson lines. However, this problem can be effectively resolved locally, that is, by passing to an appropriately chosen Carlson–Simpson space. The philosophy is identical to that in Subsection 8.3.1 and the argument proceeds by applying the following three basic steps.

Step 1. By Szemerédi's regularity method, we show that a given dense set D of words over A is sufficiently pseudorandom. This enables us to model the set D as a family of measurable events $\{\mathscr{D}_t : t\in V\}$ in a probability space (Ω,Σ,μ) indexed by a Carlson–Simpson space V of $A^{<\mathbb{N}}$ of sufficiently large dimension. The main tools in this step are Lemmas 6.24 and 9.11.

Step 2. We use coloring arguments and our inductive assumptions to show that there exists a Carlson–Simpson subspace V' of V of prescribed dimension such that the events in the subfamily $\{\mathscr{D}_t : t \in V'\}$ are highly correlated. This step is based on Theorem 4.21 and Proposition 9.21.

Step 3. Let B be a sub-alphabet of A with k letters. We use a double counting argument to locate a Carlson–Simpson space U of $A^{<\mathbb{N}}$ with $\dim(U) = \dim(V')$ and satisfying one of the following alternatives. Either the density of D in U is increased (and so, we can directly proceed to the next iteration), or the density of D in U is preserved and, moreover, the set D contains plenty of Carlson–Simpson lines of $U \upharpoonright B$.

Finally, regarding the effectiveness of the proof, we notice that there exist primitive recursive (and fairly reasonable) upper bounds for all the results used in the steps described above. However, the argument yields very poor lower bounds for the correlation of the events $\{\mathscr{D}_t : t \in V'\}$ in the second step. These lower bounds are partly responsible for the weak estimate of the numbers $\mathrm{DCS}(k, m, \delta)$.

9.4.2. The main dichotomy. Let k, m be positive integers with $k \geqslant 2$ and assume that for every $0 < \beta \leqslant 1$ the number $\mathrm{DCS}(k, m, \beta)$ has been estimated. Hence, for every $0 < \delta \leqslant 1$ we may set

$$\vartheta = \vartheta(k, m, \delta) = \Theta(k, m, \delta/8) \ \text{ and } \ \eta = \eta(k, m, \delta) = \frac{\delta\vartheta}{30k} \tag{9.76}$$

where $\Theta(k, m, \delta/8)$ is as in (9.49). Moreover, let

$$\Lambda' = \Lambda(k, m, \delta/8) \overset{(9.48)}{=} \lceil 8\delta^{-1}\mathrm{DCS}(k, m, \delta/8) \rceil \tag{9.77}$$

and for every $n \in \mathbb{N}$ set

$$\ell(n, m) = \mathrm{CS}(k + 1, n + \Lambda', m, 2) + 1. \tag{9.78}$$

Next, define $g \colon \mathbb{N} \times \mathbb{N} \times (0, 1] \to \mathbb{N}$ by the rule

$$g(n, m, \varepsilon) = \mathrm{RegCS}(k + 1, 1, \ell(n, m), \varepsilon). \tag{9.79}$$

Finally, if A is a finite alphabet with $|A| \geqslant 2$, then for every finite-dimensional Carlson–Simpson space V of $A^{<\mathbb{N}}$ and every $1 \leqslant m \leqslant i \leqslant \dim(V)$ we set

$$\mathrm{SubCS}^0_m(V, i) = \big\{U \in \mathrm{SubCS}_m(V) : U(0) = V(0) \text{ and } U(m) \subseteq V(i)\big\}. \tag{9.80}$$

As in Subsection 1.5.3, for every sub-alphabet B of A with $|B| \geqslant 2$ and every finite-dimensional Carlson–Simpson space V of $A^{<\mathbb{N}}$ let $V \upharpoonright B$ be the restriction of V on B. Recall that if I_V is the canonical isomorphism associated with V, then the map $\mathrm{I}_V \colon B^{<\dim(V)+1} \to V \upharpoonright B$ is a bijection, and so we may identify $V \upharpoonright B$ with $B^{<\dim(V)+1}$. Taking into account these remarks, we set

$$\mathrm{SubCS}^0_m(V \upharpoonright B, i) = \big\{\mathrm{I}_V(U) : U \in \mathrm{SubCS}^0_m(B^{<\dim(V)+1}, i)\big\}.$$

We are now ready to state the main result of this section.

Proposition 9.31. *Let k,m be positive integers with $k \geqslant 2$ and assume that for every $0<\beta\leqslant 1$ the number* $\mathrm{DCS}(k,m,\beta)$ *has been estimated.*

Let $0<\delta\leqslant 1$ and define ϑ, η and Λ' as in (9.76) *and* (9.77) *respectively. Also let $n\in\mathbb{N}$ with $n\geqslant 1$ and L_0 a finite subset of $\mathbb{N}$ such that*

$$|L_0| \geqslant g\big(\lceil \eta^{-4} n\rceil, m, \eta^2/2\big) \tag{9.81}$$

where g is as in (9.79)*. If A is an alphabet with $|A|=k+1$, B is a subset of A with $|B|=k$ and $D\subseteq A^{<\mathbb{N}}$ satisfies $\mathrm{dens}_{A^l}(D)\geqslant\delta$ for every $l\in L_0$, then there exist a Carlson–Simpson space W of $A^{<\mathbb{N}}$ with $\dim(W)=\lceil\eta^{-4}n\rceil+\Lambda'$ and $I\subseteq\{m,\dots,\dim(W)\}$ with $|I|\geqslant n$ such that either*

(a) *for every $i\in I$ we have $\mathrm{dens}_{W(i)}(D)\geqslant \delta+\eta^2/2$, or*
(b) *for every $i\in I$ we have $\mathrm{dens}_{W(i)}(D)\geqslant\delta-2\eta$ and, moreover,*

$$\mathrm{dens}\big(\{V\in\mathrm{SubCS}^0_m(W\upharpoonright B,i) : V\subseteq D\}\big)\geqslant\vartheta/2.$$

The following lemma is the first step of the proof of Proposition 9.31.

Lemma 9.32. *Let $k,m,\delta,\vartheta,\eta,\Lambda',n$ and L_0 be as in Proposition* 9.31*. Also let A be an alphabet with $|A|=k+1$, $B\subseteq A$ with $|B|=k$ and $D\subseteq A^{<\mathbb{N}}$ such that $\mathrm{dens}_{A^l}(D)\geqslant\delta$ for every $l\in L_0$. Then there exist a subset L of L_0 and a Carlson–Simpson subspace S of $A^{<|L|}$ of dimension $\lceil\eta^{-4}n\rceil+\Lambda'$ such that, denoting by $\mathrm{c}_L\colon A^{<|L|}\times\Omega_L\to A^{<\mathbb{N}}$ the convolution operation associated with L and setting $\mathscr{D}=\mathrm{c}_L^{-1}(D)$, the following are satisfied.*

(a) *For every $t\in A^{<|L|}$ we have $\mathrm{dens}_{\Omega_L}(\mathscr{D}_t)\geqslant\delta-\eta^2/2$ where $\mathscr{D}_t$ is the section of $\mathscr{D}$ at t.*
(b) *For every $U\in\mathrm{SubCS}_m(S\upharpoonright B)$ we have*

$$\mathrm{dens}_{\Omega_L}\Big(\bigcap_{t\in U}\mathscr{D}_t\Big)\geqslant\vartheta.$$

Proof. By (9.81), the definition of the function g in (9.79) and Lemma 6.24, there exists a subset L of L_0 with

$$|L| = \ell(\lceil\eta^{-4}n\rceil,m) \overset{(9.78)}{=} \mathrm{CS}(k+1,\lceil\eta^{-4}n\rceil+\Lambda',m,2)+1 \tag{9.82}$$

and such that the family $\mathcal{F}\coloneqq\{A\}$ is $(\eta^2/2,L)$-regular. Hence, by Lemma 9.11, we see that part (a) is satisfied for the set L.

Next, let $V=A^{<|L|}$ and $d=\lceil\eta^{-4}n\rceil+\Lambda'$. By (9.77) and (9.82), we have

$$d\geqslant\Lambda(k,m,\delta/8)-1 \quad\text{and}\quad \dim(V)=|L|-1=\mathrm{CS}(k+1,d,m,2).$$

Moreover, by part (a), for every $t\in V$ we have

$$\mathrm{dens}_{\Omega_L}(\mathscr{D}_t)\geqslant\delta-\eta^2/2 \overset{(9.76)}{\geqslant} \delta/2.$$

Therefore, by Corollary 9.23, there exists a d-dimensional Carlson–Simpson subspace S of $A^{<|L|}$ such that for every $U\in\mathrm{SubCS}_m(S)$ we have

$$\mathrm{dens}_{\Omega_L}\Big(\bigcap_{t\in U\upharpoonright B}\mathscr{D}_t\Big)\geqslant\Theta(k,m,\delta/8)$$

where $\Theta(k,m,\delta/8)$ is as in (9.49). Since

$$\mathrm{SubCS}_m(S\upharpoonright B)\subseteq\{U\upharpoonright B: U\in\mathrm{SubCS}_m(S)\}$$

and $\vartheta \overset{(9.76)}{=} \Theta(k,m,\delta/8)$, we conclude that part (b) is also satisfied. The proof of Lemma 9.32 is completed. $\square$

Now let L be a finite subset of $\mathbb{N}$ with $|L|\geqslant 2$ and consider the convolution operation $\mathrm{c}_L\colon A^{<|L|}\times\Omega_L\to A^{<\mathbb{N}}$ associated with L. Also let S be a Carlson–Simpson subspace of $A^{<|L|}$. As in Lemma 9.8, for every $\omega\in\Omega_L$ we set

$$S_\omega=\{\mathrm{c}_L(t,\omega): t\in S\} \tag{9.83}$$

and we recall that S_ω is a Carlson–Simpson space of $A^{<\mathbb{N}}$ with $\dim(S_\omega)=\dim(S)$. We have the following lemma.

LEMMA 9.33. *Let $k\in\mathbb{N}$ with $k\geqslant 2$ and A an alphabet with $|A|=k+1$. Also let L be a finite subset of $\mathbb{N}$ with $|L|\geqslant 2$ and $\mathrm{c}_L\colon A^{<|L|}\times\Omega_L\to A^{<\mathbb{N}}$ the convolution operation associated with L. Finally, let S be a Carlson–Simpson subspace of $A^{<|L|}$ and $D\subseteq A^{<\mathbb{N}}$. For every $\omega\in\Omega_L$ let S_ω be as in (9.83), set $\mathscr{D}=\mathrm{c}_L^{-1}(D)$ and for every $t\in A^{<|L|}$ let $\mathscr{D}_t$ be the section of $\mathscr{D}$ at t. Then the following hold.*

(a) *For every $i\in\{0,\dots,\dim(S)\}$ we have*

$$\mathbb{E}_{\omega\in\Omega_L}\mathrm{dens}_{S_\omega(i)}(D)=\mathbb{E}_{t\in S(i)}\mathrm{dens}_{\Omega_L}(\mathscr{D}_t).$$

(b) *For every $1\leqslant m\leqslant i\leqslant\dim(S)$ and every $B\subseteq A$ with $|B|\geqslant 2$ we have*

$$\mathbb{E}_{\omega\in\Omega_L}\mathrm{dens}\big(\{V\in\mathrm{SubCS}^0_m(S_\omega\upharpoonright B,i): V\subseteq D\}\big)= \mathbb{E}_{U\in\mathrm{SubCS}^0_m(S\upharpoonright B,i)}\,\mathrm{dens}_{\Omega_L}\Big(\bigcap_{t\in U}\mathscr{D}_t\Big).$$

PROOF. (a) Let $i\in\{0,\dots,\dim(S)\}$ be arbitrary. By Lemma 9.10, we have

$$\mathbb{E}_{\omega\in\Omega_L}\mathrm{dens}_{S_\omega(i)}(D)=\mathrm{dens}_{\mathrm{c}_L(S(i)\times\Omega_L)}(D). \tag{9.84}$$

As in (9.6), for every $t\in A^{<|L|}$ we set $\mathcal{C}_t=\{\mathrm{c}_L(t,\omega):\omega\in\Omega_L\}$. Notice that

$$\mathrm{c}_L(S(i)\times\Omega_L)=\bigcup_{t\in S(i)}\mathrm{c}_L(\{t\}\times\Omega_L)=\bigcup_{t\in S(i)}\mathcal{C}_t. \tag{9.85}$$

Next observe that $S(i)\subseteq A^l$ for some $l\in\{0,\dots,|L|-1\}$. Therefore, by Fact 9.4, we have $|\mathcal{C}_t|=|\mathcal{C}_{t'}|$ for every $t,t'\in S(i)$. It follows that the family $\{\mathcal{C}_t: t\in S(i)\}$ is an equipartition of $\mathrm{c}_L(S(i)\times\Omega_L)$ and so

$$\mathrm{dens}_{\mathrm{c}_L(S(i)\times\Omega_L)}(D)=\mathbb{E}_{t\in S(i)}\,\mathrm{dens}_{\mathcal{C}_t}(D). \tag{9.86}$$

Finally, by Lemma 9.7, for every $t\in S(i)$ we have

$$\mathrm{dens}_{\mathcal{C}_t}(D)=\mathrm{dens}_{\{t\}\times\Omega_L}(\mathscr{D})=\mathrm{dens}_{\Omega_L}(\mathscr{D}_t). \tag{9.87}$$

Combining (9.84), (9.86) and (9.87), we conclude that part (a) is satisfied.

(b) Fix $1\leqslant m\leqslant i\leqslant\dim(S)$ and $B\subseteq A$ with $|B|\geqslant 2$. We set

$$\mathcal{P}=\big\{(U,\omega)\in\mathrm{SubCS}^0_m(S\upharpoonright B,i)\times\Omega_L: U_\omega\subseteq D\big\}$$

where, as in (9.83), we have $U_\omega = \{\mathrm{c}_L(u,\omega) : u \in U\}$. Moreover, for every $U \in \mathrm{SubCS}^0_m(S \upharpoonright B, i)$ and every $\omega \in \Omega_L$ let

$$\mathcal{P}_U = \{\omega \in \Omega_L : (U,\omega) \in \mathcal{P}\} \ \text{ and } \ \mathcal{P}_\omega = \big\{U \in \mathrm{SubCS}^0_m(S \upharpoonright B, i) : (U,\omega) \in \mathcal{P}\big\}$$

be the sections of $\mathcal{P}$ at U and ω respectively. Notice that

$$\omega \in \mathcal{P}_U \Leftrightarrow (U,\omega) \in \mathcal{P} \Leftrightarrow U_\omega \subseteq D \Leftrightarrow \omega \in \bigcap_{t\in U} \mathscr{D}_t$$

which implies that for every $U \in \mathrm{SubCS}^0_m(S \upharpoonright B, i)$ we have

$$\mathcal{P}_U = \bigcap_{t\in U} \mathscr{D}_t. \tag{9.88}$$

Also observe that for every $\omega \in \Omega_L$ the map

$$\mathrm{SubCS}^0_m(S \upharpoonright B, i) \ni U \mapsto U_\omega \in \mathrm{SubCS}^0_m(S_\omega \upharpoonright B, i)$$

is a bijection. Hence, for every $\omega \in \Omega_L$ we have

$$\begin{aligned}
\mathrm{dens}\big(\{V \in \mathrm{SubCS}^0_m(S_\omega \upharpoonright B, i) : V \subseteq D\}\big) &= \\
&= \frac{|\{V \in \mathrm{SubCS}^0_m(S_\omega \upharpoonright B, i) : V \subseteq D\}|}{|\mathrm{SubCS}^0_m(S_\omega \upharpoonright B, i)|} \\
&= \frac{|\{U \in \mathrm{SubCS}^0_m(S \upharpoonright B, i) : U_\omega \subseteq D\}|}{|\mathrm{SubCS}^0_m(S \upharpoonright B, i)|} = \mathrm{dens}(\mathcal{P}_\omega). \qquad (9.89)
\end{aligned}$$

Therefore, we conclude that

$$\begin{aligned}
\mathbb{E}_{\omega\in\Omega_L} \mathrm{dens}\big(\{V \in \mathrm{SubCS}^0_m(S_\omega \upharpoonright B, i) : V \subseteq D\}\big) &= \\
&\overset{(9.89)}{=} \mathbb{E}_{\omega\in\Omega_L}\, \mathrm{dens}(\mathcal{P}_\omega) = \mathbb{E}_{U\in \mathrm{SubCS}^0_m(S\upharpoonright B,i)}\, \mathrm{dens}_{\Omega_L}(\mathcal{P}_U) \\
&\overset{(9.88)}{=} \mathbb{E}_{U\in \mathrm{SubCS}^0_m(S\upharpoonright B,i)}\, \mathrm{dens}_{\Omega_L}\Big(\bigcap_{t\in U} \mathscr{D}_t\Big).
\end{aligned}$$

The proof of Lemma 9.33 is completed. $\square$

We are now ready to give the proof of Proposition 9.31.

Proof of Proposition 9.31. We fix $D \subseteq A^{<\mathbb{N}}$ such that $\mathrm{dens}_{A^l}(D) \geqslant \delta$ for every $l \in L_0$. Set $d = \lceil \eta^{-4} n \rceil + \Lambda'$ and let L and S be as in Lemma 9.32 when applied to the set D. Notice, in particular, that $\dim(S) = d$. Invoking the first parts of Lemmas 9.32 and 9.33, for every $i \in \{0, \dots, d\}$ we have

$$\mathbb{E}_{\omega\in\Omega_L} \mathrm{dens}_{S_\omega(i)}(D) \geqslant \delta - \eta^2/2. \tag{9.90}$$

On the other hand, by the second parts of the aforementioned lemmas, we see that

$$\mathbb{E}_{\omega\in\Omega_L} \mathrm{dens}\big(\{V \in \mathrm{SubCS}^0_m(S_\omega \upharpoonright B, i) : V \subseteq D\}\big) \geqslant \vartheta \tag{9.91}$$

for every $i \in \{m, \dots, d\}$.

Now set $J = \{m, \dots, d\}$ and observe that

$$|J| = d - m + 1 \geqslant \lceil \eta^{-4} n \rceil. \tag{9.92}$$

Moreover, for every $i \in J$ set

(a) $\Omega_{i,0} = \{\omega \in \Omega_L : \mathrm{dens}_{S_\omega(i)}(D) \geqslant \delta + \eta^2/2\}$,
(b) $\Omega_{i,1} = \{\omega \in \Omega_L : \mathrm{dens}_{S_\omega(i)}(D) \geqslant \delta - 2\eta\}$, and
(c) $\Omega_{i,2} = \big\{\omega \in \Omega_L : \mathrm{dens}\big(\{V \in \mathrm{SubCS}^0_m(S_\omega \upharpoonright B, i) : V \subseteq D\}\big) \geqslant \vartheta/2\big\}$.

Finally, let $J_0 = \{i \in J : \mathrm{dens}_{\Omega_L}(\Omega_{i,0}) \geqslant \eta^3\}$. We consider the following cases.

Case 1: *we have* $|J_0| \geqslant |J|/2$. By Lemma E.4, there exists $\omega_0 \in \Omega_L$ such that

$$|\{i \in J_0 : \omega_0 \in \Omega_{i,0}\}| \geqslant \eta^3 |J_0| \geqslant \eta^3 \frac{|J|}{2} \overset{(9.92)}{\geqslant} \frac{\eta^3 \lceil \eta^{-4} n \rceil}{2} \overset{(9.76)}{\geqslant} n.$$

We set $I = \{i \in J : \omega_0 \in \Omega_{i,0}\}$ and $W = S_{\omega_0}$. Clearly, with these choices the first part of the proposition is satisfied.

Case 2: *we have* $|J_0| < |J|/2$. In this case we set $K_0 = J \setminus J_0$. Let $i \in K_0$ be arbitrary and notice that

$$\mathrm{dens}_{\Omega_L}(\Omega_{i,0}) = \mathrm{dens}\big(\{\omega \in \Omega_L : \mathrm{dens}_{S_\omega(i)}(D) \geqslant \delta + \eta^2/2\}\big) < \eta^3. \tag{9.93}$$

By (9.90), (9.93) and Lemma E.3, we obtain that $\mathrm{dens}_{\Omega_L}(\Omega_{i,1}) \geqslant 1 - \eta$. On the other hand, by (9.91), we have $\mathrm{dens}_{\Omega_L}(\Omega_{i,2}) \geqslant \vartheta/2$. Therefore, by the choice of η in (9.76), we conclude that $\mathrm{dens}_{\Omega_L}(\Omega_{i,1} \cap \Omega_{i,2}) \geqslant \vartheta/4$ for every $i \in K_0$. By another application of Lemma E.4, we see that there exists $\omega_1 \in \Omega_L$ such that

$$|\{i \in K_0 : \omega_1 \in \Omega_{i,1} \cap \Omega_{i,2}\}| \geqslant \frac{\vartheta}{4}|K_0| \geqslant \frac{\vartheta |J|}{8} \overset{(9.92)}{\geqslant} \frac{\vartheta \lceil \eta^{-4} n \rceil}{8} \overset{(9.76)}{\geqslant} n.$$

We set $I = \{i \in K_0 : \omega_1 \in \Omega_{i,1} \cap \Omega_{i,2}\}$ and $W = S_{\omega_1}$ and we notice that with these choices part (b) is satisfied. The above cases are exhaustive and so the proof of Proposition 9.31 is completed. □

9.4.3. Obtaining insensitive sets. Let A be a finite alphabet with $|A| \geqslant 2$ and $a, b \in A$ with $a \neq b$. Recall that, by Definition 2.2, a set S of words over A is said to (a,b)-insensitive provided that for every $z \in S$ and every $y \in A^{<\mathbb{N}}$ if z and y are (a,b)-equivalent, then $y \in S$. This concept can be relativized to any Carlson–Simpson space of $A^{<\mathbb{N}}$ as follows.

Definition 9.34. *Let A be a finite alphabet with $|A| \geqslant 2$ and $a, b \in A$ with $a \neq b$. Also let $S \subseteq A^{<\mathbb{N}}$ and V a Carlson–Simpson space of $A^{<\mathbb{N}}$. We say that S is* (a,b)-insensitive in V *if $\mathrm{I}_V^{-1}(S \cap V)$ is an (a,b)-insensitive subset of $A^{<\mathbb{N}}$ where I_V is the canonical isomorphism associated with V (see Definition* 1.10*).*

Now let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every $0 < \beta \leqslant 1$ the number $\mathrm{DCS}(k,1,\beta)$ has been estimated. For every $0 < \delta \leqslant 1$ we set

$$\vartheta_1 = \vartheta_1(k,\delta) = \vartheta(k,1,\delta) \ \text{ and } \ \eta_1 = \eta_1(k,\delta) = \eta(k,1,\delta) \tag{9.94}$$

where $\vartheta(k,1,\delta)$ and $\eta(k,1,\delta)$ are as in (9.76). Also define $g_1 \colon \mathbb{N} \times (0,1] \to \mathbb{N}$ by

$$g_1(n,\varepsilon) = g(n,1,\varepsilon) \tag{9.95}$$

where g is as in (9.79). We have the following lemma.

LEMMA 9.35. *Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every $0 < \beta \leqslant 1$ the number* $\mathrm{DCS}(k,1,\beta)$ *has been estimated.*

Let $0<\delta\leqslant 1$, $n\in\mathbb{N}$ with $n\geqslant 1$ and L_0 a finite subset of $\mathbb{N}$ such that

$$|L_0| \geqslant g_1\big(\lceil \eta_1^{-4}(k+1)n\rceil, \eta_1^2/2\big)$$

where η_1 and g_1 are as in (9.94) *and* (9.95) *respectively. Also let A be an alphabet with $|A| = k+1$, $a \in A$ and set $B = A\setminus\{a\}$. Finally, let $D \subseteq A^{<\mathbb{N}}$ be such that $\mathrm{dens}_{A^l}(D) \geqslant \delta$ for every $l \in L_0$ and assume that D contains no Carlson–Simpson line of $A^{<\mathbb{N}}$. Assume, moreover, that for every finite-dimensional Carlson–Simpson space W of $A^{<\mathbb{N}}$ we have*

$$\big|\big\{i \in \{0,\dots,\dim(W)\} : \mathrm{dens}_{W(i)}(D) \geqslant \delta + \eta_1^2/2\big\}\big| < n \tag{9.96}$$

Then there exist a finite-dimensional Carlson–Simpson space V of $A^{<\mathbb{N}}$, a subset C of V and a subset J of $\{0,\dots,\dim(V)\}$ with the following properties.

(a) *We have $|J| \geqslant n$.*
(b) *We have $C = \bigcap_{b\in B} C_b$ where C_b is (a,b)-insensitive in V for every $b \in B$. Moreover, $\mathrm{dens}_{V(j)}(C) \geqslant \vartheta_1/2$ for every $j \in J$ where ϑ_1 is as in* (9.94).
(c) *The sets D and C are disjoint.*
(d) *For every $j \in J$ we have $\mathrm{dens}_{V(j)}(D) \geqslant \delta - 5k\eta_1$.*

The proof of Lemma 9.35 is based on Proposition 9.31. Before we proceed to the details we need to introduce some pieces of notation and some terminology. Let A be an alphabet with $|A| = k+1$ and fix $a \in A$. Also let d be a positive integer and let W be a d-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$. Consider the canonical isomorphism $\mathrm{I}_W\colon A^{<d+1} \to W$ associated with W and set

$$W[a] = \big\{\mathrm{I}_W(s) : s \in A^{<d+1} \text{ with } |s| \geqslant 1 \text{ and } s(0) = a\big\}. \tag{9.97}$$

Notice that if $\dim(W) \geqslant 2$, then $W[a]$ is a Carlson–Simpson subspace of W with $\dim(W[a]) = \dim(W) - 1$. On the other hand, if W is a Carlson–Simpson line, then $W[a]$ is the singleton $\{\mathrm{I}_W(a)\}$; we will identify in this case $W[a]$ with $\mathrm{I}_W(a)$. Next, set $B = A\setminus\{a\}$ and note that, by Fact 1.14, for every Carlson–Simpson subspace[1] R of $W \upharpoonright B$ there exists a unique Carlson–Simpson subspace U of W such that $R = U \upharpoonright B$. We will call this unique Carlson–Simpson space U as the *extension* of R and we will denote it by $\overline{R}$. We have the following elementary fact.

FACT 9.36. *Let A, a and B be as in Lemma 9.35 and W a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ with $\dim(W) \geqslant 2$. We set $V = W[a]$. Then for every $j \in \{0,\dots,\dim(V)\}$ and every $R \in \mathrm{SubCS}_1^0(W\upharpoonright B, j+1)$ we have $\overline{R}[a] \in V(j)$. Moreover, the map*

$$\mathrm{SubCS}_1^0(W\upharpoonright B, j+1) \ni R \mapsto \overline{R}[a] \in V(j)$$

is a bijection.

We are ready to give the proof of Lemma 9.35.

[1] Recall that, by (1.41), every Carlson–Simpson subspace of $W \upharpoonright B$ is of the form $\mathrm{I}_W(S)$ for some (unique) Carlson–Simpson subspace S of $B^{<d+1}$.

Proof of Lemma 9.35. By our assumptions and Proposition 9.31 applied to the set D and "$m=1$", there exist a finite-dimensional Carlson–Simpson space W of $A^{<\mathbb{N}}$ and $I \subseteq \{1,\dots,\dim(W)\}$ with $|I| \geqslant (k+1)n$ such that

$$\mathrm{dens}_{W(i)}(D) \geqslant \delta - 2\eta_1 \ \text{ and } \ \mathrm{dens}\big(\{R \in \mathrm{SubCS}_1^0(W \upharpoonright B, i) : R \subseteq D\}\big) \geqslant \vartheta_1/2$$

for every $i \in I$. For every $b \in B$ let $V_b = W[b]$. We set

$$V = W[a], \tag{9.98}$$

$$C = \bigcup_{i\in I} \big\{\overline{R}[a] : R \in \mathrm{SubCS}_1^0(W \upharpoonright B, i) \text{ with } R \subseteq D\big\} \tag{9.99}$$

and

$$J = \big\{j \in \{0,\dots,\dim(V)\} : \mathrm{dens}_{V(j)}(D) \geqslant \delta - 5k\eta_1 \text{ and } j+1 \in I\big\} \tag{9.100}$$

and we claim that V, C and J are as desired. First we will show that $|J| \geqslant n$. To this end, set $J_0 = \{j \in \{0,\dots,\dim(V)\} : j+1 \in I\}$ and observe that $J \subseteq J_0$ and

$$|J_0| \geqslant (k+1)n. \tag{9.101}$$

Let $j \in J_0 \setminus J$ be arbitrary. Notice that $\mathrm{dens}_{V(j)}(D) < \delta - 5k\eta_1$. On the other hand, we have $j+1 \in I$ and so, by the choice of I,

$$\frac{1}{k+1}\Big(\mathrm{dens}_{V(j)}(D) + \sum_{b\in B} \mathrm{dens}_{V_b(j)}(D)\Big) = \mathrm{dens}_{W(j+1)}(D) \geqslant \delta - 2\eta_1.$$

It follows that there exists $b_j \in B$ such that $\mathrm{dens}_{V_{b_j}(j)}(D) \geqslant \delta + \eta_1$. Since $|B| = k$, by the classical pigeonhole principle, there exists $b_0 \in B$ such that

$$|\{j \in J_0 \setminus J : \mathrm{dens}_{V_{b_0}(j)}(D) \geqslant \delta + \eta_1\}| \geqslant \frac{|J_0 \setminus J|}{k} \overset{(9.101)}{\geqslant} n + \frac{n - |J|}{k}. \tag{9.102}$$

Moreover, by (9.96), we have

$$|\{j \in J_0 \setminus J : \mathrm{dens}_{V_{b_0}(j)}(D) \geqslant \delta + \eta_1\}| < n. \tag{9.103}$$

Combining (9.102) and (9.103) we conclude that $|J| \geqslant n$.

We continue with the proof of part (b). Let $b \in B$ be arbitrary. For every $l \in \mathbb{N}$ and every $s \in A^l$ let $s^{a\to b}$ be the unique element of B^l obtained by replacing all appearances of a in s by b. We set

$$C_b = \Big\{\mathrm{I}_V(s) : s \in \bigcup_{i\in I} A^{i-1} \text{ and } s^{a\to b} \in \mathrm{I}_{V_b}^{-1}(D)\Big\}$$

where I_V and I_{V_b} are the canonical isomorphisms associated with V and V_b respectively (see Definition 1.10). Observe that C_b is (a,b)-insensitive in V. We will show that C coincides with $\bigcap_{b\in B} C_b$. Indeed, notice first that $C \subseteq \bigcap_{b\in B} C_b$. To see the other inclusion, fix $t \in \bigcap_{b\in B} C_b$ and set $s = \mathrm{I}_V^{-1}(t)$. Let i be the unique element of I such that $s \in A^{i-1}$ and define

$$R_t = \{W(0)\} \cup \big\{\mathrm{I}_{V_b}(s^{a\to b}) : b \in B\big\}.$$

Observe that $R_t \in \mathrm{SubCS}_1^0(W \upharpoonright B, i)$. By the choice of W, we have $W(0) \in D$ while the fact that $t \in \bigcap_{b\in B} C_b$ yields that $\mathrm{I}_{V_b}(s^{a\to b}) \in D$ for every $b \in B$. Thus, we see that $R_t \subseteq D$. Since $t = \overline{R}_t[a]$ we conclude that $t \in C$. Finally, let $j \in J$. Recall that $j+1 \in I$ and notice that

$$\mathrm{dens}_{V(j)}(C) = \mathrm{dens}_{V(j)}\big(\{\overline{R}[a] : R \in \mathrm{SubCS}^0_1(W \upharpoonright B, j+1) \text{ with } R \subseteq D\}\big).$$

Hence, by Fact 9.36 and the choice of I, we obtain that

$$\mathrm{dens}_{V(j)}(C) = \mathrm{dens}\big(\{R \in \mathrm{SubCS}^0_1(W \upharpoonright B, j+1) : R \subseteq D\}\big) \geqslant \vartheta_1/2$$

as desired.

Now the fact that D and C are disjoint follows from our assumption that the set D contains no Carlson–Simpson line of $A^{<\mathbb{N}}$ and the definition of C in (9.99). Finally, part (d) is an immediate consequence of (9.100). The proof of Lemma 9.35 is completed. $\square$

9.4.4. Consequences. In this subsection we will summarize what we have achieved in Proposition 9.31 and Lemma 9.35. The resulting statement is the first main step of the proof of the inductive scheme described in (9.75).

Corollary 9.37. *Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every $0 < \beta \leqslant 1$ the number $\mathrm{DCS}(k,1,\beta)$ has been estimated.*

Let $0 < \delta \leqslant 1$, $n \in \mathbb{N}$ with $n \geqslant 1$ and L_0 a finite subset of $\mathbb{N}$ such that

$$|L_0| \geqslant g_1\big(\lceil \eta_1^{-4}(k+1)kn\rceil, \eta_1^2/2\big) \tag{9.104}$$

where η_1 and g_1 are as in (9.94) and (9.95) respectively. Also let A be an alphabet with $|A| = k+1$, $a \in A$ and set $B = A \setminus \{a\}$. Finally, let $D \subseteq A^{<\mathbb{N}}$ be such that $\mathrm{dens}_{A^l}(D) \geqslant \delta$ for every $l \in L_0$. Assume that D contains no Carlson–Simpson line of $A^{<\mathbb{N}}$. Then there exist a finite-dimensional Carlson–Simpson space V of $A^{<\mathbb{N}}$, a subset S of V and a subset I of $\{0,\dots,\dim(V)\}$ with the following properties.

(a) *We have $|I| \geqslant n$.*
(b) *We have $S = \bigcap_{b\in B} S_b$ where S_b is (a,b)-insensitive in V for every $b \in B$.*
(c) *For every $i \in I$ we have $\mathrm{dens}_{V(i)}(D \cap S) \geqslant (\delta + \eta_1^2/2)\,\mathrm{dens}_{V(i)}(S)$ and $\mathrm{dens}_{V(i)}(S) \geqslant \eta_1^2/2$.*

Proof. Assume, first, that there exists a finite-dimensional Carlson–Simpson space W of $A^{<\mathbb{N}}$ such that

$$|\{i \in \{0,\dots,\dim(W)\} : \mathrm{dens}_{W(i)}(D) \geqslant \delta + \eta_1^2/2\}| \geqslant kn.$$

In this case we set $V = W$, $I = \{i \in \{0,\dots,\dim(W)\} : \mathrm{dens}_{W(i)}(D) \geqslant \delta + \eta_1^2/2\}$ and $S_b = V$ for every $b \in B$. It is clear that with these choices the result follows.

Otherwise, by Lemma 9.35, there exist a finite-dimensional Carlson–Simpson space V of $A^{<\mathbb{N}}$, a subset J of $\{0,\dots,\dim(V)\}$ with $|J| \geqslant kn$ and a set $C = \bigcap_{b\in B} C_b$ such that: (i) $D \cap C = \emptyset$, (ii) C_b is (a,b)-insensitive in V for every $b \in B$ and $\mathrm{dens}_{V(j)}(C) \geqslant \vartheta_1/2$ for every $j \in J$, and (iii) $\mathrm{dens}_{V(j)}(D) \geqslant \delta - 5k\eta_1$ for every $j \in J$. (Here, ϑ_1 and η_1 are as in (9.94).) In particular, for every $j \in J$ we have

$$\frac{\mathrm{dens}_{V(j)}(D)}{\mathrm{dens}_{V(j)}(V \setminus C)} \geqslant \frac{\delta - 5k\eta_1}{1 - \vartheta_1/2} \geqslant (\delta - 5k\eta_1)(1 + \vartheta_1/2) \geqslant \delta + 7k\eta_1.$$

Let $\{b_1, \dots, b_k\}$ be an enumeration of the alphabet B. We set $Q_1 = V \setminus C_{b_1}$ and $Q_r = C_{b_1} \cap \dots \cap C_{b_{r-1}} \cap (V \setminus C_{b_r})$ for every $r \in \{2, \dots, k\}$, and we observe that the family $\{Q_1, \dots, Q_r\}$ is a partition of $V \setminus C$. Let $j \in J$ be arbitrary. By Lemma E.6 applied for "$\varepsilon = k\eta_1$", there exists $r_j \in [k]$ such that

$$\mathrm{dens}_{V(j)}(D \cap Q_{r_j}) \geqslant (\delta + 6k\eta_1)\, \mathrm{dens}_{V(j)}(Q_{r_j}) \tag{9.105}$$

and

$$\mathrm{dens}_{V(j)}(Q_{r_j}) \geqslant (\delta - 5k\eta_1)\, \eta_1/4. \tag{9.106}$$

Hence, by the classical pigeonhole principle, there exist $r_0 \in [k]$ and $I \subseteq J$ with $|I| \geqslant |J|/k \geqslant n$ and such that $r_i = r_0$ for every $i \in I$. We set $S = Q_{r_0}$, $S_r = C_r$ if $r < r_0$, $S_{r_0} = V \setminus C_{r_0}$ and $S_r = V$ if $r > r_0$. Notice, first, that S_r is (a, b_r)-insensitive in V for every $r \in [k]$. Also observe that $S_1 \cap \dots \cap S_k = S$. Finally, by the choice of η_1 in (9.94), for every $i \in I$ we have

$$\mathrm{dens}_{V(i)}(D \cap S) \overset{(9.105)}{\geqslant} (\delta + 6k\eta_1)\, \mathrm{dens}_{V(i)}(S) \geqslant (\delta + \eta_1^2/2)\, \mathrm{dens}_{V(i)}(S)$$

and

$$\mathrm{dens}_{V(i)}(S) \overset{(9.106)}{\geqslant} (\delta - 5k\eta_1)\, \eta_1/4 \geqslant \eta_1^2/2.$$

The proof of Corollary 9.37 is completed. □

9.5. An exhaustion procedure: achieving the density increment

9.5.1. Motivation. This section is devoted to the proof of the second part of the inductive scheme described in (9.75). Recall that the first part of this inductive scheme is the content of Corollary 9.37. Specifically, by Corollary 9.37, if A is an alphabet with $k+1$ letters and D is a dense set of words over A not containing a Carlson–Simpson line, then there exist a finite-dimensional Carlson–Simpson space V of $A^{<\mathbb{N}}$ and a "simple" subset S of V (precisely, S is the intersection of few insensitive sets) which correlates with D more than expected in many levels of V. Our goal in this section is to use this information to achieve density increment for the set D. In order to do so, a natural strategy is to produce an "almost tiling" of S, that is, to construct a collection $\mathcal{V}$ of pairwise disjoint Carlson–Simpson spaces of sufficiently large dimension which are all contained in S and are such that the set $S \setminus \cup\mathcal{V}$ is essentially negligible[2]. Once this is done, one then expects to be able to find a Carlson–Simpson space W which belongs to the family $\mathcal{V}$ and is such that the density of D has been increased in sufficiently many levels of W. However, this is not possible in general, as is shown in the following example.

EXAMPLE 9.1. Fix a positive integer m and $0 < \varepsilon \leqslant 1$. Let A be a finite alphabet with $|A| \geqslant 2$ and set $k = |A|$. Also let q, ℓ be positive integers with $q \geqslant \ell$ and such that $(k^\ell - 1)k^{\ell - q} \leqslant \varepsilon$. With these choices it is possible to select for every $t \in A^{<\ell}$ an element $s_t \in A^{q-\ell}$ such that $s_t \neq s_{t'}$ for every $t, t' \in A^{<\ell}$ with

[2]This method was invented by Ajtai and Szemerédi [**ASz**]. It was used, for instance, in Section 8.3 in the proof of the density Hales–Jewett theorem.

$t \neq t'$. Next, for every $i \in \{0, \dots, \ell - 1\}$, every $t \in A^i$ and every $a \in A$ we set $z_t^a = t^\frown (a^{\ell-i})^\frown s_t$ where $a^{\ell-i}$ is as in (2.1). Moreover, set

$$Z = \{z_t^a : t \in A^{<\ell} \text{ and } a \in A\}.$$

and notice that $Z \subseteq A^q$ and $\mathrm{dens}(Z) \leqslant \varepsilon$. Also observe that if $t \neq t'$, then we have $\{z_t^a : a \in A\} \cap \{z_{t'}^a : a \in A\} = \emptyset$. Finally, set

$$S = A^{<\ell} \cup \bigcup_{i=q}^{q+m-1} A^i \quad \text{and} \quad D = A^{<\ell} \cup \bigcup_{y \in A^q \setminus Z} \{y^\frown s : s \in A^{<m}\}.$$

It is clear that S is a highly structured subset of $A^{<\mathbb{N}}$ (it is the union of certain levels of $A^{<\mathbb{N}}$) and D is a subset of S of relative density at least $1 - \varepsilon$. Now for every $t \in A^{<\ell}$ let

$$V_t = \{t\} \cup \{z_t^{a\frown} s : a \in A \text{ and } s \in A^{<m}\}$$

and observe that $\mathcal{V} \coloneqq \{V_t : t \in A^{<\ell}\}$ is a family of pairwise disjoint m-dimensional Carlson–Simpson spaces which are all contained in S. Also notice that, regardless of how large ℓ is, $\mathcal{V}$ is maximal, that is, the set $S \setminus \cup\mathcal{V}$ contains no Carlson–Simpson space of dimension m. Nevertheless, $V_t \cap D$ is the singleton $\{t\}$ for every $t \in A^{<\ell}$.

The above example shows that, in the context of the density Carlson–Simpson theorem, one cannot achieve the density increment merely by producing an almost tiling of the "simple" set S. (In particular, a greedy algorithm will be inefficient.) To overcome this obstacle, an exhaustion procedure is used which can be roughly described as follows. At each step of the process, we are given a subset S' of S and we produce a collection $\mathcal{U}$ of Carlson–Simpson spaces of sufficiently large dimension which are all contained in S'. These Carlson–Simpson spaces are not pairwise disjoint since we are not aiming at producing a tiling. Instead, we are mainly interested in whether a sufficient portion of them behaves as expected, in the sense that for "many" $U \in \mathcal{U}$ the density of the set D in U is close enough to the relative density of D in S. If this is the case, then we can easily achieve the density increment. Otherwise, using coloring arguments, we show that for "almost every" Carlson–Simpson space $U \in \mathcal{U}$ the restriction of D on U is quite "thin". We then remove from S' an appropriately chosen subset of $\cup\mathcal{U}$ and we repeat this process for the resulting set. Finally, it is shown that this algorithm will indeed terminate, thus completing the proof of this step.

We also note that in order to execute the steps described above, we will represent any finite subset of $A^{<\mathbb{N}}$ as a family of measurable events indexed by an appropriately chosen finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$. The philosophy is identical to that in Section 9.4, though the details are somewhat different since we need to work with iterated convolutions. In particular, the reader is advised to review the material in Section 9.2.

9.5.2. The main result. Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every positive integer l and every $0 < \beta \leqslant 1$ the number $\mathrm{DCS}(k, l, \beta)$ has been estimated.

This assumption permits us to introduce some numerical invariants. Specifically, for every positive integer m and every $0<\gamma\leqslant 1$ we set

$$\bar{m}=\bar{m}(m,\gamma)=\lceil 256\gamma^{-3}m\rceil \tag{9.107}$$

and

$$M=\Lambda(k,\bar{m},\gamma^2/32)\overset{(9.48)}{=}\lceil 32\gamma^{-2}\mathrm{DCS}(k,\bar{m},\gamma^2/32)\rceil. \tag{9.108}$$

Also let

$$\alpha=\alpha(k,m,\gamma)=\Theta(k,\bar{m},\gamma^2/32)\ \text{ and }\ p_0=p_0(k,m,\gamma)=\lfloor\alpha^{-1}\rfloor \tag{9.109}$$

where $\Theta(k,\bar{m},\gamma^2/32)$ is as in (9.49). Finally, we define three sequences (n_p), (ν_p) and (N_p) in $\mathbb{N}$—also depending on the parameters k,m and γ—recursively by the rule $n_0=\nu_0=N_0=0$ and

$$\begin{cases} n_{p+1}=\bar{m}+(\bar{m}+1)N_p,\\ \nu_{p+1}=\mathrm{CS}(k+1,n_{p+1},\bar{m},\bar{m}+1),\\ N_{p+1}=\mathrm{CS}(k+1,\max\{\nu_{p+1},M\},\bar{m},2).\end{cases} \tag{9.110}$$

The following theorem is the main result of this section.

Theorem 9.38. *Let $k\in\mathbb{N}$ with $k\geqslant 2$ and assume that for every positive integer l and every $0<\beta\leqslant 1$ the number $\mathrm{DCS}(k,l,\beta)$ has been estimated.*

Let $0<\gamma,\delta\leqslant 1$, A an alphabet with $|A|=k+1$ and $a,b\in A$ with $a\neq b$. Also let V be a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ and I a nonempty subset of $\{0,\dots,\dim(V)\}$. Assume that we are given three subsets S,T and D of $A^{<\mathbb{N}}$ with the following properties.

(a) *The set S is (a,b)-insensitive in V.*
(b) *For every $i\in I$ we have $\mathrm{dens}_{V(i)}(S\cap T\cap D)\geqslant(\delta+2\gamma)\,\mathrm{dens}_{V(i)}(S\cap T)$ and $\mathrm{dens}_{V(i)}(S\cap T)\geqslant 2\gamma$.*

Finally, let m be a positive integer and suppose that

$$|I|\geqslant \mathrm{RegCS}(k+1,2,N_{p_0}+1,\gamma^2/2)$$

where p_0 and N_{p_0} are defined in (9.109) *and* (9.110) *respectively for the parameters k,m and γ. Then there exist a finite-dimensional Carlson–Simpson subspace W of V and a subset I' of $\{0,\dots,\dim(W)\}$ of cardinality m such that*

$$\mathrm{dens}_{W(i)}(T\cap D)\geqslant(\delta+\gamma/2)\,\mathrm{dens}_{W(i)}(T)$$

and

$$\mathrm{dens}_{W(i)}(T)\geqslant\frac{\gamma^3}{256}$$

for every $i\in I'$.

The proof of Theorem 9.38 occupies the bulk of the present section and is given in Subsections 9.5.3 and 9.5.4. Finally, in Subsection 9.5.5 we use Theorem 9.38 to complete the proof of the second step of the inductive scheme described in (9.75).

9.5.3. Preliminary lemmas. We are about to present some results which are part of the proof of Theorem 9.38 but are independent of the rest of the argument. We start with the following variant of Lemma 9.11.

LEMMA 9.39. *Let $0 < \gamma, \delta \leqslant 1$. Also let A be a finite alphabet with $|A| \geqslant 2$, V a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$, $N \in \mathbb{N}$ and $I \subseteq \{0, \dots, \dim(V)\}$ such that*

$$|I| \geqslant \mathrm{RegCS}(|A|, 2, N+1, \gamma^2/2). \tag{9.111}$$

Finally, let E, D be subsets of $A^{<\mathbb{N}}$ such that for every $i \in I$ we have

$$\mathrm{dens}_{V(i)}(E \cap D) \geqslant (\delta + 2\gamma)\, \mathrm{dens}_{V(i)}(E) \ \ \textit{and} \ \ \mathrm{dens}_{V(i)}(E) \geqslant 2\gamma. \tag{9.112}$$

Then there exists $L \subseteq I$ with $|L| = N+1$ and satisfying the following property. Let $\mathrm{c}_{L,V}$ be the convolution operation associated with (L, V), and set $\mathcal{E} = \mathrm{c}_{L,V}^{-1}(E)$ and $\mathcal{D} = \mathrm{c}_{L,V}^{-1}(D)$. Then for every $t \in A^{<|L|}$ we have

$$\mathrm{dens}_{\Omega_L}(\mathcal{E}_t \cap \mathcal{D}_t) \geqslant (\delta + \gamma)\, \mathrm{dens}_{\Omega_L}(\mathcal{E}_t) \ \ \textit{and} \ \ \mathrm{dens}_{\Omega_L}(\mathcal{E}_t) \geqslant \gamma \tag{9.113}$$

where $\mathcal{E}_t$ and $\mathcal{D}_t$ are the sections at t of $\mathcal{E}$ and $\mathcal{D}$ respectively.

PROOF. We set $\mathcal{F} = \big\{ \mathrm{I}_V^{-1}(E), \mathrm{I}_V^{-1}(E \cap D) \big\}$ where I_V is the canonical isomorphism associated with V (see Definition 1.10). By Lemma 6.24 and (9.111), there exists a subset $L = \{l_0 < \dots < l_{|L|-1}\}$ of I with $|L| = N+1$ such that the family $\mathcal{F}$ is $(\gamma^2/2, L)$-regular. We will show that the set L is as desired. Indeed, by (9.4), we have $\mathcal{E} \cap \mathcal{D} = \mathrm{c}_L^{-1}\big(\mathrm{I}_V^{-1}(E \cap D)\big)$ and $\mathcal{E} = \mathrm{c}_L^{-1}\big(\mathrm{I}_V^{-1}(E)\big)$. Fix $i \in \{0, \dots, |L|-1\}$ and let $t \in A^i$ be arbitrary. By Lemma 9.11, we see that

(a) $|\mathrm{dens}_{\Omega_L}(\mathcal{E}_t \cap \mathcal{D}_t) - \mathrm{dens}_{A^{l_i}}\big(\mathrm{I}_V^{-1}(E \cap D)\big)| \leqslant \gamma^2/2$ and
(b) $|\mathrm{dens}_{\Omega_L}(\mathcal{E}_t) - \mathrm{dens}_{A^{l_i}}\big(\mathrm{I}_V^{-1}(E)\big)| \leqslant \gamma^2/2$.

Since $\mathrm{dens}_{A^{l_i}}\big(\mathrm{I}_V^{-1}(X)\big) = \mathrm{dens}_{V(l_i)}(X)$ for every $X \subseteq V(l_i)$, we obtain that

$$|\mathrm{dens}_{\Omega_L}(\mathcal{E}_t \cap \mathcal{D}_t) - \mathrm{dens}_{V(l_i)}(E \cap D)| \leqslant \gamma^2/2 \tag{9.114}$$

and

$$|\mathrm{dens}_{\Omega_L}(\mathcal{E}_t) - \mathrm{dens}_{V(l_i)}(E)| \leqslant \gamma^2/2. \tag{9.115}$$

Combining (9.112), (9.114) and (9.115), we conclude that the two estimates in (9.113) are satisfied. The proof of Lemma 9.39 is completed. □

The next lemma asserts that certain metric properties are preserved by iterated convolutions operations.

LEMMA 9.40. *Let A be a finite alphabet with $|A| \geqslant 2$, d a positive integer and $(\mathbf{L}, \mathbf{V}) = \big((L_n)_{n=0}^d, (V_n)_{n=0}^d\big)$ an A-compatible pair. Also let $E, D \subseteq A^{<\mathbb{N}}$. For every $p \in \{0, \dots, d\}$ set $(\mathbf{L}_p, \mathbf{V}_p) = \big((L_n)_{n=0}^p, (V_n)_{n=0}^p\big)$, $\mathcal{E}^p = \mathrm{c}_{\mathbf{L}_p, \mathbf{V}_p}^{-1}(E)$ and $\mathcal{D}^p = \mathrm{c}_{\mathbf{L}_p, \mathbf{V}_p}^{-1}(D)$, and for every $t \in A^{<|L_p|}$ let $\mathcal{E}_t^p$ and $\mathcal{D}_t^p$ be the sections at t of $\mathcal{E}^p$ and $\mathcal{D}^p$ respectively. Finally, let $\lambda > 0$ and $0 < \gamma \leqslant 1$. Then the following hold.*

(a) *If $\mathrm{dens}_{\Omega_{L_0}}(\mathcal{E}_s^0 \cap \mathcal{D}_s^0) \geqslant \lambda \cdot \mathrm{dens}_{\Omega_{L_0}}(\mathcal{E}_s^0)$ for every $s \in A^{<|L_0|}$, then for every $p \in [d]$ and every $t \in A^{<|L_p|}$ we have $\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(\mathcal{E}_t^p \cap \mathcal{D}_t^p) \geqslant \lambda \cdot \mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(\mathcal{E}_t^p)$.*
(b) *If $\mathrm{dens}_{\Omega_{L_0}}(\mathcal{E}_s^0) \geqslant \gamma$ for every $s \in A^{<|L_0|}$, then for every $p \in [d]$ and every $t \in A^{<|L_p|}$ we have $\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(\mathcal{E}_t^p) \geqslant \gamma$.*

PROOF. Let $p \in [d]$ and $t \in A^{<|L_p|}$ be arbitrary. By Corollary 9.18, we have

$$\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(\mathcal{E}_t^p \cap \mathcal{D}_t^p) = \mathbb{E}_{s\in\mathcal{C}_t}\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_{p-1}}}(\mathcal{E}_s^{p-1} \cap \mathcal{D}_s^{p-1}) \tag{9.116}$$

and

$$\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(\mathcal{E}_t^p) = \mathbb{E}_{s\in\mathcal{C}_t}\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_{p-1}}}(\mathcal{E}_s^{p-1}) \tag{9.117}$$

where, as in (9.40), $\mathcal{C}_t = \{\mathrm{c}_{L_p,V_p}(t,\omega) : \omega \in \Omega_{L_p}\}$. Therefore, the result follows by induction on p and using (9.116) and (9.117). □

We proceed with the following variant of Lemma 9.17.

LEMMA 9.41. *Let A, d and $(\mathbf{L},\mathbf{V})$ be as in Lemma 9.40. Let $p \in \{0,\dots,d-1\}$ and set $(\mathbf{L}_p,\mathbf{V}_p) = \big((L_n)_{n=0}^p,(V_n)_{n=0}^p\big)$ and $(\mathbf{L}_{p+1},\mathbf{V}_{p+1}) = \big((L_n)_{n=0}^{p+1},(V_n)_{n=0}^{p+1}\big)$. Denote by q_{p+1} the quotient map associated with $(\mathbf{L}_{p+1},\mathbf{V}_{p+1})$ defined in (9.33). Let $\mathcal{X}$ be a subset of $A^{<|L_p|}\times\mathbf{\Omega}_{\mathbf{L}_p}$. Also let $t \in A^{<|L_{p+1}|}$ and $\mathcal{Y}$ a nonempty subset of $\mathcal{C}_t \times \mathbf{\Omega}_{\mathbf{L}_p}$ where $\mathcal{C}_t = \{\mathrm{c}_{L_{p+1},V_{p+1}}(t,\omega) : \omega \in \Omega_{L_{p+1}}\}$. We set $\mathscr{X} = \mathrm{q}_{p+1}^{-1}(\mathcal{X})$ and $\mathscr{Y} = \mathrm{q}_{p+1}^{-1}(\mathcal{Y})$. Then $\mathscr{Y}$ is a subset of $\{t\}\times\mathbf{\Omega}_{\mathbf{L}_{p+1}}$ and*

$$\mathrm{dens}_{\mathscr{Y}}(\mathscr{X}) = \mathrm{dens}_{\mathcal{Y}}(\mathcal{X}). \tag{9.118}$$

In particular, for every $X \subseteq V_0$ we have

$$\mathrm{dens}_{\mathscr{Y}}\big(\mathrm{c}^{-1}_{\mathbf{L}_{p+1},\mathbf{V}_{p+1}}(X)\big) = \mathrm{dens}_{\mathcal{Y}}\big(\mathrm{c}^{-1}_{\mathbf{L}_p,\mathbf{V}_p}(X)\big). \tag{9.119}$$

PROOF. Recall that, by Lemma 9.17, we have $\mathrm{q}_{p+1}^{-1}(\mathcal{C}_t\times\mathbf{\Omega}_{\mathbf{L}_p}) = \{t\}\times\mathbf{\Omega}_{\mathbf{L}_{p+1}}$. This implies that $\mathscr{Y} \subseteq \{t\}\times\mathbf{\Omega}_{\mathbf{L}_{p+1}}$. Moreover, $\mathscr{X}\cap\mathscr{Y} = \mathrm{q}_{p+1}^{-1}(\mathcal{X}\cap\mathcal{Y})$ and so

$$\begin{aligned}\mathrm{dens}_{\mathscr{Y}}(\mathscr{X}) = \frac{|\mathscr{X}\cap\mathscr{Y}|}{|\mathscr{Y}|} \quad &= \quad \frac{\mathrm{dens}_{\{t\}\times\mathbf{\Omega}_{\mathbf{L}_{p+1}}}(\mathscr{X}\cap\mathscr{Y})}{\mathrm{dens}_{\{t\}\times\mathbf{\Omega}_{\mathbf{L}_{p+1}}}(\mathscr{Y})}\\ &\overset{(9.41)}{=} \quad \frac{\mathrm{dens}_{\mathcal{C}_t\times\mathbf{\Omega}_{\mathbf{L}_p}}(\mathcal{X}\cap\mathcal{Y})}{\mathrm{dens}_{\mathcal{C}_t\times\mathbf{\Omega}_{\mathbf{L}_p}}(\mathcal{Y})}\\ &= \quad \frac{|\mathcal{X}\cap\mathcal{Y}|}{|\mathcal{Y}|} = \mathrm{dens}_{\mathcal{Y}}(\mathcal{X}).\end{aligned}$$

Finally, by Fact 9.14, we have $\mathrm{c}^{-1}_{\mathbf{L}_{p+1},\mathbf{V}_{p+1}}(X) = \mathrm{q}_{p+1}^{-1}\big(\mathrm{c}^{-1}_{\mathbf{L}_p,\mathbf{V}_p}(X)\big)$ for every $X \subseteq V_0$, and so (9.119) follows from (9.118). The proof of Lemma 9.41 is completed. □

The last result of this subsection shows that iterated convolutions operations are compatible with the notion of (a,b)-equivalence introduced in Subsection 2.1.1.

LEMMA 9.42. *Let A, d and $(\mathbf{L},\mathbf{V})$ be as in Lemma 9.40. Let $p \in \{0,\dots,d\}$ and set $(\mathbf{L}_p,\mathbf{V}_p) = \big((L_n)_{n=0}^p,(V_n)_{n=0}^p\big)$. Also let $t,t' \in A^{<|L_p|}$ and $a,b \in A$ with $a \neq b$. Then the following hold.*

(a) *Let $\boldsymbol{\omega} \in \mathbf{\Omega}_{\mathbf{L}_p}$ and set $s = \mathrm{c}_{\mathbf{L}_p,\mathbf{V}_p}(t,\boldsymbol{\omega})$ and $s' = \mathrm{c}_{\mathbf{L}_p,\mathbf{V}_p}(t',\boldsymbol{\omega})$. Then t and t' are (a,b)-equivalent if and only if s and s' are (a,b)-equivalent.*
(b) *Let S be a subset of $A^{<\mathbb{N}}$. Set $\mathcal{S}^p = \mathrm{c}^{-1}_{\mathbf{L}_p,\mathbf{V}_p}(S)$ and let $\mathcal{S}_t^p$ and $\mathcal{S}_{t'}^p$ be the sections of $\mathcal{S}^p$ at t and t' respectively. If S is (a,b)-insensitive in V_0 and t,t' are (a,b)-equivalent, then $\mathcal{S}_t^p$ and $\mathcal{S}_{t'}^p$ coincide.*

Proof. (a) By induction on p. The initial case "$p = 0$" follows immediately by Definition 9.3. Let $p \in \{0, \dots, d-1\}$ and assume that the result has been proved up to p. Fix $\boldsymbol{\omega} \in \boldsymbol{\Omega}_{\mathbf{L}_{p+1}}$ and write $\boldsymbol{\omega}$ as $(\boldsymbol{\omega}_0, \omega)$ where $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}_{\mathbf{L}_p}$ and $\omega \in \Omega_{L_{p+1}}$. Set $y = \mathrm{c}_{L_{p+1},V_p}(t,\omega)$ and $y' = \mathrm{c}_{L_{p+1},V_p}(t',\omega)$ and notice that

$$t, t' \text{ are } (a,b)\text{-equivalent} \Leftrightarrow y, y' \text{ are } (a,b)\text{-equivalent}. \tag{9.120}$$

On the other hand, by (9.31), we have $s = \mathrm{c}_{\mathbf{L}_p,\mathbf{V}_p}(y, \boldsymbol{\omega}_0)$ and $s' = \mathrm{c}_{\mathbf{L}_p,\mathbf{V}_p}(y', \boldsymbol{\omega}_0)$. Thus, invoking our inductive assumptions, we obtain that

$$y, y' \text{ are } (a,b)\text{-equivalent} \Leftrightarrow s, s' \text{ are } (a,b)\text{-equivalent}. \tag{9.121}$$

By (9.120) and (9.121), the proof of the first part of the lemma is completed.

(b) Let $\boldsymbol{\omega} \in \boldsymbol{\Omega}_{\mathbf{L}_p}$ be arbitrary. By part (a) and using the fact that S is (a,b)-insensitive in V_0, we see that $\mathrm{c}_{\mathbf{L}_p,\mathbf{V}_p}(t,\boldsymbol{\omega}) \in S$ if and only if $\mathrm{c}_{\mathbf{L}_p,\mathbf{V}_p}(t',\boldsymbol{\omega}) \in S$, which is equivalent to saying that $\boldsymbol{\omega} \in \mathcal{S}^p_t$ if and only if $\boldsymbol{\omega} \in \mathcal{S}^p_{t'}$. It follows that $\mathcal{S}^p_t = \mathcal{S}^p_{t'}$ and the proof of Lemma 9.42 is completed. □

9.5.4. Proof of Theorem 9.38. First we apply Lemma 9.39 and we obtain $L \subseteq I$ with $|L| = N_{p_0} + 1$ such that, setting $\mathcal{S} = \mathrm{c}^{-1}_{L,V}(S)$, $\mathcal{T} = \mathrm{c}^{-1}_{L,V}(T)$ and $\mathcal{D} = \mathrm{c}^{-1}_{L,V}(D)$, for every $s \in A^{<|L|}$ we have

$$\mathrm{dens}_{\Omega_L}(\mathcal{S}_s \cap \mathcal{T}_s \cap \mathcal{D}_s) \geqslant (\delta + \gamma)\, \mathrm{dens}_{\Omega_L}(\mathcal{S}_s \cap \mathcal{T}_s) \tag{9.122}$$

and

$$\mathrm{dens}_{\Omega_L}(\mathcal{S}_s \cap \mathcal{T}_s) \geqslant \gamma \tag{9.123}$$

where $\mathcal{S}_s$, $\mathcal{T}_s$ and $\mathcal{D}_s$ are the sections at s of $\mathcal{S}$, $\mathcal{T}$ and $\mathcal{D}$ respectively.

We now argue by contradiction. Specifically, if Theorem 9.38 is not satisfied, then we will determine an integer $d \in [p_0]$ and we will select

(a) an A-compatible pair $\big((L_n)^d_{n=0}, (V_n)^d_{n=0}\big)$ with $L_0 = L$ and $V_0 = V$, and
(b) for every $p \in [d]$, every $\ell \in [p]$ and every $s \in A^{<|L_p|}$ a family $\mathcal{Q}^{\ell,p}_s$ of subsets of $\boldsymbol{\Omega}_{\mathbf{L}_p}$ where $\mathbf{L}_p = (L_n)^p_{n=0}$ and $\mathbf{V}_p = (V_n)^p_{n=0}$.

The selection is done recursively so that, setting

$$\mathcal{S}^p = \mathrm{c}^{-1}_{\mathbf{L}_p,\mathbf{V}_p}(S), \quad \mathcal{T}^p = \mathrm{c}^{-1}_{\mathbf{L}_p,\mathbf{V}_p}(T) \quad \text{and} \quad \mathcal{D}^p = \mathrm{c}^{-1}_{\mathbf{L}_p,\mathbf{V}_p}(D)$$

for every $p \in [d]$, the following conditions are satisfied.

(C1) The set L_p has cardinality $N_{p_0-p} + 1$.
(C2) For every $s \in A^{<|L_p|}$ and every $\ell \in [p]$ the family $\mathcal{Q}^{\ell,p}_s$ consists of pairwise disjoint subsets of the section $\mathcal{S}^p_s$ of $\mathcal{S}^p$ at s.
(C3) If $p \geqslant 2$, then for every $s \in A^{<|L_p|}$ the sets $\cup\mathcal{Q}^{1,p}_s, \dots, \cup\mathcal{Q}^{p,p}_s$ are pairwise disjoint.
(C4) For every $s, s' \in A^{<|L_p|}$ with $|s| = |s'|$, every $\ell \in [p]$ and every $Q \in \mathcal{Q}^{\ell,p}_s$ and $Q' \in \mathcal{Q}^{\ell,p}_{s'}$ we have $\mathrm{dens}_{\boldsymbol{\Omega}_{\mathbf{L}_p}}(Q) = \mathrm{dens}_{\boldsymbol{\Omega}_{\mathbf{L}_p}}(Q')$.
(C5) For every $s \in A^{<|L_p|}$ and every $\ell \in [p]$ we say that an element Q of $\mathcal{Q}^{\ell,p}_s$ is *good* provided that

$$\mathrm{dens}_Q(\mathcal{S}^p_s \cap \mathcal{T}^p_s \cap \mathcal{D}^p_s) \geqslant (\delta + \gamma/2)\, \mathrm{dens}_Q(\mathcal{S}^p_s \cap \mathcal{T}^p_s)$$

and
$$\mathrm{dens}_Q(\mathcal{S}^p_s \cap \mathcal{T}^p_s) \geqslant \gamma^3/256.$$
Then, setting
$$\mathcal{G}^{\ell,p}_s = \{Q \in \mathcal{Q}^{\ell,p}_s : Q \text{ is good}\},$$
we have
$$\frac{|\mathcal{G}^{\ell,p}_s|}{|\mathcal{Q}^{\ell,p}_s|} < \frac{\gamma^3}{256}. \tag{9.124}$$

(C6) For every $s \in A^{<|L_p|}$ and every $\ell \in [p]$ we have $\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(\cup\mathcal{Q}^{\ell,p}_s) \geqslant \alpha$ where α is as in (9.109).

(C7) For every $s, s' \in A^{<|L_p|}$ if s and s' are (a,b)-equivalent, then for every $\ell \in [p]$ we have $\mathcal{Q}^{\ell,p}_s = \mathcal{Q}^{\ell,p}_{s'}$.

(C8) If $p = d$, then there exists $t_0 \in A^{<|L_d|}$ such that
$$\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_d}}\Big(\mathcal{S}^d_{t_0} \setminus \bigcup_{\ell=1}^{d} \cup\mathcal{Q}^{\ell,d}_{t_0}\Big) < \gamma^2/8. \tag{9.125}$$

Assuming that the above selection has been carried out, the proof of the theorem is completed as follows. Let t_0 be as in condition (C8). Since $L_0 = L$ and $V_0 = V$, by (9.122), (9.123) and Lemma 9.40, we see that
$$\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_d}}(\mathcal{S}^d_{t_0} \cap \mathcal{T}^d_{t_0} \cap \mathcal{D}^d_{t_0}) \geqslant (\delta+\gamma)\,\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_d}}(\mathcal{S}^d_{t_0} \cap \mathcal{T}^d_{t_0}) \tag{9.126}$$
and
$$\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_d}}(\mathcal{S}^d_{t_0} \cap \mathcal{T}^d_{t_0}) \geqslant \gamma. \tag{9.127}$$
For every $\ell \in [d]$ we set $\boldsymbol{Q}_\ell = \cup\mathcal{Q}^{\ell,d}_{t_0}$. By (9.125), the family $\{\boldsymbol{Q}_\ell : \ell \in [d]\}$ is an "almost cover" of $\mathcal{S}^d_{t_0}$, and a fortiori an "almost cover" of $\mathcal{S}^d_{t_0} \cap \mathcal{T}^d_{t_0}$. Therefore, by (9.126), (9.127) and applying Lemma E.6 for "$\lambda = \delta + \gamma$", "$\beta = \gamma$" and "$\varepsilon = \gamma/4$", there exists $\ell_0 \in [d]$ such that
$$\mathrm{dens}_{\boldsymbol{Q}_{\ell_0}}(\mathcal{S}^d_{t_0} \cap \mathcal{T}^d_{t_0} \cap \mathcal{D}^d_{t_0}) \geqslant (\delta+3\gamma/4)\,\mathrm{dens}_{\boldsymbol{Q}_{\ell_0}}(\mathcal{S}^d_{t_0} \cap \mathcal{T}^d_{t_0}) \tag{9.128}$$
and
$$\mathrm{dens}_{\boldsymbol{Q}_{\ell_0}}(\mathcal{S}^d_{t_0} \cap \mathcal{T}^d_{t_0}) \geqslant \gamma^2/16. \tag{9.129}$$
Next observe that, by conditions (C2) and (C4), the family $\mathcal{Q}^{\ell_0,d}_{t_0}$ is a partition of $\boldsymbol{Q}_{\ell_0}$ into sets of equal size. Taking into account this observation and the estimates in (9.128) and (9.129), by a second application of Lemma E.6 for "$\lambda = \delta + 3\gamma/4$", "$\beta = \gamma^2/16$" and "$\varepsilon = \gamma/4$", we conclude that
$$\frac{|\mathcal{G}^{\ell_0,d}_{t_0}|}{|\mathcal{Q}^{\ell_0,d}_{t_0}|} \geqslant \frac{\gamma^3}{256}.$$
This contradicts (9.124), as desired.

The rest of the proof is devoted to the description of the recursive selection. For "$p = 0$" we set $L_0 = L$ and $V_0 = V$. Let $p \in \{0, \dots, p_0\}$ and assume that the selection has been carried out up to p. We consider the following cases.

Case 1: *we have* $p = p_0$. Notice first that, by condition (C1), the set L_p is a singleton, and so $A^{<|L_p|}$ consists only of the empty word. We set $t_0 = \emptyset$ and $d = p_0$.

With these choices the recursive selection will be completed once we show that the estimate in (9.125) is satisfied. Indeed, by conditions (C3) and (C6), we have

$$\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_d}}\Big(\bigcup_{\ell=1}^{d} \cup\mathcal{Q}^{\ell,d}_{t_0}\Big) \geqslant d\cdot\alpha = p_0\cdot\alpha \overset{(9.109)}{>} 1-\gamma^2/8$$

which implies, of course, the estimate in (9.125).

Case 2: *we have* $1\leqslant p<p_0$ *and there exists* $t_0\in A^{<|L_p|}$ *such that*

$$\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}\Big(\mathcal{S}^p_{t_0}\setminus\bigcup_{\ell=1}^{p} \cup\mathcal{Q}^{\ell,p}_{t_0}\Big) < \gamma^2/8.$$

In this case we set $d=p$ and we terminate the construction.

Case 3: *either* $p=0$, *or* $1\leqslant p<p_0$ *and*

$$\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}\Big(\mathcal{S}^p_{s}\setminus\bigcup_{\ell=1}^{p} \cup\mathcal{Q}^{\ell,p}_{s}\Big) \geqslant \gamma^2/8 \tag{9.130}$$

for every $s\in A^{<|L_p|}$. If $p=0$, then for every $s\in A^{<|L_0|}$ we set

$$\Gamma_s=\mathcal{S}^0_s. \tag{9.131}$$

Otherwise, for every $s\in A^{<|L_p|}$ let

$$\Gamma_s=\mathcal{S}^p_s\setminus\Big(\bigcup_{\ell=1}^{p}\cup\mathcal{Q}^{\ell,p}_s\Big). \tag{9.132}$$

We have the following fact.

Fact 9.43. *For every* $s\in A^{<|L_p|}$ *we have* $\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(\Gamma_s)\geqslant\gamma^2/8$. *Moreover, if* $s,s'\in A^{<|L_p|}$ *are* (a,b)*-equivalent, then* $\Gamma_s=\Gamma_{s'}$.

Proof. Let $s\in A^{<|L_p|}$. If $p=0$, then we have

$$\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(\Gamma_s)=\mathrm{dens}_{\Omega_{L_0}}(\mathcal{S}^0_s)\geqslant \mathrm{dens}_{\Omega_{L_0}}(\mathcal{S}^0_s\cap\mathcal{T}^0_s)\overset{(9.123)}{\geqslant}\gamma>\gamma^2/8.$$

On the other hand, if $p\geqslant 1$, then the desired estimate follows from (9.130). Finally, let $s,s'\in A^{<|L_p|}$ be (a,b)-equivalent. By Lemma 9.42 (and condition (C7) if $p\geqslant1$), we obtain that $\mathcal{S}^p_s=\mathcal{S}^p_{s'}$. The proof of Fact 9.43 is completed. □

For every Carlson–Simpson subspace U of $A^{<|L_p|}$ we set

$$\Gamma_U=\bigcap_{s\in U}\Gamma_s. \tag{9.133}$$

Also let $\bar{m}$ and M be as in (9.107) and (9.108) respectively and notice that, by condition (C1), we have

$$|L_p|=N_{p_0-p}+1\overset{(9.110)}{=}\mathrm{CS}(k+1,\max\{\nu_{p_0-p},M\},\bar{m},2)+1. \tag{9.134}$$

Therefore, by Fact 9.43 and (9.134), we may apply Corollary 9.24 and we obtain a Carlson–Simpson subspace X of $A^{<|L_p|}$ with $\dim(X)=\nu_{p_0-p}$ such that

$$\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(\Gamma_U)\geqslant\Theta(k,\bar{m},\gamma^2/32)\overset{(9.109)}{=}\alpha \tag{9.135}$$

for every $\bar{m}$-dimensional Carlson–Simpson subspace U of X.

Next, for every $i \in \{0, \dots, \bar{m}\}$ and every Carlson–Simpson subspace U of X with $\dim(U) = \bar{m}$ let $G_{i,U}$ be the set of all $\boldsymbol{\omega} \in \Gamma_U$ satisfying

(P1) $\mathrm{dens}_{U(i)\times\{\boldsymbol{\omega}\}}(\mathcal{S}^p \cap \mathcal{T}^p \cap \mathcal{D}^p) \geqslant (\delta + \gamma/2)\, \mathrm{dens}_{U(i)\times\{\boldsymbol{\omega}\}}(\mathcal{S}^p \cap \mathcal{T}^p)$, and

(P2) $\mathrm{dens}_{U(i)\times\{\boldsymbol{\omega}\}}(\mathcal{S}^p \cap \mathcal{T}^p) \geqslant \gamma^3/256$.

Our assumption that Theorem 9.38 does not hold true, reduces to the following property of the sets $G_{i,U}$.

FACT 9.44. *For every $\bar{m}$-dimensional Carlson–Simpson subspace U of X there exists $i \in \{0, \dots, \bar{m}\}$ such that* $\mathrm{dens}_{\Gamma_U}(G_{i,U}) < \gamma^3/256$.

PROOF. Assume, towards a contradiction, that there exists a $\bar{m}$-dimensional Carlson–Simpson subspace U of X such that for every $i \in \{0, \dots, \bar{m}\}$ we have $\mathrm{dens}_{\Gamma_U}(G_{i,U}) \geqslant \gamma^3/256$. For every $\boldsymbol{\omega} \in \Gamma_U$ we set

$$I_{\boldsymbol{\omega}} = \big\{ i \in \{0, \dots, \bar{m}\} : \boldsymbol{\omega} \in G_{i,U} \big\}.$$

By Lemma E.4, there exists $\boldsymbol{\omega}_0 \in \Gamma_U$ such that $|I_{\boldsymbol{\omega}_0}| \geqslant (\bar{m}+1)\gamma^3/256$. Hence, by the choice of $\bar{m}$ in (9.107), we have $|I_{\boldsymbol{\omega}_0}| \geqslant m$. We define $U_{\boldsymbol{\omega}_0} = \{\mathrm{c}_{\mathbf{L}_p, \mathbf{V}_p}(s, \boldsymbol{\omega}_0) : s \in U\}$ and we observe that, by Lemma 9.15, the set $U_{\boldsymbol{\omega}_0}$ is a Carlson–Simpson subspace of $V_0 = V$ of dimension $\bar{m}$. By Lemma 9.16 applied to the sets $S \cap T \cap D$ and $S \cap T$, for every $i \in \{0, \dots, \bar{m}\}$ we have

$$\mathrm{dens}_{U_{\boldsymbol{\omega}_0}(i)}(S \cap T \cap D) = \mathrm{dens}_{U(i)\times\{\boldsymbol{\omega}_0\}}(\mathcal{S}^p \cap \mathcal{T}^p \cap \mathcal{D}^p)$$

and

$$\mathrm{dens}_{U_{\boldsymbol{\omega}_0}(i)}(S \cap T) = \mathrm{dens}_{U(i)\times\{\boldsymbol{\omega}_0\}}(\mathcal{S}^p \cap \mathcal{T}^p).$$

The above equalities and the fact that $\boldsymbol{\omega}_0 \in G_{i,U}$ for every $i \in I_{\boldsymbol{\omega}_0}$ yield that

$$\mathrm{dens}_{U_{\boldsymbol{\omega}_0}(i)}(S \cap T \cap D) \geqslant (\delta + \gamma/2)\, \mathrm{dens}_{U_{\boldsymbol{\omega}_0}(i)}(S \cap T)$$

and

$$\mathrm{dens}_{U_{\boldsymbol{\omega}_0}(i)}(S \cap T) \geqslant \frac{\gamma^3}{256}$$

for every $i \in I_{\boldsymbol{\omega}_0}$. Finally, observe that $U_{\boldsymbol{\omega}_0}$ is contained in S since $\boldsymbol{\omega}_0 \in \Gamma_U$. Therefore, the Carlson–Simpson space $U_{\boldsymbol{\omega}_0}$ and the set $I_{\boldsymbol{\omega}_0}$ satisfy the conclusion of Theorem 9.38, in contradiction with our assumption. The proof of Fact 9.44 is completed. $\square$

We are now in a position to define the new objects of the recursive selection.

Step 1: selection of V_{p+1} and L_{p+1}. First, we will use a coloring argument to control the integer i obtained by Fact 9.44. Specifically, by Theorem 4.21 and the fact that

$$\dim(X) = \nu_{p_0 - p} \overset{(9.110)}{=} \mathrm{CS}(k+1, n_{p_0-p}, \bar{m}, \bar{m}+1),$$

there exist $i_0 \in \{0, \dots, \bar{m}\}$ and a Carlson–Simpson subspace Y of X with

$$\dim(Y) = n_{p_0-p} \overset{(9.110)}{=} \bar{m} + (\bar{m}+1) N_{p_0-(p+1)} \tag{9.136}$$

and such that for every $\bar{m}$-dimensional Carlson–Simpson subspace U of Y we have

$$\mathrm{dens}_{\Gamma_U}(G_{i_0,U}) < \gamma^3/256.$$

We define

$$V_{p+1} = Y \quad \text{and} \quad L_{p+1} = \{i_0 + (i_0+1)i : 0 \leqslant i \leqslant N_{p_0-(p+1)}\}. \tag{9.137}$$

Notice that

$$|L_{p+1}| = N_{p_0-(p+1)} + 1 \tag{9.138}$$

and so with these choices condition (C1) is satisfied. Also observe that, by (9.136) and (9.137), we have $L_{p+1} \subseteq \{0, \dots, \dim(V_{p+1})\}$ and $V_{p+1} \subseteq X \subseteq A^{<|L_p|}$. Thus, the pair $(\mathbf{L}_{p+1}, \mathbf{V}_{p+1}) = \big((L_n)_{n=0}^{p+1}, (V_n)_{n=0}^{p+1}\big)$ is A-compatible. *In what follows, for notational simplicity, we shall denote by* q_{p+1} *the quotient map associated with the pair* $(\mathbf{L}_{p+1}, \mathbf{V}_{p+1})$ *defined in* (9.33).

Step 2: selection of the families $\mathcal{Q}_t^{p+1,p+1}$. This is the most important part of the recursive selection. The members of the families $\mathcal{Q}_t^{p+1,p+1}$ are, essentially, the sets Γ_U where U varies over all $\bar{m}$-dimensional Carlson–Simpson subspaces of V_{p+1}. However, in order to carry out the next steps of the selection, we have to group them in a canonical way.

We proceed to the details. Let $t \in A^{<|L_{p+1}|}$ and let $i \in \{0, \dots, |L_{p+1}|-1\}$ be the unique integer such that $t \in A^i$. As in (9.6), we set

$$\mathcal{C}_t = \big\{\mathrm{c}_{L_{p+1},V_{p+1}}(t,\omega) : \omega \in \Omega_{L_{p+1}}\big\}. \tag{9.139}$$

By Fact 9.4, we have $\mathcal{C}_t \subseteq V_{p+1}\big(i_0 + (i_0+1)i\big)$. If $i \geqslant 1$, then let

$$Z_t = \big\{\mathrm{c}_{L_{p+1}}(t \upharpoonright (i-1), \omega)^\frown t(i-1) : \omega \in \Omega_{L_{p+1}}\big\} \tag{9.140}$$

and observe that $Z_t \subseteq A^{(i_0+1)i}$. On the other hand, if $i = 0$ (that is, if t is the empty word), then we set $Z_\emptyset = A^0 = \{\emptyset\}$. Next, for every $z \in Z_t$ let

$$R^z = \{z^\frown x : x \in A^{<\bar{m}+1}\}. \tag{9.141}$$

Notice that the family $\{R^z : z \in Z_t\}$ consists of pairwise disjoint $\bar{m}$-dimensional Carlson–Simpson spaces and observe that for every $z \in Z_t$ we have

$$R^z \subseteq A^{<\bar{m}+(i_0+1)i+1} \overset{(9.136)}{\subseteq} A^{<\dim(V_{p+1})+1}.$$

We also set

$$\mathcal{U}_t = \big\{\mathrm{I}_{V_{p+1}}(R^z) : z \in Z_t\big\} \tag{9.142}$$

where $\mathrm{I}_{V_{p+1}}$ is the canonical isomorphism associated with V_{p+1} (see Definition 1.10).

Before we analyze the above definitions, let us give a specific example. Consider the alphabet $A = \{a, b, c, d\}$ and assume for simplicity that $V_{p+1} = A^{<n+1}$ where the integer n is large enough compared to i_0 (hence, the map $\mathrm{I}_{V_{p+1}}$ is the identity). Let $t = (a, b, a) \in A^3$ and observe that $t \upharpoonright 2 = (a, b)$. Also notice that $\mathcal{C}_t$ is the set of all $z \in A^{4i_0+3}$ such that $z(i_0) = t(0) = a$, $z(2i_0+1) = t(1) = b$ and $z(3i_0+2) = t(2) = a$. On the other hand, the set Z_t consists of all $z \in A^{3i_0+3}$ such that $z(i_0) = t(0) = a$, $z(2i_0+1) = t(1) = b$ and $z(3i_0+2) = t(2) = a$. It is easy to see that in this example the family $\{U(i_0) : U \in \mathcal{U}_t\}$ is a partition of the set $\mathcal{C}_t$. This is actually a general property as is shown in the following fact.

FACT 9.45. *Let $t \in A^{<|L_{p+1}|}$. Then we have*

$$\{c_{L_{p+1}}(t,\omega) : \omega \in \Omega_{L_{p+1}}\} = \{z^\frown x : z \in Z_t \text{ and } x \in A^{i_0}\}. \tag{9.143}$$

Moreover, the set $\mathcal{U}_t$ consists of pairwise disjoint $\bar{m}$-dimensional Carlson–Simpson subspaces of V_{p+1} and

$$\mathcal{C}_t = \bigcup_{U \in \mathcal{U}_t} U(i_0). \tag{9.144}$$

PROOF. By Definition 9.3, we see that (9.143) is satisfied. It is also clear that $\mathcal{U}_t$ consists of pairwise disjoint $\bar{m}$-dimensional Carlson–Simpson subspaces of V_{p+1}. Finally, notice that

$$\begin{aligned} \mathcal{C}_t &\overset{(9.139)}{=} \mathrm{I}_{V_{p+1}}\big(\{c_{L_{p+1}}(t,\omega) : \omega \in \Omega_{L_{p+1}}\}\big) \\ &\overset{(9.143)}{=} \bigcup_{z \in Z_t} \mathrm{I}_{V_{p+1}}\big(\{z^\frown x : x \in A^{i_0}\}\big) \\ &\overset{(9.141)}{=} \bigcup_{z \in Z_t} \mathrm{I}_{V_{p+1}}\big(R^z(i_0)\big) \overset{(9.142)}{=} \bigcup_{U \in \mathcal{U}_t} U(i_0) \end{aligned}$$

and the proof of Fact 9.45 is completed. □

We are now ready to define the families $\mathcal{Q}_t^{p+1,p+1}$. Specifically, let $t \in A^{<|L_{p+1}|}$ and for every $U \in \mathcal{U}_t$ and every $\boldsymbol{\omega} \in \Gamma_U$ we define

$$Q_t^{\boldsymbol{\omega},U} = \{\boldsymbol{\omega}\} \times \big\{\omega \in \Omega_{L_{p+1}} : c_{L_{p+1},V_{p+1}}(t,\omega) \in U(i_0)\big\} \subseteq \boldsymbol{\Omega}_{\mathbf{L}_{p+1}}. \tag{9.145}$$

We isolate, for future use, the following two representations of the sets $Q_t^{\boldsymbol{\omega},U}$.

(R1) Let $\mathcal{C}_t$ be as in (9.139). Moreover, as in (9.8), for every $s \in \mathcal{C}_t$ we set $\Omega_t^s = \{\omega \in \Omega_{L_{p+1}} : c_{L_{p+1},V_{p+1}}(t,\omega) = s\}$. Then observe that

$$Q_t^{\boldsymbol{\omega},U} = \{\boldsymbol{\omega}\} \times \bigcup_{s \in U(i_0)} \Omega_t^s. \tag{9.146}$$

(R2) By Fact 9.45, the set $U(i_0)$ is contained in $\mathcal{C}_t$. Therefore,

$$\{t\} \times Q_t^{\boldsymbol{\omega},U} = \mathrm{q}_{p+1}^{-1}\big(U(i_0) \times \{\boldsymbol{\omega}\}\big). \tag{9.147}$$

Finally, we define

$$\mathcal{Q}_t^{p+1,p+1} = \{Q_t^{\boldsymbol{\omega},U} : U \in \mathcal{U}_t \text{ and } \boldsymbol{\omega} \in \Gamma_U\}. \tag{9.148}$$

The second step of the recursive selection is completed.

Step 3: selection of the families $\mathcal{Q}_t^{\ell,p+1}$ for every $\ell \in [p]$. In this step we will not introduce something new, but only "copy" in the space $\boldsymbol{\Omega}_{\mathbf{L}_{p+1}}$ what we have constructed so far. In particular, this step is meaningful only if $p \geqslant 1$.

So let $p \geqslant 1$. Fix $t \in A^{<|L_{p+1}|}$ and let $\mathcal{C}_t$ be as in (9.139). For every $s \in \mathcal{C}_t$, every $\ell \in [p]$ and every $Q \in \mathcal{Q}_s^{\ell,p}$ we define

$$C_t^{s,\ell,Q} = Q \times \{\omega \in \Omega_{L_{p+1}} : c_{L_{p+1},V_{p+1}}(t,\omega) = s\} \subseteq \boldsymbol{\Omega}_{\mathbf{L}_{p+1}}. \tag{9.149}$$

As in the previous step, we have the following representations of these sets.

(R3) Since $\Omega_t^s = \{\omega \in \Omega_{L_{p+1}} : \mathrm{c}_{L_{p+1},V_{p+1}}(t,\omega) = s\}$, we see that

$$C_t^{s,\ell,Q} = Q \times \Omega_t^s. \tag{9.150}$$

(R4) We have

$$\{t\} \times C_t^{s,\ell,Q} = \mathrm{q}_{p+1}^{-1}\big(\{s\} \times Q\big). \tag{9.151}$$

Finally, let

$$\mathcal{Q}_t^{\ell,p+1} = \{C_t^{s,\ell,Q} : s \in \mathcal{C}_t \text{ and } Q \in \mathcal{Q}_s^{\ell,p}\}. \tag{9.152}$$

The recursive selection is completed.

Step 4: verification of the inductive assumptions. Recall that condition (C1) has already been checked in Step 1. Conditions (C2)–(C7) will be verified in the following series of claims.

Claim 9.46. *Let $t \in A^{<|L_{p+1}|}$. Then for every $\ell \in [p+1]$ the family $\mathcal{Q}_t^{\ell,p+1}$ consists of pairwise disjoint subsets of $\mathcal{S}_t^{p+1}$. That is, condition* (C2) *is satisfied.*

Proof. Assume that $p \geqslant 1$ and $\ell \in [p]$. First we will show that the family $\mathcal{Q}_t^{\ell,p+1}$ consists of pairwise disjoint sets. To this end, let $s, s' \in \mathcal{C}_t$, $Q \in \mathcal{Q}_s^{\ell,p}$ and $Q' \in \mathcal{Q}_{s'}^{\ell,p}$ such that $(s,Q) \neq (s',Q')$. If $s = s'$, then we have $Q \neq Q'$ which implies, by our inductive assumptions, that $Q \cap Q' = \emptyset$. By (9.150), we obtain that $C_t^{s,\ell,Q} \cap C_t^{s',\ell,Q'} = \emptyset$. Otherwise, if $s \neq s'$, then by Fact 9.5 we have $\Omega_t^s \cap \Omega_t^{s'} = \emptyset$. Invoking (9.150) once again, we conclude that $C_t^{s,\ell,Q} \cap C_t^{s',\ell,Q'} = \emptyset$. Next, let $C_t^{s,\ell,Q} \in \mathcal{Q}_t^{\ell,p+1}$ for some $s \in \Omega_t$ and $Q \in \mathcal{Q}_s^{\ell,p}$. By our inductive assumptions, we have $Q \subseteq \mathcal{S}_s^p$ or equivalently $\{s\} \times Q \subseteq \mathcal{S}^p$. Hence, by Fact 9.14,

$$\{t\} \times C_t^{s,\ell,Q} \overset{(9.151)}{=} \mathrm{q}_{p+1}\big(\{s\} \times Q\big) \subseteq \mathrm{q}_{p+1}^{-1}(\mathcal{S}^p) = \mathcal{S}^{p+1}.$$

We now consider the case "$\ell = p+1$". Let $U, U' \in \mathcal{U}_t$, $\boldsymbol{\omega} \in \Gamma_U$ and $\boldsymbol{\omega}' \in \Gamma_{U'}$, and assume that $(U,\boldsymbol{\omega}) \neq (U',\boldsymbol{\omega}')$. We need to show that the sets $Q_t^{\boldsymbol{\omega},U}$ and $Q_t^{\boldsymbol{\omega}',U'}$ are disjoint. Indeed, if $U = U'$, then we have $\boldsymbol{\omega} \neq \boldsymbol{\omega}'$ which implies, by (9.146), that $Q_t^{\boldsymbol{\omega},U} \cap Q_t^{\boldsymbol{\omega}',U'} = \emptyset$. Otherwise, by Fact 9.45, the Carlson–Simpson spaces U and U' are disjoint. This implies, in particular, that $U(i_0) \cap U'(i_0) = \emptyset$ and so, by (9.147), we see that $Q_t^{\boldsymbol{\omega},U} \cap Q_t^{\boldsymbol{\omega}',U'} = \emptyset$. Finally, let $U \in \mathcal{U}_t$ and $\boldsymbol{\omega} \in \Gamma_U$. By (9.132) and (9.133), we have $U(i_0) \times \{\boldsymbol{\omega}\} \subseteq \mathcal{S}^p$. Therefore, by Fact 9.14, we conclude that

$$\{t\} \times Q_t^{\boldsymbol{\omega},U} \overset{(9.147)}{=} \mathrm{q}_{p+1}\big(U(i_0) \times \{\boldsymbol{\omega}\}\big) \subseteq \mathrm{q}_{p+1}^{-1}(\mathcal{S}^p) = \mathcal{S}^{p+1}.$$

The proof of Claim 9.46 is completed. □

Claim 9.47. *Let $t \in A^{<|L_{p+1}|}$. If $p \geqslant 1$, then the sets $\cup\mathcal{Q}_t^{1,p+1}, \dots, \cup\mathcal{Q}_t^{p+1,p+1}$ are pairwise disjoint. That is, condition* (C3) *is satisfied.*

Proof. Let $\ell \in [p]$. Also let $\ell' \in [p+1]$ with $\ell' \neq \ell$. We need to show that the sets $\cup\mathcal{Q}_t^{\ell,p+1}$ and $\cup\mathcal{Q}_t^{\ell',p+1}$ are disjoint. If $\ell' \leqslant p$, then this follows immediately from (9.150) and our inductive assumptions. Next, assume that $\ell' = p+1$ and let $U \in \mathcal{U}_t$ and $\boldsymbol{\omega} \in \Gamma_U$ be arbitrary. By (9.132) and (9.133), we see that $\boldsymbol{\omega} \notin \cup\mathcal{Q}_s^{\ell,p}$ for every $s \in U$. Using this observation, the result follows from (9.147) and (9.151). □

CLAIM 9.48. *Let* $t,t'\in A^{<|L_{p+1}|}$ *with* $|t|=|t'|$. *Then for every* $\ell\in[p+1]$, *every* $Q\in\mathcal{Q}^{\ell,p+1}_t$ *and every* $Q'\in\mathcal{Q}^{\ell,p+1}_{t'}$ *we have* $\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_{p+1}}}(Q)=\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_{p+1}}}(Q')$. *That is, condition* (C4) *is satisfied.*

PROOF. Fix $\ell\in[p+1]$, $Q\in\mathcal{Q}^{\ell,p+1}_t$ and $Q'\in\mathcal{Q}^{\ell,p+1}_{t'}$. Assume that $p\geqslant 1$ and $\ell\in[p]$. By (9.150), there exist $s\in\mathcal{C}_t$, $s'\in\mathcal{C}_{t'}$, $Q_0\in\mathcal{Q}^{\ell,p}_s$ and $Q_0'\in\mathcal{Q}^{\ell,p}_{s'}$ such that

$$Q=Q_0\times\Omega^s_t \quad\text{and}\quad Q'=Q_0'\times\Omega^{s'}_{t'}.$$

Since $|t|=|t'|$, by Fact 9.4, we see that $|\mathcal{C}_t|=|\mathcal{C}_{t'}|$ and $|s|=|s'|$. By our inductive assumptions, this implies that

$$\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(Q_0)=\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(Q_0') \tag{9.153}$$

and moreover, by Fact 9.5,

$$\mathrm{dens}_{\Omega_{L_{p+1}}}(\Omega^s_t)=\frac{1}{|\mathcal{C}_t|}=\frac{1}{|\mathcal{C}_{t'}|}=\mathrm{dens}_{\Omega_{L_{p+1}}}(\Omega^{s'}_{t'}). \tag{9.154}$$

Therefore, combining (9.153) and (9.154), we obtain that

$$\begin{aligned}
\mathrm{dens}_{\mathbf{\Omega}_{L_{p+1}}}(Q) &= \mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_{p+1}}}(Q_0\times\Omega^s_t)\\
&= \mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(Q_0)\cdot\mathrm{dens}_{\Omega_{L_{p+1}}}(\Omega^s_t)\\
&= \mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}(Q_0')\cdot\mathrm{dens}_{\Omega_{L_{p+1}}}(\Omega^{s'}_{t'})=\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_{p+1}}}(Q').
\end{aligned}$$

Next we deal with the case "$\ell=p+1$". By (9.146), there exist $U\in\mathcal{U}_t$, $U'\in\mathcal{U}_{t'}$, $\boldsymbol{\omega}\in\Gamma_U$ and $\boldsymbol{\omega}'\in\Gamma_{U'}$ such that

$$Q=\{\boldsymbol{\omega}\}\times\bigcup_{s\in U(i_0)}\Omega^s_t \quad\text{and}\quad Q'=\{\boldsymbol{\omega}'\}\times\bigcup_{s'\in U'(i_0)}\Omega^{s'}_{t'}.$$

By Fact 9.45, we have $U(i_0)\subseteq\mathcal{C}_t$, $U'(i_0)\subseteq\mathcal{C}_{t'}$ and $|U(i_0)|=|U'(i_0)|=|A^{i_0}|$. On the other hand, by Fact 9.5, the families $\{\Omega^s_t: s\in\mathcal{C}_t\}$ and $\{\Omega^{s'}_{t'}: s'\in\mathcal{C}_{t'}\}$ consist of pairwise disjoint sets. Also recall that $|\mathcal{C}_t|=|\mathcal{C}_{t'}|$. Thus, by (9.154),

$$\mathrm{dens}_{\Omega_{L_{p+1}}}\Big(\bigcup_{s\in U(i_0)}\Omega^s_t\Big)=\frac{|U(i_0)|}{|\mathcal{C}_t|}=\frac{|U'(i_0)|}{|\mathcal{C}_{t'}|}=\mathrm{dens}_{\Omega_{L_{p+1}}}\Big(\bigcup_{s'\in U'(i_0)}\Omega^{s'}_{t'}\Big)$$

which implies that

$$\begin{aligned}
\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_{p+1}}}(Q) &= \mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_{p+1}}}\Big(\{\boldsymbol{\omega}\}\times\bigcup_{s\in U(i_0)}\Omega^s_t\Big)\\
&= \mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}\big(\{\boldsymbol{\omega}\}\big)\cdot\mathrm{dens}_{\Omega_{L_{p+1}}}\Big(\bigcup_{s\in U(i_0)}\Omega^s_t\Big)\\
&= \mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_p}}\big(\{\boldsymbol{\omega}'\}\big)\cdot\mathrm{dens}_{\Omega_{L_{p+1}}}\Big(\bigcup_{s'\in U'(i_0)}\Omega^{s'}_{t'}\Big)=\mathrm{dens}_{\mathbf{\Omega}_{\mathbf{L}_{p+1}}}(Q').
\end{aligned}$$

The proof of Claim 9.48 is completed. □

CLAIM 9.49. *Let* $t\in A^{<|L_{p+1}|}$. *Then for every* $\ell\in[p+1]$ *we have*

$$\frac{|\mathcal{G}^{\ell,p+1}_t|}{|\mathcal{Q}^{\ell,p+1}_t|}<\frac{\gamma^3}{256}.$$

That is, condition (C5) *is satisfied.*

PROOF. First assume that $p \geqslant 1$ and $\ell \in [p]$. For every $s \in \mathcal{C}_t$ let

$$\mathcal{Q}_s = \big\{C_t^{s,\ell,Q} : Q \in \mathcal{Q}_s^{\ell,p}\big\} \ \text{ and } \ \mathcal{G}_s = \mathcal{G}_t^{\ell,p+1} \cap \mathcal{Q}_s.$$

By (9.151), the family $\{\mathcal{Q}_s : s \in \mathcal{C}_t\}$ is a partition of $\mathcal{Q}_t^{\ell,p+1}$. Also notice that the family $\{\mathcal{G}_s : s \in \mathcal{C}_t\}$ is the induced partition of $\mathcal{G}_t^{\ell,p+1}$. Therefore,

$$|\mathcal{Q}_t^{\ell,p+1}| = \sum_{s\in\mathcal{C}_t} |\mathcal{Q}_s| \ \text{ and } \ |\mathcal{G}_t^{\ell,p+1}| = \sum_{s\in\mathcal{C}_t} |\mathcal{G}_s|. \tag{9.155}$$

We have the following representation of the family $\mathcal{G}_s$.

SUBCLAIM 9.50. *For every $s \in \mathcal{C}_t$ we have $\mathcal{G}_s = \{C_t^{s,\ell,Q} : Q \in \mathcal{G}_s^{\ell,p}\}$.*

PROOF OF SUBCLAIM 9.50. Fix $s \in \mathcal{C}_t$ and let $C \in \mathcal{Q}_s$ be arbitrary. By (9.150), we see that the map $\mathcal{Q}_s^{\ell,p} \ni Q \mapsto C_t^{s,\ell,Q} \in \mathcal{Q}_s$ is a bijection. Hence, there exists a unique $Q \in \mathcal{Q}_s^{\ell,p}$ such that $C = C_t^{\ell,s,Q}$. In particular, by (9.151), we have $\{t\} \times C = \mathrm{q}_{p+1}^{-1}\big(\{s\} \times Q\big)$ and so, by Fact 9.14 and Lemma 9.41, we obtain that

$$\begin{aligned}
\mathrm{dens}_C(\mathcal{S}_t^{p+1} \cap \mathcal{T}_t^{p+1} \cap \mathcal{D}_t^{p+1}) &= \mathrm{dens}_{\{t\}\times C}(\mathcal{S}^{p+1} \cap \mathcal{T}^{p+1} \cap \mathcal{D}^{p+1}) \\
&= \mathrm{dens}_{\mathrm{q}_{p+1}^{-1}(\{s\}\times Q)}\big(\mathrm{q}_{p+1}^{-1}(\mathcal{S}^p \cap \mathcal{T}^p \cap \mathcal{D}^p)\big) \\
&= \mathrm{dens}_{\{s\}\times Q}(\mathcal{S}^p \cap \mathcal{T}^p \cap \mathcal{D}^p) \\
&= \mathrm{dens}_Q(\mathcal{S}_s^p \cap \mathcal{T}_s^p \cap \mathcal{D}_s^p)
\end{aligned}$$

and, similarly,

$$\mathrm{dens}_C(\mathcal{S}_t^{p+1} \cap \mathcal{T}_t^{p+1}) = \mathrm{dens}_Q(\mathcal{S}_s^p \cap \mathcal{T}_s^p).$$

By the above equalities and the definition of a good set described in condition (C5), we conclude that $C \in \mathcal{G}_s$ if and only if $Q \in \mathcal{G}_s^{\ell,p}$ which is equivalent to saying that $\mathcal{G}_s = \{C_t^{s,\ell,Q} : Q \in \mathcal{G}_s^{\ell,p}\}$. The proof of Subclaim 9.50 is completed. □

We are in a position to complete the proof for the case "$\ell \in [p]$". As we have already mentioned, for every $s \in \mathcal{C}_t$ the map $\mathcal{Q}_s^{\ell,p} \ni Q \mapsto C_t^{s,\ell,Q} \in \mathcal{Q}_s$ is a bijection and so $|\mathcal{Q}_s| = |\mathcal{Q}_s^{\ell,p}|$. Moreover, by Subclaim 9.50, we have $|\mathcal{G}_s| = |\mathcal{G}_s^{\ell,p}|$ for every $s \in \mathcal{C}_t$. Hence, by (9.155), we see that

$$|\mathcal{Q}_t^{\ell,p+1}| = \sum_{s\in\mathcal{C}_t} |\mathcal{Q}_s^{\ell,p}| \ \text{ and } \ |\mathcal{G}_t^{\ell,p+1}| = \sum_{s\in\mathcal{C}_t} |\mathcal{G}_s^{\ell,p}|. \tag{9.156}$$

On the other hand, by our inductive assumptions, for every $s \in \mathcal{C}_t$ we have

$$|\mathcal{G}_s^{\ell,p}| < (\gamma^3/256) \cdot |\mathcal{Q}_s^{\ell,p}|. \tag{9.157}$$

Therefore,

$$\frac{|\mathcal{G}_t^{\ell,p+1}|}{|\mathcal{Q}_t^{\ell,p+1}|} \overset{(9.156)}{=} \frac{\sum_{s\in\mathcal{C}_t} |\mathcal{G}_s^{\ell,p}|}{\sum_{s\in\mathcal{C}_t} |\mathcal{Q}_s^{\ell,p}|} \overset{(9.157)}{<} \frac{\gamma^3}{256}.$$

Next we consider the case "$\ell = p+1$". The argument is similar. Specifically, for every $U \in \mathcal{U}_t$ let

$$\mathcal{Q}_U = \big\{Q_t^{\boldsymbol{\omega},U} : \boldsymbol{\omega} \in \Gamma_U\big\} \ \text{ and } \ \mathcal{G}_U = \mathcal{G}_t^{p+1,p+1} \cap \mathcal{Q}_U.$$

By Fact 9.45 and (9.147), the family $\{\mathcal{Q}_U : U \in \mathcal{U}_t\}$ is a partition of $\mathcal{Q}_t^{p+1,p+1}$ while the family $\{\mathcal{G}_U : U \in \mathcal{U}_t\}$ is the induced partition of $\mathcal{G}_t^{p+1,p+1}$. Hence,

$$|\mathcal{Q}_t^{p+1,p+1}| = \sum_{U \in \mathcal{U}_t} |\mathcal{Q}_U| \ \text{ and } \ |\mathcal{G}_t^{p+1,p+1}| = \sum_{U \in \mathcal{U}_t} |\mathcal{G}_U|. \tag{9.158}$$

Also recall that $G_{i_0,U}$ is the set of all $\boldsymbol{\omega} \in \Gamma_U$ satisfying properties (P1) and (P2). We have the following analogue of Subclaim 9.50.

Subclaim 9.51. *For every $U \in \mathcal{U}_t$ we have $\mathcal{G}_U = \{Q_t^{\boldsymbol{\omega},U} : \boldsymbol{\omega} \in G_{i_0,U}\}$.*

Proof of Subclaim 9.51. Fix $U \in \mathcal{U}_t$ and let $C \in \mathcal{Q}_U$ be arbitrary. By (9.146), the map $\Gamma_U \ni \boldsymbol{\omega} \mapsto Q_t^{\boldsymbol{\omega},U} \in \mathcal{Q}_U$ is a bijection and so there exists a unique $\boldsymbol{\omega} \in \Gamma_U$ such that $C = Q_t^{\boldsymbol{\omega},U}$. By (9.147), we see that $\{t\} \times C = \mathrm{q}_{p+1}^{-1}\big(U(i_0) \times \{\boldsymbol{\omega}\}\big)$. Moreover, by Fact 9.45, we have $U(i_0) \subseteq \mathcal{C}_t$. By Fact 9.14, Lemma 9.41 and arguing precisely as in the proof of Subclaim 9.50, we conclude that $C \in \mathcal{G}_U$ if and only if $\boldsymbol{\omega} \in G_{i_0,U}$. The proof of Subclaim 9.51 is completed. □

We are ready to estimate the size of $\mathcal{G}_t^{p+1,p+1}$. By Subclaim 9.51 and the fact that for every $U \in \mathcal{U}_t$ the map $\Gamma_U \ni \boldsymbol{\omega} \mapsto Q_t^{\boldsymbol{\omega},U} \in \mathcal{Q}_U$ is a bijection, we obtain that $|\mathcal{Q}_U| = |\Gamma_U|$ and $|\mathcal{G}_U| = |G_{i_0,U}|$ for every $U \in \mathcal{U}_t$. Hence, by (9.158),

$$|\mathcal{Q}_t^{p+1,p+1}| = \sum_{U \in \mathcal{U}_t} |\Gamma_U| \ \text{ and } \ |\mathcal{G}_t^{p+1,p+1}| = \sum_{U \in \mathcal{U}_t} |G_{i_0,U}|. \tag{9.159}$$

By the choice of i_0 in Step 1, we have $|G_{i_0,U}| < (\gamma^3/256) \cdot |\Gamma_U|$ for every $U \in \mathcal{U}_t$. Therefore,

$$\frac{|\mathcal{G}_t^{p+1,p+1}|}{|\mathcal{Q}_t^{p+1,p+1}|} \overset{(9.159)}{=} \frac{\sum_{U \in \mathcal{U}_t} |G_{i_0,U}|}{\sum_{U \in \mathcal{U}_t} |\Gamma_U|} < \frac{\gamma^3}{256}$$

and the proof of Claim 9.49 is completed. □

Claim 9.52. *Let $t \in A^{<|L_{p+1}|}$. Then for every $\ell \in [p+1]$ we have*

$$\mathrm{dens}_{\boldsymbol{\Omega}_{\mathbf{L}_{p+1}}}(\cup \mathcal{Q}_t^{\ell,p+1}) \geqslant \alpha.$$

That is, condition (C6) *is satisfied.*

Proof. First assume that $p \geqslant 1$ and $\ell \in [p]$. By our inductive assumptions,

$$\mathrm{dens}_{\mathcal{C}_t \times \boldsymbol{\Omega}_{\mathbf{L}_p}}\Big(\bigcup_{s \in \mathcal{C}_t} \{s\} \times \cup \mathcal{Q}_s^{\ell,p}\Big) \geqslant \alpha. \tag{9.160}$$

On the other hand,

$$\begin{aligned} \mathrm{q}_{p+1}^{-1}\Big(\bigcup_{s \in \mathcal{C}_t} \{s\} \times \cup \mathcal{Q}_s^{\ell,p}\Big) &= \mathrm{q}_{p+1}^{-1}\Big(\bigcup_{s \in \mathcal{C}_t} \bigcup_{Q \in \mathcal{Q}_s^{\ell,p}} \{s\} \times Q\Big) \\ &= \bigcup_{s \in \mathcal{C}_t} \bigcup_{Q \in \mathcal{Q}_s^{\ell,p}} \mathrm{q}_{p+1}^{-1}\big(\{s\} \times Q\big) \\ &\overset{(9.151)}{=} \bigcup_{s \in \mathcal{C}_t} \bigcup_{Q \in \mathcal{Q}_s^{\ell,p}} \{t\} \times C_t^{s,\ell,Q} \\ &\overset{(9.152)}{=} \{t\} \times \cup \mathcal{Q}_t^{\ell,p+1}. \end{aligned} \tag{9.161}$$

By (9.160), (9.161) and Lemma 9.17, we obtain that

$$\mathrm{dens}_{\boldsymbol{\Omega}_{\mathbf{L}_{p+1}}}(\cup\mathcal{Q}_t^{\ell,p+1}) = \mathrm{dens}_{\{t\}\times\boldsymbol{\Omega}_{\mathbf{L}_{p+1}}}\big(\{t\}\times\cup\mathcal{Q}_t^{\ell,p+1}\big) \geqslant \alpha.$$

Next assume that $\ell = p+1$. By Fact 9.45, the family $\{U(i_0) : U \in \mathcal{U}_t\}$ is a partition of $\mathcal{C}_t$ and, by (9.135), we have $\mathrm{dens}_{\boldsymbol{\Omega}_{\mathbf{L}_p}}(\Gamma_U) \geqslant \alpha$ for every $U \in \mathcal{U}_t$. Hence,

$$\mathrm{dens}_{\mathcal{C}_t\times\boldsymbol{\Omega}_{\mathbf{L}_p}}\Big(\bigcup_{U\in\mathcal{U}_t} U(i_0)\times\Gamma_U\Big) \geqslant \alpha. \tag{9.162}$$

Notice that

$$\begin{aligned}
\mathrm{q}_{p+1}^{-1}\Big(\bigcup_{U\in\mathcal{U}_t} U(i_0)\times\Gamma_U\Big) &= \mathrm{q}_{p+1}^{-1}\Big(\bigcup_{U\in\mathcal{U}_t}\bigcup_{\boldsymbol{\omega}\in\Gamma_U} U(i_0)\times\{\boldsymbol{\omega}\}\Big) \\
&= \bigcup_{U\in\mathcal{U}_t}\bigcup_{\boldsymbol{\omega}\in\Gamma_U} \mathrm{q}_{p+1}^{-1}\big(U(i_0)\times\{\boldsymbol{\omega}\}\big) \\
&\overset{(9.147)}{=} \bigcup_{U\in\mathcal{U}_t}\bigcup_{\boldsymbol{\omega}\in\Gamma_U} \{t\}\times Q_t^{\boldsymbol{\omega},U} \\
&\overset{(9.148)}{=} \{t\}\times\cup\mathcal{Q}_t^{p+1,p+1}. \qquad (9.163)
\end{aligned}$$

Combining (9.162) and (9.163) and applying Lemma 9.17, we conclude that

$$\mathrm{dens}_{\boldsymbol{\Omega}_{\mathbf{L}_{p+1}}}(\cup\mathcal{Q}_t^{p+1,p+1}) = \mathrm{dens}_{\{t\}\times\boldsymbol{\Omega}_{\mathbf{L}_{p+1}}}\big(\{t\}\times\cup\mathcal{Q}_t^{p+1,p+1}\big) \geqslant \alpha.$$

The proof of Claim 9.52 is completed. □

CLAIM 9.53. *Let $t, t' \in A^{<|L_{p+1}|}$ such that t and t' are (a,b)-equivalent. Then for every $\ell \in [p+1]$ we have $\mathcal{Q}_t^{\ell,p+1} = \mathcal{Q}_{t'}^{\ell,p+1}$. That is, condition* (C7) *is satisfied.*

PROOF. Assume that $p \geqslant 1$ and $\ell \in [p]$. For every $s \in \mathcal{C}_t$ and every $s' \in \mathcal{C}_{t'}$ let

$$\mathcal{Q}_s^t = \{C_t^{s,\ell,Q} : Q \in \mathcal{Q}_s^{\ell,p}\} \ \text{ and } \ \mathcal{Q}_{s'}^{t'} = \{C_{t'}^{s',\ell,Q'} : Q' \in \mathcal{Q}_{s'}^{\ell,p}\}. \tag{9.164}$$

By (9.151), the families $\{\mathcal{Q}_s^t : s \in \mathcal{C}_t\}$ and $\{\mathcal{Q}_{s'}^{t'} : s' \in \mathcal{C}_{t'}\}$ are partitions of $\mathcal{Q}_t^{\ell,p+1}$ and $\mathcal{Q}_{t'}^{\ell,p+1}$ respectively. Let $r_{t,t'}\colon \mathcal{C}_t \to \mathcal{C}_{t'}$ be the bijection defined in (9.10) and let $s \in \mathcal{C}_t$ be arbitrary. By Fact 9.6, we have

$$\Omega_t^s = \Omega_{t'}^{r_{t,t'}(s)} \tag{9.165}$$

and moreover, since t and t' are (a,b)-equivalent, the words s and $r_{t,t'}(s)$ are (a,b)-equivalent. Thus, invoking our inductive assumptions, we obtain that

$$\mathcal{Q}_s^{\ell,p} = \mathcal{Q}_{r_{t,t'}(s)}^{\ell,p}. \tag{9.166}$$

Therefore,

$$\begin{aligned}
\mathcal{Q}_s^t &\overset{(9.164)}{=} \{C_t^{s,\ell,Q} : Q \in \mathcal{Q}_s^{\ell,p}\} \overset{(9.150)}{=} \{Q\times\Omega_t^s : Q \in \mathcal{Q}_s^{\ell,p}\} \\
&\overset{(9.165)}{=} \{Q\times\Omega_{t'}^{r_{t,t'}(s)} : Q \in \mathcal{Q}_s^{\ell,p}\} \overset{(9.166)}{=} \{Q\times\Omega_{t'}^{r_{t,t'}(s)} : Q \in \mathcal{Q}_{r_{t,t'}(s)}^{\ell,p}\} \\
&\overset{(9.150)}{=} \{C_{t'}^{r_{t,t'}(s),\ell,Q} : Q \in \mathcal{Q}_{r_{t,t'}(s)}^{\ell,p}\} \overset{(9.164)}{=} \mathcal{Q}_{r_{t,t'}(s)}^{t'}.
\end{aligned}$$

Since the map $r_{t,t'}\colon \mathcal{C}_t \to \mathcal{C}_{t'}$ is a bijection, we conclude that $\mathcal{Q}_t^{\ell,p+1} = \mathcal{Q}_{t'}^{\ell,p+1}$.

We proceed to the case "$\ell = p+1$". To this end, we need to do some preparatory work. Specifically, let U and U' be two finite-dimensional Carlson–Simpson spaces of $A^{<\mathbb{N}}$ with $\dim(U)=\dim(U')$ and let I_U and $\mathrm{I}_{U'}$ be the canonical isomorphisms associated with U and U' respectively (see Definition 1.10). Notice that the map

$$U \ni u \mapsto \mathrm{I}_{U'}\big(\mathrm{I}_U^{-1}(u)\big) \in U'$$

is a bijection. This bijection will be called the *canonical isomorphism associated with the pair* U, U' and will be denoted by $\mathrm{I}_{U,U'}$. Observe that $\mathrm{I}_{U,U'}\big(U(i)\big)=U'(i)$ for every $i\in\{0,\dots,\dim(U)\}$.

SUBCLAIM 9.54. *Let $\mathcal{U}_t$ and $\mathcal{U}_{t'}$ be as in* (9.142). *Then there exists a map $f_{t,t'}\colon \cup\mathcal{U}_t \to \cup\mathcal{U}_{t'}$ with the following properties.*

(a) *The map $f_{t,t'}$ is a bijection. Moreover, for every $s\in\cup\mathcal{U}_t$ the words s and $f_{t,t'}(s)$ are (a,b)-equivalent.*

(b) *For every $U\in\mathcal{U}_t$ we have $f_{t,t'}(U)\in\mathcal{U}_{t'}$ and the restriction of $f_{t,t'}$ on U coincides with the canonical isomorphism associated with the pair $U, f_{t,t'}(U)$.*

(c) *For every $s\in\mathcal{C}_t$ we have $f_{t,t'}(s)=r_{t,t'}(s)$.*

PROOF OF SUBCLAIM 9.54. Let $i\in\{0,\dots,|L_{p+1}|-1\}$ be the unique integer such that $t,t'\in A^i$. If $i=0$ (that is, if $t=t'=\emptyset$), then the desired map is the identity. So assume that $i\geqslant 1$ and let Z_t and $Z_{t'}$ be as in (9.140). First we define a map $h^0_{t,t'}\colon Z_t\to Z_{t'}$ as follows. For every $z\in Z_t$ we select $\omega_z\in\Omega_{L_{p+1}}$ such that $z=\mathrm{c}_{L_{p+1}}(t\upharpoonright(i-1),\omega_z)^\frown t(i-1)$ and we set

$$h^0_{t,t'}(z)=\mathrm{c}_{L_{p+1}}(t'\upharpoonright(i-1),\omega_z)^\frown t'(i-1).$$

Note that: (i) $h^0_{t,t'}(z)$ is independent of the choice of ω_z, and (ii) the map $h^0_{t,t'}$ is a bijection. Also observe that, since t and t' are (a,b)-equivalent, for every $z\in Z_t$ the words z and $h^0_{t,t'}(z)$ are (a,b)-equivalent too.

Next we set

$$\mathcal{R}_t=\{R^z : z\in Z_t\} \ \text{ and } \ \mathcal{R}_{t'}=\{R^{z'} : z'\in Z_{t'}\}$$

where, as in (9.141), $R^z=\{z^\frown x : x\in A^{<\bar{m}+1}\}$ and $R^{z'}=\{z'^\frown x : x\in A^{<\bar{m}+1}\}$ for every $z\in Z_t$ and every $z'\in Z_{t'}$ respectively. We define $h_{t,t'}\colon \cup\mathcal{R}_t\to\cup\mathcal{R}_{t'}$ by the rule $h_{t,t'}(z^\frown x)=h^0_{t,t'}(z)^\frown x$. Using the aforementioned properties of $h^0_{t,t'}$, we see that the following are satisfied.

(H1) The map $h_{t,t'}$ is a bijection. Moreover, for every $r\in\cup\mathcal{R}_t$ the words r and $h_{t,t'}(r)$ are (a,b)-equivalent.

(H2) For every $z\in Z_t$ the restriction of $h_{t,t'}$ on R^z is onto $R^{h^0_{t,t'}(z)}$ and coincides with the canonical isomorphism associated with the pair $R^z, R^{h^0_{t,t'}(z)}$.

(H3) For every $\omega\in\Omega_{L_{p+1}}$ we have $h_{t,t'}\big(\mathrm{c}_{L_{p+1}}(t,\omega)\big)=\mathrm{c}_{L_{p+1}}(t',\omega)$.

We are in a position to define the map $f_{t,t'}$. Specifically, let $\mathrm{I}_{V_{p+1}}$ be the canonical isomorphism associated with V_{p+1} and for every $s\in\cup\mathcal{U}_t$ we set

$$f_{t,t'}(s)=\mathrm{I}_{V_{p+1}}\Big(h_{t,t'}\big(\mathrm{I}^{-1}_{V_{p+1}}(s)\big)\Big). \tag{9.167}$$

(Notice that, by (9.142), $f_{t,t'}(s)$ is well-defined.) It follows readily from properties (H1)–(H3) isolated above that the map $f_{t,t'}$ is as desired. The proof of Subclaim 9.54 is completed. $\square$

Now let $U \in \mathcal{U}_t$ be arbitrary. By Subclaim 9.54, we have $f_{t,t'}(U) \in \mathcal{U}_{t'}$. Moreover, for every $s \in U$ the words s and $f_{t,t'}(s)$ are (a,b)-equivalent. Thus, by Fact 9.43 and (9.133), we obtain that

$$\Gamma_U = \Gamma_{f_{t,t'}(U)}. \tag{9.168}$$

Next recall that, by Fact 9.45, we have $U(i_0) \subseteq \mathcal{C}_t$. Hence, invoking Subclaim 9.54 once again, we conclude that

$$f_{t,t'}(U)(i_0) = f_{t,t'}\big(U(i_0)\big) = \{r_{t,t'}(s) : s \in U(i_0)\}. \tag{9.169}$$

We are ready for the last step of the argument. For every $U \in \mathcal{U}_t$ and every $U' \in \mathcal{U}_{t'}$ we set

$$\mathcal{Q}^t_U = \{Q^{\boldsymbol{\omega},U}_t : \boldsymbol{\omega} \in \Gamma_U\} \ \text{ and } \ \mathcal{Q}^{t'}_{U'} = \{Q^{\boldsymbol{\omega}',U'}_{t'} : \boldsymbol{\omega}' \in \Gamma_{U'}\}. \tag{9.170}$$

By Fact 9.45 and (9.147), the families $\{\mathcal{Q}^t_U : U \in \mathcal{U}_t\}$ and $\{\mathcal{Q}^{t'}_{U'} : U' \in \mathcal{U}_{t'}\}$ are partitions of $\mathcal{Q}^{p+1,p+1}_t$ and $\mathcal{Q}^{p+1,p+1}_{t'}$ respectively. Moreover, for every $U \in \mathcal{U}_t$,

$$\begin{aligned}
\mathcal{Q}^t_U \ &\overset{(9.170)}{=} \ \{Q^{\boldsymbol{\omega},U}_t : \boldsymbol{\omega} \in \Gamma_U\} \overset{(9.146)}{=} \Big\{\{\boldsymbol{\omega}\} \times \bigcup_{s \in U(i_0)} \Omega^s_t : \boldsymbol{\omega} \in \Gamma_U\Big\} \\
&\overset{(9.168)}{=} \ \Big\{\{\boldsymbol{\omega}'\} \times \bigcup_{s \in U(i_0)} \Omega^s_t : \boldsymbol{\omega}' \in \Gamma_{f_{t,t'}(U)}\Big\} \\
&\overset{(9.165)}{=} \ \Big\{\{\boldsymbol{\omega}'\} \times \bigcup_{s \in U(i_0)} \Omega^{r_{t,t'}(s)}_{t'} : \boldsymbol{\omega}' \in \Gamma_{f_{t,t'}(U)}\Big\} \\
&\overset{(9.169)}{=} \ \Big\{\{\boldsymbol{\omega}'\} \times \bigcup_{s' \in f_{t,t'}(U)(i_0)} \Omega^{s'}_{t'} : \boldsymbol{\omega}' \in \Gamma_{f_{t,t'}(U)}\Big\} = \mathcal{Q}^{t'}_{f_{t,t'}(U)}.
\end{aligned} \tag{9.171}$$

Finally, observe that, by Subclaim 9.54, the map $\mathcal{U}_t \ni U \mapsto f_{t,t'}(U) \in \mathcal{U}_{t'}$ is a bijection. Hence, by (9.171), we conclude that $\mathcal{Q}^{p+1,p+1}_t = \mathcal{Q}^{p+1,p+1}_{t'}$ and the proof of Claim 9.53 is completed. □

By Claims 9.46 up to 9.53, the pair (V_{p+1}, L_{p+1}) and the families $\mathcal{Q}^{\ell,p+1}_t$ constructed in Steps 1, 2 and 3, satisfy all required conditions. This completes the recursive selection, and as we have already indicated, the entire proof of Theorem 9.38 is completed.

9.5.5. Consequences. Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every positive integer l and every $0 < \beta \leqslant 1$ the number $\mathrm{DCS}(k,l,\beta)$ has been estimated. We define $H \colon \mathbb{N} \times (0,1] \to \mathbb{N}$ by $H(0,\gamma) = 0$ and

$$H(m,\gamma) = \mathrm{RegCS}(k+1, 2, N_{p_0}+1, \gamma^2/2) \tag{9.172}$$

if $m \geqslant 1$, where p_0 and N_{p_0} are as in (9.109) and (9.110) respectively for the parameters k, m and γ. Next, for every $n \in \{0,\dots,k\}$ we define, recursively, $H^{(n)} \colon \mathbb{N} \times (0,1] \to \mathbb{N}$ by the rule $H^{(0)}(m,\gamma) = m$ and

$$H^{(n+1)}(m,\gamma) = H\big(H^{(n)}(m,\gamma),\gamma\big). \tag{9.173}$$

Finally, for every $0 < \gamma \leqslant 1$ let

$$\xi = \xi(\gamma) = \frac{\gamma^{3^k}}{\big(2^{1/2} \cdot 32\big)^{3^k - 1}}. \tag{9.174}$$

We have the following corollary. This result together with Corollary 9.37 form the basis of the proof of Theorem 9.2.

COROLLARY 9.55. *Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every positive integer l and every $0 < \beta \leqslant 1$ the number $\mathrm{DCS}(k, l, \beta)$ has been estimated.*

Let $0 < \gamma, \delta \leqslant 1$. Also let A be an alphabet with $|A| = k + 1$, $a \in A$, V a finite-dimensional Carlson–Simpson space of $A^{<\mathbb{N}}$ and I a nonempty subset of $\{0, \dots, \dim(V)\}$. Set $B = A \setminus \{a\}$ and assume that we are given a subset D of $A^{<\mathbb{N}}$ and a family $\{S_b : b \in B\}$ of subsets of $A^{<\mathbb{N}}$ with the following properties.

(a) *For every $b \in B$ the set S_b is (a, b)-insensitive in V.*
(b) *For every $i \in I$ we have $\mathrm{dens}_{V(i)}\big(\bigcap_{b\in B} S_b \cap D\big) \geqslant (\delta+\gamma)\mathrm{dens}_{V(i)}\big(\bigcap_{b\in B} S_b\big)$ and $\mathrm{dens}_{V(i)}\big(\bigcap_{b\in B} S_b\big) \geqslant \gamma$.*

Finally, let $m \in \mathbb{N}$ with $m \geqslant 1$ and suppose that

$$|I| \geqslant H^{(k)}(m, \xi) \tag{9.175}$$

where $H^{(k)}$ and ξ are defined in (9.173) and (9.174) respectively for the parameters k, m and γ. Then there exist a finite-dimensional Carlson–Simpson subspace W of V and a subset I' of $\{0, \dots, \dim(W)\}$ of cardinality m such that

$$\mathrm{dens}_{W(i)}(D) \geqslant \delta + \xi \tag{9.176}$$

for every $i \in I'$.

PROOF. We define a sequence (γ_r) in $\mathbb{R}$ recursively by the rule $\gamma_0 = \gamma/2$ and $\gamma_{r+1} = \gamma_r^3/512$. Note that for every $r \in \mathbb{N}$ we have

$$\gamma_r = \frac{\gamma^{3^r}}{2 \cdot \big(2^{1/2} \cdot 32\big)^{3^r - 1}}. \tag{9.177}$$

In particular, by (9.174) and (9.177), we see that $2\gamma_k = \xi$.

Next let $\{b_1, \dots, b_k\}$ be an enumeration of B and let, for notational simplicity, $S_r = S_{b_r}$ for every $r \in [k]$. We also set $S_{k+1} = V$. Using Theorem 9.38, we select a finite sequence $V_0 \supseteq V_1 \supseteq \cdots \supseteq V_k$ of Carlson–Simpson subspaces of V and a finite sequence $I_0, I_1 \dots, I_k$ of finite subsets of $\mathbb{N}$ with $V_0 = V$ and $I_0 = I$, and satisfying the following conditions for every $r \in \{0, \dots, k\}$.

(C1) We have $I_r \subseteq \{0, \dots, \dim(V_r)\}$ and $|I_r| = H^{(k-r)}(m, \xi)$.
(C2) For every $i \in I_r$ we have

$$\mathrm{dens}_{V_r(i)}\Big(\bigcap_{j=r+1}^{k+1} S_j \cap D\Big) \geqslant (\delta + 2\gamma_r)\, \mathrm{dens}_{V_r(i)}\Big(\bigcap_{j=r+1}^{k+1} S_j\Big)$$

and

$$\mathrm{dens}_{V_r(i)}\Big(\bigcap_{j=r+1}^{k+1} S_j\Big) \geqslant 2\gamma_r.$$

Finally, we set $W = V_k$ and $I' = I_k$ and we claim that with these choices the result follows. Indeed notice first that, by condition (C3), we have $I' \subseteq \{0, \dots, \dim(W)\}$ and $|I'| = H^{(0)}(m, \xi) = m$. On the other hand, by condition (C2), we see that $\mathrm{dens}_{W(i)}(V \cap D) \geqslant (\delta + 2\gamma_k)\, \mathrm{dens}_{W(i)}(V)$ for every $i \in I'$. Since $W = V_k \subseteq V$, this implies that $\mathrm{dens}_{W(i)}(D) \geqslant \delta + 2\gamma_k = \delta + \xi$ for every $i \in I'$. The proof of Corollary 9.55 is thus completed. □

9.6. Proof of Theorem 9.2

In this section we will complete the proof of Theorem 9.2 following the inductive scheme outlined in Subsection 9.4.1. First we recall that the numbers $\mathrm{DCS}(2, 1, \delta)$ are estimated in Proposition 9.25. By induction on m and using Theorem 9.29, we may also estimate the numbers $\mathrm{DCS}(2, m, \delta)$.

Next we argue for the general inductive step. Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every positive integer l and every $0 < \beta \leqslant 1$ the number $\mathrm{DCS}(k, l, \beta)$ has been estimated. We fix $0 < \delta \leqslant 1$. Let η_1 be as in (9.94). Recall that

$$\eta_1 = \frac{\delta^2}{120k \cdot |\mathrm{SubCS}_1\big([k]^{<\Lambda}\big)|}$$

where $\Lambda = \lceil 8\delta^{-1}\mathrm{DCS}(k, 1, \delta/8)\rceil$. We set

$$\varrho = \xi(\eta_1^2/2) \overset{(9.174)}{=} \frac{(\eta_1^2/2)^{3^k}}{\big(2^{1/2} \cdot 32\big)^{3^k - 1}} \tag{9.178}$$

and we define $F_\delta \colon \mathbb{N} \to \mathbb{N}$ by the rule

$$F_\delta(m) = g_1\Big(\big\lceil \eta_1^{-4}(k+1)k \cdot H^{(k)}(m, \varrho)\big\rceil, \eta_1^2/2\Big) \tag{9.179}$$

where g_1 and $H^{(k)}(m, \varrho)$ are as in (9.95) and (9.173) respectively. The following proposition is the heart of the density increment strategy. It is a straightforward consequence of Corollaries 9.37 and 9.55.

Proposition 9.56. *Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every positive integer l and every $0 < \beta \leqslant 1$ the number* $\mathrm{DCS}(k, l, \beta)$ *has been estimated.*

Let $0 < \delta \leqslant 1$, A an alphabet with $|A| = k+1$ and L a nonempty finite subset of $\mathbb{N}$. Also let $D \subseteq A^{<\mathbb{N}}$ such that $\mathrm{dens}_{A^l}(D) \geqslant \delta$ for every $l \in L$ and assume that D contains no Carlson–Simpson line of $A^{<\mathbb{N}}$. Finally, let m be a positive integer and suppose that $|L| \geqslant F_\delta(m)$ where F_δ is as in (9.179)*. Then there exist a finite-dimensional Carlson–Simpson space W of $A^{<\mathbb{N}}$ and a subset I of $\{0, \dots, \dim(W)\}$ with $|I| = m$ such that $\mathrm{dens}_{W(i)}(D) \geqslant \delta + \varrho$ for every $i \in I$ where ϱ is as in* (9.178)*.*

With Proposition 9.56 at our disposal, the numbers $\mathrm{DCS}(k+1, 1, \delta)$ can be estimated easily. Specifically, we have the following corollary.

Corollary 9.57. *Let $k \in \mathbb{N}$ with $k \geqslant 2$ and assume that for every positive integer l and every $0 < \beta \leqslant 1$ the number* $\mathrm{DCS}(k, l, \beta)$ *has been estimated. Then for every $0 < \delta \leqslant 1$ we have*

$$\mathrm{DCS}(k+1, 1, \delta) \leqslant F_\delta^{(\lceil \varrho^{-1} \rceil)}(1). \tag{9.180}$$

Finally, as in the case "$k = 2$", the numbers $\mathrm{DCS}(k+1, m, \delta)$ can be estimated using Theorem 9.29 and Corollary 9.57. This completes the proof of the general inductive step, and so the entire proof of Theorem 9.2 is completed.

9.7. Proof of Theorem 9.1

As we have already mentioned, the proof is based on Theorem 9.2. Specifically, for every $k \in \mathbb{N}$ with $k \geqslant 2$ and every $0 < \delta \leqslant 1$ let $\Lambda_0(k,\delta)$ and $\Theta_0(k,\delta)$ be as in (9.54). Also let $h_{k,\delta}(n)\colon \mathbb{N} \to \mathbb{N}$ be as in (9.55). Recall that

$$\Lambda_0(k,\delta) = \lceil 16\delta^{-2}\mathrm{DCS}(k,1,\delta^2/16)\rceil \ \text{ and } \ \Theta_0(k,\delta) = \frac{\delta^2/8}{|\mathrm{SubCS}_1\big([k]^{<\Lambda_0(k,\delta)}\big)|}$$

and

$$h_{k,\delta}(n) = \Lambda_0(k,\delta) + \lceil 2\Theta_0(k,\delta)^{-1} n\rceil.$$

We have the following variant of Lemma 9.30.

Lemma 9.58. *Let $k \in \mathbb{N}$ with $k \geqslant 2$ and $0 < \delta \leqslant 1$. Also let A be an alphabet with $|A| = k$, M an infinite subset of $\mathbb{N}$ and $D \subseteq A^{<\mathbb{N}}$ such that $\mathrm{dens}_{A^m}(D) \geqslant \delta$ for every $m \in M$. Then there exist a Carlson–Simpson line V of $A^{<\mathbb{N}}$ with $V \subseteq D$ and an infinite subset M' of M with the following property. If m_1 is the unique integer with $V(1) \subseteq A^{m_1}$, then $m_1 < \min(M')$ and for every $m \in M'$ we have*

$$\mathrm{dens}_{A^{m-m_1}}\big(\{w \in A^{<\mathbb{N}} : v^\frown w \in D \text{ for every } v \in V(1)\}\big) \geqslant 2^{-1}\Theta_0(k,\delta/2)$$

where $\Theta_0(k,\delta/2)$ is as in (9.54).

Proof. For every $s \in A^{<\mathbb{N}}$ let $D_s = \{w \in A^{<\mathbb{N}} : s^\frown w \in D\}$ and define

$$\delta_s = \limsup_{m\in M} \mathrm{dens}_{A^{m-|s|}}(D_s).$$

Set $\delta^* = \sup\{\delta_s : s \in A^{<\mathbb{N}}\}$ and note that $\delta \leqslant \delta^* \leqslant 1$. We fix $0 < \delta_0 \leqslant 1$ with

$$\delta/2 < \delta_0 < \delta^* < \delta_0 + \delta_0^2/8 \tag{9.181}$$

and we select $s_0 \in A^{<\mathbb{N}}$ and an infinite subset N of $\{m \in M : m > |s_0|\}$ such that

$$\delta_0 \leqslant \mathrm{dens}_{A^{m-|s_0|}}(D_{s_0}) \tag{9.182}$$

for every $m \in N$. Let I_0 be the initial segment of N with $|I_0| = \Lambda_0(k,\delta_0)$ where $\Lambda_0(k,\delta_0)$ is as in (9.54). By the definition of δ^* and (9.181), there exists $m_0 \in N$ with $m_0 \geqslant \max(I_0)$ such that for every $t \in \bigcup_{m\in I_0} A^{m-|s_0|}$ and every $m \in N$ with $m \geqslant m_0$ we have

$$\mathrm{dens}_{A^{m-|s_0^\frown t|}}(D_{s_0^\frown t}) < \delta_0 + \delta_0^2/8. \tag{9.183}$$

We also fix a sequence (J_n) of pairwise disjoint subsets of $\{m \in N : m \geqslant m_0\}$ such that $|J_n| = \lceil 2\Theta_0(k,\delta_0)^{-1}\rceil$ for every $n \in \mathbb{N}$.

Let $n \in \mathbb{N}$ be arbitrary. We set $K_n = I_0 \cup J_n$ and we observe that

$$|K_n| = |I_0| + |J_n| = \Lambda_0(k,\delta_0) + \lceil 2\Theta_0(k,\delta_0)^{-1}\rceil = h_{k,\delta_0}(1). \tag{9.184}$$

Set $D' = D_{s_0}$, $L_0 = \{m - |s_0| : m \in I_0\}$ and $L = \{m - |s_0| : m \in K_n\}$ and notice that, by (9.182), we have $\mathrm{dens}_{A^\ell}(D') \geqslant \delta_0$ for every $\ell \in L$. Moreover, L_0 is the initial segment of L with $|L_0| = \Lambda_0(k,\delta_0)$ and, by (9.184), we have $|L| = h_{k,\delta_0}(1)$.

It follows, in particular, that Lemma 9.30 can be applied to D', L and L_0. The first alternative of Lemma 9.30 does not hold true since, by (9.183), we have

$$\mathrm{dens}_{A^{\ell-|t|}}\big(\{w \in A^{<\mathbb{N}} : t^\frown w \in D'\}\big) < \delta_0 + \delta_0^2/8$$

for every $t \in \bigcup_{\ell \in L_0} A^\ell$ and every $\ell \in L \setminus L_0$. Therefore, there exist $l_n \in L \setminus L_0$ and a Carlson–Simpson line S_n of $A^{<\mathbb{N}}$ contained in D' with $L(S_n) \subseteq L_0$ (where, as in (1.33), $L(S_n)$ is the level set of S_n) and such that, denoting by $\ell_{1,n}$ the unique integer in L_0 with $S_n(1) \subseteq A^{\ell_{1,n}}$, we have

$$\mathrm{dens}_{A^{l_n-\ell_{1,n}}}\big(\{w \in A^{<\mathbb{N}} : s^\frown w \in D' \text{ for every } s \in S_n(1)\}\big) \geqslant 2^{-1}\Theta_0(k,\delta_0).$$

Set $V_n = \{s_0{}^\frown s : s \in S_n\}$, $i_n = \ell_{1,n} + |s_0|$ and $j_n = l_n + |s_0|$. Notice that: (i) $V_n \subseteq D$, (ii) $i_n \in I_0$, (iii) $j_n \in J_n$, (iv) $L(V_n) \subseteq I_0$, (v) $V_n(1) \subseteq A^{i_n}$ and (vi) $i_n < j_n$. On the other hand, the fact that $\delta/2 < \delta_0$ yields that $\Theta_0(k,\delta_0) \geqslant \Theta_0(k,\delta/2)$. Hence, we obtain that

$$\mathrm{dens}_{A^{j_n-i_n}}\big(\{w \in A^{<\mathbb{N}} : v^\frown w \in D \text{ for every } v \in V_n(1)\}\big) \geqslant 2^{-1}\Theta_0(k,\delta/2).$$

By the classical pigeonhole principle, there exist an infinite subset P of $\mathbb{N}$ and a Carlson–Simpson line V of $A^{<\mathbb{N}}$ such that $V_n = V$ for every $n \in P$. Thus, setting $M' = \{j_n : n \in P\}$, we see that V and M' are as desired. The proof of Lemma 9.58 is completed. □

We are now in a position to complete the proof of Theorem 9.1. Let A be an alphabet with $|A| \geqslant 2$. Also let $0 < \delta \leqslant 1$ and $D \subseteq A^{<\mathbb{N}}$ such that

$$\limsup_{n\to\infty} \frac{|D \cap A^n|}{|A^n|} > \delta.$$

We fix an infinite subset M of $\mathbb{N}$ such that $\mathrm{dens}_{A^m}(D) \geqslant \delta$ for every $m \in M$. Also, we define a sequence (δ_n) in $(0,1]$ by the rule

$$\delta_0 = \delta \ \text{ and } \ \delta_{n+1} = 2^{-1}\Theta_0(|A|,\delta_n/2) \tag{9.185}$$

where $\Theta_0(|A|,\delta_n/2)$ is as in (9.54). Using Lemma 9.58 we select, recursively,

(a) a sequence (D_n) of subsets of $A^{<\mathbb{N}}$ with $D_0 = D$,
(b) a sequence (V_n) of Carlson–Simpson lines of $A^{<\mathbb{N}}$, and
(c) two sequences (M_n) and (M'_n) of infinite subsets of $\mathbb{N}$ with $M_0 = M$,

such that for every $n \in \mathbb{N}$ the following conditions are satisfied.

(C1) For every $m \in M_n$ we have $\mathrm{dens}_{A^m}\big(D_n\big) \geqslant \delta_n$.
(C2) We have $V_n \subseteq D_n$.
(C3) If m_n is the unique integer with $V_n(1) \subseteq A^{m_n}$, then

$$M'_n \subseteq \{m \in M_n : m > m_n\} \ \text{ and } \ M_{n+1} = \{m' - m_n : m' \in M'_n\}.$$

(C4) We have $D_{n+1} = \big\{w \in A^{<\mathbb{N}} : v^\frown w \in D_n \text{ for every } v \in V_n(1)\big\}$.

Next, for every $n \in \mathbb{N}$ let $\langle v^n, v_0^n\rangle$ be the Carlson–Simpson system generating V_n and set $w = v^0$ and $u_n = v_0^n{}^\frown v^{n+1}$. Observe that u_n is a left variable word over A for every $n \in \mathbb{N}$. We will show that

$$\{w\} \cup \big\{w^\frown u_0(a_0)^\frown \dots ^\frown u_n(a_n) : n \in \mathbb{N} \text{ and } a_0, \dots, a_n \in A\big\} \subseteq D.$$

To this end notice first that, by condition (C2), we have $w \in D_0 = D$. Moreover, by condition (C4), we see that $V_n(1)^\frown D_{n+1} \subseteq D_n$ for every $n \in \mathbb{N}$. By induction, this inclusion yields that

$$V_0(1)^\frown \dots ^\frown V_n(1)^\frown D_{n+1} \subseteq D. \tag{9.186}$$

Now let $n \in \mathbb{N}$ and $a_0, \dots, a_n \in A$ be arbitrary. By the choice of w and $u_0, \dots, u_n$, we obtain that

$$w^\frown u_0(a_0)^\frown \dots ^\frown u_n(a_n) \in V_0(1)^\frown \dots ^\frown V_n(1)^\frown V_{n+1}(0). \tag{9.187}$$

On the other hand, by condition (C2), we have $V_{n+1}(0) \subseteq D_{n+1}$. Thus, combining (9.186) and (9.187), we conclude that $w^\frown u_0(a_0)^\frown \dots ^\frown u_n(a_n) \in D$. The proof of Theorem 9.1 is completed.

9.8. Applications

9.8.1. Connections with other results. In this section we will present some applications of Theorems 9.1 and 9.2. We begin with a discussion on the relation of the density Carlson–Simpson theorem with other results in Ramsey theory. In this direction, we first observe that the density Hales–Jewett theorem follows from Theorem 9.1 with a standard compactness argument. In fact, we have the following finer quantitative information.

PROPOSITION 9.59. *For every integer $k \geqslant 2$ and every $0 < \delta \leqslant 1$ we have*

$$\mathrm{DHJ}(k, \delta) \leqslant \mathrm{DCS}(k, 1, \delta). \tag{9.188}$$

PROOF. Fix an integer $k \geqslant 2$ and $0 < \delta \leqslant 1$, and let A be an alphabet with $|A| = k$. Also let $n \in \mathbb{N}$ with $n \geqslant \mathrm{DCS}(k, 1, \delta)$ and $D \subseteq A^n$ with $|D| \geqslant \delta |A^n|$. For every $\ell \in [n]$ and every $t \in A^{n-\ell}$ let $D_t = \{s \in A^\ell : t^\frown s \in D\}$ be the section of D at t and observe that

$$\mathbb{E}_{t \in A^{n-\ell}} \mathrm{dens}(D_t) = \mathrm{dens}(D) \geqslant \delta.$$

Hence, for every $\ell \in [n]$ we may select $t_\ell \in A^{n-\ell}$ such that $\mathrm{dens}(D_{t_\ell}) \geqslant \delta$. Let

$$D' = \bigcup_{\ell \in [n]} D_{t_\ell}$$

and notice that $\mathrm{dens}_{A^\ell}(D') = \mathrm{dens}_{A^\ell}(D_{t_\ell}) \geqslant \delta$ for every $\ell \in [n]$. By the choice of n, there exists a Carlson–Simpson line W of $A^{<\mathbb{N}}$ which is contained in D'. Let $\langle w, w_0 \rangle$ be the Carlson–Simpson system generating W. Also let $\ell \in [n]$ be the unique integer such that $W(1) \subseteq A^\ell$ and set

$$V = \{t_\ell{}^\frown w^\frown w_0(a) : a \in A\}.$$

Then observe that V is a combinatorial line of A^n and $V \subseteq D$. This shows that the estimate in (9.188) is satisfied and the proof of Proposition 9.59 is completed. □

The next result deals with dense subsets of products of homogeneous trees and is due to Dodos, Kanellopoulos and Tyros [**DKT1**].

THEOREM 9.60. *For every integer $d \geqslant 1$, every $b_1, \dots, b_d \in \mathbb{N}$ with $b_i \geqslant 2$ for all $i \in [d]$, every integer $\ell \geqslant 1$ and every $0 < \delta \leqslant 1$ there exists a positive integer N with the following property. If $\mathbf{T} = (T_1, \dots, T_d)$ is a vector homogeneous tree with $h(\mathbf{T}) \geqslant N$ and $b_{T_i} = b_i$ for all $i \in [d]$, L is a subset of $\{0, \dots, h(\mathbf{T}) - 1\}$ with $|L| \geqslant N$ and D is a subset of the level product $\otimes \mathbf{T}$ of $\mathbf{T}$ satisfying*

$$|D \cap \big(T_1(n) \times \cdots \times T_d(n)\big)| \geqslant \delta |T_1(n) \times \cdots \times T_d(n)|$$

for every $n \in L$, then there exists $\mathbf{S} \in \mathrm{Str}_\ell(\mathbf{T})$ such that $\otimes \mathbf{S} \subseteq D$. The least positive integer with this property will be denoted by $\mathrm{UDHL}(b_1, \dots, b_d \,|\, \ell, \delta)$.

Of course, the main point in Theorem 9.60 is that the result is independent of the position of the finite set L. As in Proposition 9.59, we will obtain upper bounds for the numbers $\mathrm{UDHL}(b_1, \dots, b_d \,|\, \ell, \delta)$ which are expressed in terms of the numbers $\mathrm{DCS}(k, m, \delta)$.

PROOF OF THEOREM 9.60. It is similar to the proof of Corollary 4.11. Fix the parameters $d, b_1, \dots, b_d, \ell, \delta$. Since $\mathrm{UDHL}(b_1, \dots, b_d \,|\, 1, \delta) = 1$, we may assume that $\ell \geqslant 2$. We claim that

$$\mathrm{UDHL}(b_1, \dots, b_d \,|\, \ell, \delta) \leqslant \mathrm{DCS}\Big(\prod_{i=1}^{d} b_i, \ell - 1, \delta\Big). \tag{9.189}$$

Indeed, set $N = \mathrm{DCS}\big(\prod_{i=1}^{d} b_i, \ell - 1, \delta\big)$ and let $\mathbf{T} = (T_1, \dots, T_d)$ be a vector homogeneous tree with $h(\mathbf{T}) \geqslant N$ and $b_{T_i} = b_i$ for all $i \in [d]$. Clearly, we may additionally assume that $h(\mathbf{T})$ is finite and $T_i = [b_i]^{<h(\mathbf{T})}$ for every $i \in [d]$. For every $i \in [d]$ let $\pi_i \colon \mathbb{A} \to [b_i]$ and $\bar{\pi}_i \colon \mathbb{A}^{<h(\mathbf{T})} \to [b_i]^{<h(\mathbf{T})}$ be as in Corollary 4.11 and define the map $\mathrm{I} \colon \mathbb{A}^{<h(\mathbf{T})} \to \otimes \mathbf{T}$ by $\mathrm{I}(w) = \big(\bar{\pi}_1(w), \dots, \bar{\pi}_d(w)\big)$. Recall that I is a bijection and satisfies $\mathrm{I}(\mathbb{A}^n) = [b_1]^n \times \cdots \times [b_d]^n$ for every $n \in \{0, \dots, h(\mathbf{T}) - 1\}$.

Now let $D \subseteq \otimes \mathbf{T}$ and $L \subseteq \{0, \dots, h(\mathbf{T}) - 1\}$ with $|L| \geqslant N$ and assume that $|D \cap \otimes \mathbf{T}(n)| \geqslant \delta |\otimes \mathbf{T}(n)|$ for every $n \in L$. We set $D' = \mathrm{I}^{-1}(D)$ and we observe that $|D' \cap \mathbb{A}^n| \geqslant \delta |\mathbb{A}^n|$ for every $n \in L$. Hence, by the choice of N, there exists a Carlson–Simpson space of $\mathbb{A}^{<\mathbb{N}}$ of dimension $\ell - 1$ which is contained in D'. For every $i \in [d]$ set $S_i = \bar{\pi}_i(S)$ and notice that S_i is a Carlson–Simpson subspace of $[b_i]^{<\mathbb{N}}$ having the same level set as S. Therefore, $\mathbf{S} = (S_1, \dots, S_d)$ is a vector strong subtree of $\mathbf{T}$ with $h(\mathbf{S}) = \ell$. Since $\mathrm{I}(S) = \otimes \mathbf{S}$, we conclude that $\otimes \mathbf{S}$ is contained in D and the proof of Theorem 9.60 is completed. $\square$

By Theorem 9.1 and arguing precisely as above, we obtain the following infinite version of Theorem 9.60 due to Dodos, Kanellopoulos and Karagiannis [**DKK**].

THEOREM 9.61. *For every vector homogeneous tree $\mathbf{T} = (T_1, \dots, T_d)$ of infinite height and every subset D of the level product $\otimes \mathbf{T}$ of $\mathbf{T}$ satisfying*

$$\limsup_{n \to \infty} \frac{|D \cap \big(T_1(n) \times \cdots \times T_d(n)\big)|}{|T_1(n) \times \cdots \times T_d(n)|} > 0$$

there exists a vector strong subtree $\mathbf{S}$ of $\mathbf{T}$ of infinite height whose level product is contained in D.

9.8.2. Homogeneous trees. Let $H\colon [0,1] \to [0,1]$ be the binary entropy function—see (E.21)—and for every integer $b \geqslant 2$ and every $0 < \delta \leqslant 1$ let $c(b,\delta)$ be the unique[3] real in $(0,1/2]$ such that

$$c(b,\delta) = H^{-1}\Big(\delta \cdot \log_2 \big(\frac{b}{b-1}\big)\Big). \tag{9.190}$$

The following theorem is the main result of this subsection. It is a quantitative refinement of Theorem 9.60 for the case of a single homogeneous tree.

Theorem 9.62. *Let $b \in \mathbb{N}$ with $b \geqslant 2$ and $0 < \delta \leqslant 1$. Also let T be a nonempty homogeneous tree of finite height and with branching number b. Then for every nonempty subset L of $\{0, \dots, h(T)-1\}$ and every $D \subseteq T$ satisfying*

$$\mathbb{E}_{n \in L}\, \mathrm{dens}_{T(n)}(D) \geqslant \delta$$

there exists a strong subtree S of T which is contained in D and satisfies $L_T(S) \subseteq L$ and $|h(S)| > c(b,\delta)|L|$ where $c(b,\delta)$ is as in (9.190).

In particular, for every positive integer ℓ we have

$$\mathrm{UDHL}(b\,|\,\ell,\delta) \leqslant \frac{\ell}{c(b,\delta)}. \tag{9.191}$$

For the proof of Theorem 9.62 we need to do some preparatory work. First, we will introduce two invariants which are associated with subsets of trees and are of independent interest. Specifically, let T be a nonempty tree of finite height and for every subset D of T we set

$$\mathcal{L}_T(D) = \{L_T(S) : S \text{ is a strong subtree of } T \text{ with } S \subseteq D\} \tag{9.192}$$

and

$$w_T(D) = \sum_{n=0}^{h(T)-1} \mathrm{dens}_{T(n)}(D). \tag{9.193}$$

(Notice that $|\mathcal{L}_T(D)| \geqslant 1$ since the empty tree is a strong subtree of any tree T.) The following lemma is due to Pach, Solymosi and Tardos [**PST**] and relates the quantities $|\mathcal{L}_T(D)|$ and $w_T(D)$. It is the main tool for the proof of Theorem 9.62.

Lemma 9.63. *Let T be a nonempty homogeneous tree of finite height and let b be the branching number of T. Then for every $D \subseteq T$ we have*

$$|\mathcal{L}_T(D)| \geqslant \Big(\frac{b}{b-1}\Big)^{w_T(D)}. \tag{9.194}$$

Proof. By induction on the height of the tree T. The initial case "$h(T) = 1$" is straightforward, and so let n be a positive integer and assume that the result has been proved for every nonempty homogeneous tree with height less than or equal to n. Fix a homogeneous tree T with $h(T) = n+1$ and let b be the branching

[3] Recall that the restriction of H on $[0,1/2]$ is strictly increasing and onto $[0,1]$, and so the constant $c(b,\delta)$ is well-defined.

number of T. Also let $\{t_1,\dots,t_b\}$ be an enumeration of the 1-level $T(1)$ of T. For every $i\in[b]$ set $T_i=\mathrm{Succ}_T(t_i)$ and $D_i=D\cap T_i$, and notice that

$$\mathcal{L}_T\Big(\bigcup_{i=1}^{b} D_i\Big)=\bigcup_{i=1}^{b}\mathcal{L}_T(D_i) \ \text{ and } \ w_T\Big(\bigcup_{i=1}^{b} D_i\Big)=\sum_{i=1}^{b} w_T(D_i). \tag{9.195}$$

Also observe that for every $i\in[b]$ we have: (i) $h(T_i)=n$, (ii) $|\mathcal{L}_{T_i}(D_i)|=|\mathcal{L}_T(D_i)|$, and (iii) $w_{T_i}(D_i)=b\cdot w_T(D_i)$. Hence, by our inductive assumptions, we have

$$|\mathcal{L}_T(D_i)|{=}|\mathcal{L}_{T_i}(D_i)|\geqslant\Big(\frac{b}{b-1}\Big)^{w_{T_i}(D_i)}=\Big(\frac{b}{b-1}\Big)^{b\cdot w_T(D_i)} \tag{9.196}$$

for every $i\in[b]$. Let $T(0)$ be the root of T and consider the following cases.

Case 1: *we have $T(0)\notin D$.* In this case we see that $D=D_1\cup\dots\cup D_b$ and $w_T(D)=w_T(D_1)+\dots+w_T(D_b)$. Therefore,

$$\begin{aligned}|\mathcal{L}_T(D)|\overset{(9.195)}{=}|\bigcup_{i=1}^{b}\mathcal{L}_T(D_i)| \quad &\geqslant \quad \max\big\{|\mathcal{L}_T(D_i)|:1\leqslant i\leqslant b\big\}\\ &\overset{(9.196)}{=} \quad \Big(\frac{b}{b-1}\Big)^{b\cdot\max\{w_T(D_i):1\leqslant i\leqslant b\}}\\ &\geqslant \quad \Big(\frac{b}{b-1}\Big)^{\sum_{i=1}^{b}w_T(D_i)}=\Big(\frac{b}{b-1}\Big)^{w_T(D)}\end{aligned}$$

as desired.

Case 2: *we have $T(0)\in D$.* Our assumption in this case implies that

$$|\mathcal{L}_T(D)|=|\bigcup_{i=1}^{b}\mathcal{L}_T(D_i)|+|\bigcap_{i=1}^{d}\mathcal{L}_T(D_i)| \tag{9.197}$$

and

$$w_T(D)=1+\sum_{i=1}^{b}w_T(D_i)=1+\frac{1}{b}\sum_{i=1}^{b}w_{T_i}(D_i). \tag{9.198}$$

We will need the following simple fact in order to estimate the size of $\mathcal{L}_T(D)$.

Fact 9.64. *Let b be a positive integer. Also let X be a set and let $\mathcal{L}_1,\dots,\mathcal{L}_b$ be subsets of X. Then we have*

$$\sum_{i=1}^{b}|\mathcal{L}_i|\leqslant(b-1)|\bigcup_{i=1}^{b}\mathcal{L}_i|+|\bigcap_{i=1}^{b}\mathcal{L}_i|. \tag{9.199}$$

Proof of Fact 9.64. We may assume, of course, that $b\geqslant 2$. Notice that

$$\begin{aligned}|\mathcal{L}_b|-|\bigcap_{i=1}^{b}\mathcal{L}_i| \quad &\leqslant \quad \sum_{i=1}^{b-1}|\mathcal{L}_b\setminus\mathcal{L}_i|\leqslant\sum_{i=1}^{b-1}|\bigcup_{j=1}^{b}\mathcal{L}_j\setminus\mathcal{L}_i|\\ &= \quad \sum_{i=1}^{b-1}\Big(|\bigcup_{j=1}^{b}\mathcal{L}_j|-|\mathcal{L}_i|\Big)=(b-1)|\bigcup_{i=1}^{b}\mathcal{L}_i|-\sum_{i=1}^{b-1}|\mathcal{L}_i|\end{aligned}$$

and the proof of Fact 9.64 is completed. □

Now, using the convexity of the function $f(x) = \big(\frac{b}{b-1}\big)^x$ and our inductive assumptions, we obtain that

$$\begin{aligned}\Big(\frac{b}{b-1}\Big)^{w_T(D)} &\overset{(9.198)}{=} \Big(\frac{b}{b-1}\Big)\cdot\Big(\frac{b}{b-1}\Big)^{b^{-1}\cdot\sum_{i=1}^{b} w_{T_i}(D_i)} \\ &\leqslant \frac{b}{b-1}\cdot\frac{1}{b}\cdot\sum_{i=1}^{b}\Big(\frac{b}{b-1}\Big)^{w_{T_i}(D_i)} \overset{(9.196)}{\leqslant} \frac{1}{b-1}\sum_{i=1}^{b}|\mathcal{L}_T(D_i)|\end{aligned}$$

and so, by Fact 9.64, we conclude that

$$\begin{aligned}\Big(\frac{b}{b-1}\Big)^{w_T(D)} &\leqslant |\bigcup_{i=1}^{b}\mathcal{L}_T(D_i)| + \frac{1}{b-1}\cdot|\bigcap_{i=1}^{b}\mathcal{L}_T(D_i)| \\ &\leqslant |\bigcup_{i=1}^{b}\mathcal{L}_T(D_i)| + |\bigcap_{i=1}^{b}\mathcal{L}_T(D_i)| \overset{(9.197)}{=} |\mathcal{L}_T(D)|.\end{aligned}$$

The above cases are exhaustive and the proof of Lemma 9.63 is completed. □

We are ready to give the proof of Theorem 9.62.

Proof of Theorem 9.62. We fix a nonempty subset L of $\{0,\dots,h(T)-1\}$ and a subset D of T such that $\mathbb{E}_{n\in L}\,\mathrm{dens}_{T(n)}(D)\geqslant\delta$. Clearly, we may assume that D is contained in $\bigcup_{n\in L}T(n)$, and in particular, that the set $\mathcal{L}_T(D)$ is a collection of subsets of L. Next observe that

$$w_T(D) \overset{(9.193)}{=} \big(\mathbb{E}_{n\in L}\,\mathrm{dens}_{T(n)}(D)\big)\cdot|L| \geqslant \delta|L|$$

and so, by Lemma 9.63,

$$\begin{aligned}\log_2\big(|\mathcal{L}_T(D)|\big) &\overset{(9.194)}{\geqslant} \log_2\Big(\frac{b}{b-1}\Big)\cdot w_T(D) \\ &\geqslant \Big(\delta\cdot\log_2\big(\frac{b}{b-1}\big)\Big)\cdot|L| \overset{(9.190)}{=} H\big(c(b,\delta)\big)\cdot|L|.\end{aligned}$$

This estimate and Lemma E.7 yield that

$$|\mathcal{L}_T(D)| > \sum_{i=0}^{\lfloor c(b,\delta)\cdot|L|\rfloor}\binom{|L|}{i}.$$

It follows that the family $\mathcal{L}_T(D)$ must contain a subset of L of cardinality at least $\lfloor c(b,\delta)\cdot|L|\rfloor+1$ and the proof of Theorem 9.62 is completed. □

We close this subsection with the following "parameterized" version of Szemerédi's theorem due to Furstenberg and Weiss [**FW**].

Theorem 9.65. *Let b,ℓ be a pair of integers with $b,\ell\geqslant 2$ and $0<\delta\leqslant 1$. Also let T be a homogeneous tree of finite height, with branching number b and such that $h(T)\geqslant \mathrm{Sz}\big(\ell, c(b,\delta)\big)$ where $c(b,\delta)$ is as in (9.190). Then every $D\subseteq T$ satisfying*

$$\mathbb{E}_{n\in\{0,\dots,h(T)-1\}}\,\mathrm{dens}_{T(n)}(D)\geqslant\delta \tag{9.200}$$

contains a strong subtree S of T of height ℓ whose level set is an arithmetic progression.

For the proof of Theorem 9.65 we need the following fact. It follows from basic properties of strong subtrees.

FACT 9.66. *Let T be a nonempty finite tree of finite height. Also let R be a strong subtree of T and $L' \subseteq L_T(R)$. Then there exists a strong subtree S of R such that $L_T(R) = L'$. In particular, for every $D \subseteq T$ the family $\mathcal{L}_T(D)$ is hereditary, that is, for every $L \in \mathcal{L}_T(D)$ and every $L' \subseteq L$ we have $L' \in \mathcal{L}_T(D)$.*

We proceed to the proof of Theorem 9.65.

PROOF OF THEOREM 9.65. We follow the proof from [**PST**]. We fix $D \subseteq T$ satisfying (9.200). By Theorem 9.62 applied for "$L = \{0, \dots, h(T) - 1\}$", there exists a strong subtree R of T with $R \subseteq D$ and such that $h(R) \geqslant c(b,\delta)h(T)$. In particular, we have $L_T(R) \subseteq \{0, \dots, h(T) - 1\}$ and $|L_T(R)| \geqslant c(b,\delta)h(T)$. Since $h(T) \geqslant \mathrm{Sz}\big(\ell, c(b,\delta)\big)$, by Szemerédi's theorem, the set $L_T(R)$ contains an arithmetic progression P of length ℓ. Noticing that $L_T(R) \in \mathcal{L}_T(D)$, by Fact 9.66, we conclude that there exists $S \in \mathcal{L}_T(D)$ with $L_T(S) = P$. Clearly, S is as desired. The proof of Theorem 9.65 is completed. □

9.8.3. Patterns. Our goal in this subsection is to prove an extension of Theorem 9.2 which does not refer to left variable words but to a wider classes of variable words. Specifically, let A be a finite alphabet with $|A| \geqslant 2$ and let u, p be two variable words over A. We say that u is of *pattern* p if p is an initial segment of u. (Notice that if $p = (v)$, then u is of pattern p if and only if u is a left variable word.) More generally, given two nonempty finite sequences $(u_n)_{n=0}^{m-1}$ and $(p_n)_{n=0}^{m-1}$ of variable words over A, we say that $(u_n)_{n=0}^{m-1}$ is of *pattern* $(p_n)_{n=0}^{m-1}$ if p_n is an initial segment of u_n for every $n \in \{0, \dots, m-1\}$. We have the following theorem.

THEOREM 9.67. *Let m, r be positive integers and $0 < \delta \leqslant 1$. Also let A be a finite alphabet with $|A| \geqslant 2$ and L a finite subset of $\mathbb{N}$ with*

$$|L| \geqslant \lceil 2r\delta^{-1}\mathrm{DCS}(|A^r|, m, \delta/2)\rceil. \tag{9.201}$$

If D is a subset of $A^{<\mathbb{N}}$ with $|D \cap A^\ell| \geqslant \delta|A^\ell|$ for every $\ell \in L$, then for every finite sequence $(p_n)_{n=0}^{m-1}$ of variable words over A with $\max\{|p_n| : 0 \leqslant n < m\} \leqslant r$ there exist a word w over A and a finite sequence $(u_n)_{n=0}^{m-1}$ of variable words over A of pattern $(p_n)_{n=0}^{m-1}$ such that the set

$$\{w\} \cup \big\{w^\frown u_0(a_0)^\frown \dots ^\frown u_n(a_n) : n \in \{0, \dots, m-1\} \text{ and } a_0, \dots, a_n \in A\big\}$$

is contained in D.

We will give a proof of Theorem 9.67 which is a variant of the second proof of the multidimensional Hales–Jewett theorem presented in Section 2.2 and relies on an application of Theorem 9.2 for an appropriately chosen finite Cartesian product of A. We start with the following definition.

DEFINITION 9.68. *Let A be a finite alphabet with $|A| \geqslant 2$ and r a positive integer. We set $B = A^r$ and we define $\Phi_r \colon B^{<\mathbb{N}} \to A^{<\mathbb{N}}$ by setting $\Phi_r(\emptyset) = \emptyset$ and*

$$\Phi_r\big((\beta_0, \dots, \beta_{n-1})\big) = \beta_0{}^\frown \dots {}^\frown \beta_{n-1}$$

for every $n \geqslant 1$ *and every* $(\beta_0, \dots, \beta_{n-1}) \in B^n$*. Moreover, for every* $t \in A^{<\mathbb{N}}$ *let* $\Phi_{t,r} \colon B^{<\mathbb{N}} \to A^{<\mathbb{N}}$ *be defined by* $\Phi_{t,r}(s) = t^\frown \Phi_r(s)$ *for every* $s \in B^{<\mathbb{N}}$*.*

In the following fact we collect some basic properties of the maps Φ_r and $\Phi_{t,r}$. They are all straightforward consequences of the relevant definitions.

FACT 9.69. *Let* A *be a finite alphabet with* $|A| \geqslant 2$ *and* $t \in A^{<\mathbb{N}}$*. Also let* r *be a positive integer and set* $B = A^r$*. Then the following hold.*

(a) *If* $t = \emptyset$*, then* $\Phi_{t,r} = \Phi_r$*.*
(b) *For every* $v_1, v_2 \in B^{<\mathbb{N}}$ *we have*

$$\Phi_{t,r}(v_1^\frown v_2) = \Phi_{t,r}(v_1)^\frown \Phi_r(v_2). \tag{9.202}$$

(c) *The map* $\Phi_{t,r}$ *is an injection and satisfies* $\Phi_{t,r}\big(B^n\big) = \{t^\frown s : s \in A^{r\cdot n}\}$ *for every* $n \in \mathbb{N}$*. In particular, for every* $D \subseteq A^{<\mathbb{N}}$ *and every* $n \in \mathbb{N}$

$$\mathrm{dens}_{B^n}\big(\Phi_{t,r}^{-1}(D)\big) = \mathrm{dens}_{A^{n\cdot r}}(D_t)$$

where $D_t = \{s \in A^{<\mathbb{N}} : t^\frown s \in D\}$*.*

(d) *Let* v *be a left variable word over* B *and let* p *be a variable word over* A *with* $|p| = r$*. Then there exists a variable word* w *over* A *of pattern* p *such that* $w(a) = \Phi_r\big(v(p(a))\big)$ *for every* $a \in A$*.*

We proceed with the following lemma.

LEMMA 9.70. *Let* r *be a positive integer and* $0 < \delta \leqslant 1$*. Also let* A *be a finite alphabet with* $|A| \geqslant 2$*,* L *a nonempty finite subset of* $\mathbb{N}$ *and* $D \subseteq A^{<\mathbb{N}}$ *such that* $\mathrm{dens}_{A^\ell}(D) \geqslant \delta$ *for every* $\ell \in L$*. Then there exist an integer* $i_0 \in \{0, \dots, r-1\}$*, a word* $t_0 \in A^{i_0}$ *and a finite subset* M *of* $\mathbb{N}$ *with* $|M| \geqslant (2r)^{-1}\delta|L|$ *such that, setting* $B = A^r$*, we have* $\mathrm{dens}_{B^m}\big(\Phi_{t_0,r}^{-1}(D)\big) \geqslant \delta/2$ *for every* $m \in M$*.*

PROOF. For every $i \in \{0, \dots, r-1\}$ we set $L_i = \{\ell \in L : \ell = i \bmod r\}$ and we select $i_0 \in \{0, \dots, r-1\}$ such that $|L_{i_0}| \geqslant |L|/r$. Next, for every $t \in A^{i_0}$ we set $L_{i_0}^t = \{\ell \in L_{i_0} : \mathrm{dens}_{A^{\ell - i_0}}(D_t) \geqslant \delta/2\}$ where $D_t = \{s \in A^{<\mathbb{N}} : t^\frown s \in D\}$. Let $\ell \in L_{i_0}$ be arbitrary. Notice that $\mathbb{E}_{t \in A^{i_0}} \mathrm{dens}_{A^{\ell-i_0}}(D_t) = \mathrm{dens}_{A^\ell}(D) \geqslant \delta$ and so, by Markov's inequality, we have

$$\mathrm{dens}\big(\{t \in A^{i_0} : \mathrm{dens}_{A^{\ell-i_0}}(D_t) \geqslant \delta/2\}\big) \geqslant \delta/2.$$

Hence, by Lemma E.4, there exists $t_0 \in A^{i_0}$ such that $|L_{i_0}^{t_0}| \geqslant (\delta/2)|L_{i_0}|$. We set $M = \{m \in \mathbb{N} : mr + i_0 \in L_{i_0}^{t_0}\}$ and we observe that $|M| = |L_{i_0}^{t_0}| \geqslant (2r)^{-1}\delta|L|$. Finally, let $m \in M$ and set $\ell = mr + i_0 \in L_{i_0}^{t_0}$. By Fact 9.69, we see that $\mathrm{dens}_{B^m}\big(\Phi_{t_0,r}^{-1}(D)\big) = \mathrm{dens}_{A^{m\cdot r}}(D_{t_0}) = \mathrm{dens}_{A^{\ell-i_0}}(D_{t_0}) \geqslant \delta/2$ and the proof of Lemma 9.70 is completed. □

We are ready to give the proof of Theorem 9.67.

PROOF OF THEOREM 9.67. Fix a pair m, r of positive integers and $0 < \delta \leqslant 1$. Let A be a finite alphabet with $|A| \geqslant 2$ and L a finite subset of $\mathbb{N}$ satisfying (9.201). Also let $D \subseteq A^{<\mathbb{N}}$ such that $\mathrm{dens}_{A^\ell}(D) \geqslant \delta$ for every $\ell \in L$. We set $B = A^r$. By

Lemma 9.70 and (9.201), there exist an integer $i_0 \in \{0,\dots,r-1\}$, a word $t_0 \in A^{i_0}$ and a subset M of $\mathbb{N}$ with

$$|M| \geqslant (2r)^{-1}\delta|L| \geqslant \mathrm{DCS}(|A^r|, m, \delta/2)$$

and such that $\mathrm{dens}_{B^m}\big(\Phi^{-1}_{t_0,r}(D)\big) \geqslant \delta/2$ for every $m \in M$. Thus, by Theorem 9.2, there exists an m-dimensional Carlson–Simpson space V of $B^{<\mathbb{N}}$ with $V \subseteq \Phi^{-1}_{t_0,r}(D)$. Let $\langle s, (v_n)_{n=0}^{m-1}\rangle$ be the Carlson–Simpson system generating V. Also let $(p_n)_{n=0}^{m-1}$ be a finite sequence of variable words over A with $\max\{|p_n| : 0 \leqslant n \leqslant m-1\} \leqslant r$. Fix $\alpha \in A$ and for every $n \in \{0,\dots,m-1\}$ set

$$p'_n = p_n{}^\frown \alpha^{r-|p_n|}. \tag{9.203}$$

(Here, $\alpha^{r-|p_n|}$ is as in (2.1).) Notice that each p'_n is a variable word over A of length r and of pattern p_n. By Fact 9.69, there exists a finite sequence $(u_n)_{n=0}^{m-1}$ of variable words over A of pattern $(p'_n)_{n=0}^{m-1}$ such that

$$u_n(a) = \Phi_r\big(v_n(p'_n(a))\big) \tag{9.204}$$

for every $n \in \{0,\dots,m-1\}$ and every $a \in A$. By (9.203), we see that $(u_n)_{n=0}^{m-1}$ is of pattern $(p_n)_{n=0}^{m-1}$. Next, set $w = \Phi_{t_0,r}(s)$ and observe that $w = \Phi_{t_0,r}\big(V(0)\big)$. Moreover, for every $n \in \{0,\dots,m-1\}$ and every $a_0,\dots,a_n \in A$ we have

$$\begin{aligned} w^\frown u_0(a_0)^\frown\dots^\frown u_n(a_n) &\overset{(9.204)}{=} \Phi_{t_0,r}(s)^\frown\Phi_r(v_0(p'_0(a_0)))^\frown\dots^\frown\Phi_r\big(v_n(p'_n(a_n))\big) \\ &\overset{(9.202)}{=} \Phi_{t_0,r}\big(s^\frown v_0(p'_0(a_0))^\frown\dots^\frown v_n(p'_n(a_n))\big) \end{aligned}$$

which implies that $w^\frown u_0(a_0)^\frown\dots^\frown u_n(a_n) \in \Phi_{t_0,r}\big(V(n)\big)$. Since $V \subseteq \Phi^{-1}_{t_0,r}(D)$, we conclude that the set

$$\{w\} \cup \big\{w^\frown u_0(a_0)^\frown\dots^\frown u_n(a_n) : n \in \{0,\dots,m-1\} \text{ and } a_0,\dots,a_n \in A\big\}$$

is contained in D and the proof of Theorem 9.67 is completed. □

9.9. Notes and remarks

9.9.1. The material in Sections 9.1 up to 9.7 is taken from [**DKT3**]. We notice, however, that the class of Furstenberg–Weiss measures, introduced in Definition 9.19, appeared first in [**FW**] in a slightly less general form.

9.9.2. As we have already mentioned, Proposition 9.21 is the analogue of Proposition 8.7. In this direction, we also have the following extension of Proposition 9.21 in the spirit of Theorem 8.21.

Theorem 9.71. *For every positive integer p and every $0 < \delta \leqslant 1$ there exists a strictly positive constant $\Theta(p,\delta)$ with the following property. If k, m are positive integers with $k \geqslant 2$, then there exists a positive integer* $\mathrm{CorCS}(k,m,\delta)$ *such that for every alphabet A with $|A| = k$, every Carlson–Simpson space T of $A^{<\mathbb{N}}$ with* $\dim(T) \geqslant \mathrm{CorCS}(k,m,\delta)$ *and every family $\{D_t : t \in T\}$ of measurable events in a probability space (Ω,Σ,μ) satisfying $\mu(D_t) \geqslant \delta$ for every $t \in T$, there exists an*

m-dimensional Carlson–Simpson subspace S of T such that for every nonempty subset F of S we have

$$\mu\Big(\bigcap_{t\in F} D_t\Big) \geqslant \Theta(|F|,\delta). \tag{9.205}$$

Theorem 9.71 is based on Theorems 5.11 and 9.2. It follows the same strategy as the proof of Theorem 8.21, though the argument for the case of Carlson–Simpson spaces is somewhat more involved (see [**DKT4**] for details).

9.9.3. Theorem 9.61 is a density version of the Halpern–Läuchli theorem and was conjectured by Laver in the late 1960s. We also point out that the assumption in Theorem 9.61 that the trees are homogeneous is essentially optimal (see [**BV**]).

9.9.4. We remark that Theorem 9.67 has an infinite-dimensional version which extends Theorem 9.1 to wider classes of sequences of variable words. This extension can be found in [**DKT3**].

APPENDIX A

Primitive recursive functions

Throughout this appendix we deal with *number theoretic functions*, that is, functions from a nonempty finite Cartesian product of the natural numbers to the natural numbers. If $f\colon \mathbb{N}^k \to \mathbb{N}$ is a number theoretic function, then the (unique) positive integer k is called the *arity* of f. Two simple examples of number theoretic functions, which are relevant to our discussion, are the *successor function* $S\colon \mathbb{N} \to \mathbb{N}$ defined by $S(n) = n + 1$, and the *projection functions* $P^k_i\colon \mathbb{N}^k \to \mathbb{N}$ $(1 \leqslant i \leqslant k)$ defined by $P^k_i(x_1, \dots, x_k) = x_i$.

Let g, h and f be number theoretic functions of arities k, $k+2$ and $k+1$ respectively. Recall that f is said to be defined by *primitive recursion* from g and h provided that for every $x \in \mathbb{N}^k$ and every $n \in \mathbb{N}$ we have

$$\begin{cases} f(0, x) = g(x), \\ f(n+1, x) = h(f(n, x), n, x). \end{cases} \tag{A.1}$$

There is a simpler kind of primitive recursion appropriate for defining unary functions, namely

$$\begin{cases} f(0) = m, \\ f(n+1) = h(f(n), n) \end{cases} \tag{A.2}$$

where $m \in \mathbb{N}$ and $h\colon \mathbb{N}^2 \to \mathbb{N}$. We will include this simpler scheme when we talk of definition by primitive recursion.

Also recall that if ψ is a number theoretic function of arity k and $g_1, \dots, g_k$ are number theoretic functions all of arity m, then the *composition* of ψ with $g_1, \dots, g_k$ is the function ϕ of arity m defined by $\phi(y) = \psi\big(g_1(y), \dots, g_k(y)\big)$ where y varies over $\mathbb{N}^m$.

Definition A.1. *The class of* primitive recursive functions *is the smallest set of number theoretic functions that contains the constant zero function, the successor function and the projection functions, and is closed under composition and primitive recursion.*

We will not need the fine structure of primitive recursive functions, only their basic properties. They will be used as effective tools in order to estimate the growth of number theoretic functions coming from various inductive arguments. This point of view is very convenient from a Ramsey theoretic perspective, especially when combined with a natural hierarchy of primitive recursive functions, introduced by Grzegorczyk [**Grz**], which we are about to recall.

Let E_0 and E_1 be the number theoretic functions defined by $E_0(x,y)=x+y$ and $E_1(x)=x^2+2$ (thus E_0 is binary while E_1 is unary). Next, for every $n\in\mathbb{N}$ let E_{n+2} be the unary number theoretic function defined recursively by the rule

$$\begin{cases} E_{n+2}(0)=2, \\ E_{n+2}(x+1)=E_{n+1}\big(E_{n+2}(x)\big). \end{cases} \tag{A.3}$$

Observe that each E_n is primitive recursive. Also notice that for every $n\geqslant 1$ the function E_n is increasing.

Definition A.2. *For every $n\in\mathbb{N}$ the* Grzegorczyk's class $\mathcal{E}^n$ *is the smallest set of number theoretic functions that contains the functions E_k for $k<n$, the constant zero function, the successor function and the projection functions, and is closed under composition and limited primitive recursion (that is, if $g,h,j\in\mathcal{E}^n$ and f is defined by primitive recursion from g and h, has the same arity as j and is pointwise bounded by j, then f belongs to $\mathcal{E}^n$).*

As we have already indicated, much of our interest in Grzegorczyk's classes stems from the fact that they possess strong stability properties. We gather, below, these properties that are used throughout this book. For a proof, as well as for a detailed exposition of this material, we refer to [**Ros**].

Proposition A.3. *The following hold.*

(a) *For every $n\in\mathbb{N}$ we have $\mathcal{E}^n\subseteq\mathcal{E}^{n+1}$. Moreover, a number theoretic function f is primitive recursive if and only if $f\in\mathcal{E}^n$ for some $n\in\mathbb{N}$.*
(b) *If $g,h\in\mathcal{E}^n$ for some $n\in\mathbb{N}$ and f is defined by primitive recursion from g and h, then $f\in\mathcal{E}^{n+1}$.*
(c) *For every integer $n\geqslant 2$ and every $f\in\mathcal{E}^n$ there exists $m\in\mathbb{N}$ such that $f(x_1,\dots,x_k)\leqslant E^{(m)}_{n-1}(\max\{x_1,\dots,x_k\})$ where k is the arity of f and $x_1,\dots,x_k$ vary over $\mathbb{N}$.*

One consequence of Proposition A.3 is that every unary function in the class $\mathcal{E}^n$ ($n\in\mathbb{N}$) is majorized by a unary increasing function also belonging to $\mathcal{E}^n$. More generally we have the following corollary.

Corollary A.4. *For every $n\in\mathbb{N}$ and every $f\in\mathcal{E}^n$ of arity k there exists $F\in\mathcal{E}^n$ of arity k which dominates f pointwise and satisfies*

$$F(x_1,\dots,x_k)\leqslant F(y_1,\dots,y_k) \tag{A.4}$$

for every $x_1,\dots,x_k,y_1,\dots,y_k\in\mathbb{N}$ with $x_i\leqslant y_i$ for all $i\in[k]$.

In light of Corollary A.4 we may assume that all primitive recursive functions we are dealing with satisfy the monotonicity property described in (A.4). We will follow this assumption throughout this book, sometimes without giving an explicit reference to Corollary A.4.

APPENDIX B

Ramsey's theorem

For every triple d, m, r of positive integers with $d \geqslant m$ let $\mathrm{R}(d,m,r)$ be the *Ramsey number* for the parameters d, m, r, that is, the least integer $n \geqslant d$ such that for every n-element set X and every r-coloring of $\binom{X}{m}$ there exists $Z \in \binom{X}{d}$ such that the set $\binom{Z}{m}$ is monochromatic. The existence of the numbers $\mathrm{R}(d,m,r)$ for every choice of admissible parameters is, of course, the content of Ramsey's famous theorem [**Ra**].

It is obvious that $\mathrm{R}(d,m,1) = d$ for every $d \geqslant m \geqslant 1$. The case "$m = 1$" is a consequence of the classical pigeonhole principle. Indeed, notice that

$$\mathrm{R}(d,1,r) = r(d-1)+1 \tag{B.1}$$

for every $d \geqslant 1$ and every $r \geqslant 1$. Our goal in this appendix is to present the proof of the following general estimate for the Ramsey numbers, essentially due to Erdős and Rado [**ER**].

Theorem B.1. *For every triple d, m, r of positive integers with $d \geqslant m+1$ and $r \geqslant 2$ we have*

$$\mathrm{R}(d,m+1,r) \leqslant m-1+r^{\binom{\mathrm{R}(d,m,r)-1}{m}}. \tag{B.2}$$

In particular, the numbers $\mathrm{R}(d,m,r)$ are upper bounded by a primitive recursive functions belonging to the class $\mathcal{E}^4$.

The proof of Theorem B.1 follows the scheme we discussed in Section 2.1. It is based on the following lemma which will enable us to reduce a finite coloring of $\binom{X}{m+1}$ to a "simpler" one.

Lemma B.2. *Let ℓ, m, r be positive integers with $\ell \geqslant m+1$ and $r \geqslant 2$. Also let X be a subset of $\mathbb{N}$ and $c\colon \binom{X}{m+1} \to [r]$, and assume that*

$$|X| = m-1+r^{\binom{\ell-1}{m}}. \tag{B.3}$$

Then there exists $Y \in \binom{X}{\ell}$ such that $c(F) = c(G)$ for every $F, G \in \binom{Y}{m+1}$ satisfying $F \setminus \{\max(F)\} = G \setminus \{\max(G)\}$.

Proof. If $m \geqslant 2$, then let $Y_0 = \{x_0, \dots, x_{m-2}\}$ be the subset of X consisting of the first $m-1$ elements of X; otherwise, let $Y_0 = \emptyset$. Also set $X_{m-1} = X \setminus Y_0$ and $x_{m-1} = \min(X_{m-1})$. By (B.3), we have $|X_{m-1}| = r^{\binom{\ell-1}{m}}$. Recursively, we will select a decreasing sequence $X_m \supseteq \cdots \supseteq X_{\ell-1}$ of subsets of X_{m-1} and an increasing sequence $x_m < \cdots < x_{\ell-1}$ of elements of X such that the following conditions are satisfied.

(C1) For every $i \in \{m, \dots, \ell-1\}$ we have $|X_i| = r^{\binom{\ell-1}{m} - \binom{i}{m}}$ and $x_i \in X_i$.

(C2) For every $i \in \{m, \dots, \ell-1\}$ we have $X_i \subseteq X_{i-1} \setminus \{x_{i-1}\}$. Moreover, for every $G \subseteq \{x_0, \dots, x_{i-2}\}$ with $|G| = m-1$ and every $z_1, z_2 \in X_i$ we have
$$c\big(G \cup \{x_{i-1}\} \cup \{z_1\}\big) = c\big(G \cup \{x_{i-1}\} \cup \{z_2\}\big).$$

The first step is identical to the general one, and so let $i \in \{m, \dots, \ell-2\}$ and assume that the sets $X_m, \dots, X_i$ and the elements $x_m, \dots, x_i$ have been selected so that the above conditions are satisfied. We set
$$\mathcal{G} = \binom{\{x_0, \dots, x_{i-1}\}}{m-1}$$
and we define a coloring $C \colon X_i \setminus \{x_i\} \to [r]^{\mathcal{G}}$ by the rule
$$C(z) = \big\langle c\big(G \cup \{x_i\} \cup \{z\}\big) : G \in \mathcal{G}\big\rangle.$$
Since $r \geqslant 2$ and $|X_i| = r^{\binom{\ell-1}{m} - \binom{i}{m}}$, by the classical pigeonhole principle, there exists a subset Z of $X_i \setminus \{x_i\}$ which is monochromatic with respect to C and satisfies
$$|Z| \geqslant \left\lceil \frac{|X_i| - 1}{r^{\binom{i}{m-1}}} \right\rceil \geqslant r^{\binom{\ell-1}{m} - \binom{i}{m} - \binom{i}{m-1}} = r^{\binom{\ell-1}{m} - \binom{i+1}{m}}.$$
We set $X_{i+1} = Z$ and $x_{i+1} = \min(X_{i+1})$ and we observe that conditions (C1) and (C2) are satisfied. The recursive selection is thus completed.

Finally, we set $Y = Y_0 \cup \{x_{m-1}, \dots, x_{\ell-1}\} = \{x_0, \dots, x_{\ell-1}\}$ and we claim that this set satisfies the requirements of the lemma. Indeed, first observe that $|Y| = \ell$. Also notice that for every $F \in \binom{Y}{m+1}$ there exists $i \in \{m, \dots, \ell-1\}$ such that F is written as $G \cup \{x_{i-1}\} \cup \{z\}$ where G is the set of the first $m-1$ elements of F and $z = \max(F) \in X_i$. Using this remark and condition (C2), we see that Y is as desired. The proof of lemma B.2 is completed. □

We are ready to give the proof of Theorem B.1.

Proof of Theorem B.1. The estimate in (B.2) is an immediate consequence of Lemma B.2. On the other hand, the fact that the Ramsey numbers are upper bounded by a function belonging to the class $\mathcal{E}^4$ follows by (B.1), (B.2) and elementary properties of primitive recursive functions. The proof of Theorem B.1 is completed. □

APPENDIX C

The Baire property

We recall the following classical topological notions.

Definition C.1. *Let (X,τ) be a topological space. A subset N of X is said to be* nowhere dense *if its closure has empty interior. A subset M of X is called* meager *if it is the countable union of nowhere dense subsets of X. Finally, a subset A of X is said to have the* Baire property *if it is equal to an open set modulo a meager set, that is, if there exists an open subset U of X such that the symmetric difference*

$$A \triangle U = (A \setminus U) \cup (U \setminus A)$$

of A and U is meager.

Notice that the collection of all subsets of a topological space with the Baire property is a σ-algebra and contains all open and all meager sets. In fact, we have the following finer information (see, e.g., [**Ke**]).

Proposition C.2. *The class of sets with the Baire property of a topological space (X,τ) is the smallest σ-algebra on X containing all open and all meager sets.*

We proceed to discuss yet another important closure property of the class of sets with the Baire property. To this end we recall some definitions.

Let X be a set. A *Souslin scheme on X* is a collection $\langle F_s : s \in \mathbb{N}^{<\mathbb{N}} \rangle$ of subsets of X indexed by the set $\mathbb{N}^{<\mathbb{N}}$ of all finite sequences in $\mathbb{N}$. The *Souslin operation* applied to such a scheme produces the set

$$\mathcal{A}_s F_s = \bigcup_{x\in\mathbb{N}^{\mathbb{N}}} \bigcap_{n\in\mathbb{N}} F_{x\upharpoonright n}. \tag{C.1}$$

A family $\mathcal{F}$ of subsets of X is said to be *closed under the Souslin operation* if $\mathcal{A}_s F_s \in \mathcal{F}$ for every Souslin scheme $\langle F_s : s \in \mathbb{N}^{<\mathbb{N}} \rangle$ with $F_s \in \mathcal{F}$ for every $s \in \mathbb{N}^{<\mathbb{N}}$.

The Souslin operation is of fundamental importance in classical descriptive set theory. Its relation with the Baire property is described in the following theorem (see, e.g., [**Ke**]).

Theorem C.3. *The class of sets with the Baire property of a topological space (X,τ) is closed under the Souslin operation.*

Note that, as opposed to Proposition C.2, the above result does not characterize the family of all sets with the Baire property. Precisely, if (X,τ) is a topological space, then the smallest σ-algebra on X containing all open sets and closed under the Souslin operation may be strictly smaller than the class of all sets with the Baire property. This motivates the following definition.

DEFINITION C.4. *Let (X, τ) be a topological space. A subset of X is said to be a C-*set *if it belongs to the smallest σ-algebra on X containing all open sets and closed under the Souslin operation.*

The family of C-sets is quite extensive and contains most sets that appear in mathematical practice. Also it enjoys, by its very definition, all stability properties of the class of sets with the Baire property. We close this appendix by noting another important property of the family of C-sets (see, e.g., [**Ke**]).

PROPOSITION C.5. *Let X and Y be topological spaces and $f\colon X \to Y$ a Borel measurable function. If $A \subseteq Y$ is a C-set, then so is $f^{-1}(A)$.*

APPENDIX D

Ultrafilters

D.1. Definitions. We recall the following notion.

Definition D.1. *Let X be a nonempty set. An* ultrafilter on X *is a family* $\mathcal{U}$ *of subsets of X which satisfies the following properties.*

(U1) *The empty set does not belong to $\mathcal{U}$.*
(U2) *If $A \in \mathcal{U}$ and $A \subseteq B \subseteq X$, then $B \in \mathcal{U}$.*
(U3) *If $A, B \in \mathcal{U}$, then $A \cap B \in \mathcal{U}$.*
(U4) *For every $A \subseteq X$ we have that either $A \in \mathcal{U}$ or $X \setminus A \in \mathcal{U}$.*

The set of all ultrafilters on X will be denoted by βX.

A family of subsets of X satisfying (U1)–(U3) is called a *filter* on X. By (U4), we see that every ultrafilter on X is, in fact, a maximal filter on X. An ultrafilter $\mathcal{U}$ on X is called *principal* if there exists $x \in X$ such that $\mathcal{U} = \{A \subseteq X : x \in A\}$. The existence of non-principal ultrafilters on an infinite set X is a straightforward consequence of Zorn's lemma. More generally, we have the following fact.

Fact D.2. *Let X be an infinite set. Then the following hold.*

(a) *Every family $\mathcal{F}$ of nonempty subsets of X with the finite intersection property (that is, every finite subfamily of $\mathcal{F}$ has a nonempty intersection) is extended to an ultrafilter on X.*
(b) *Let $\mathcal{F} = \{F : X \setminus F$ is finite$\}$ be the family of all cofinite subsets of X. Then $\mathcal{F}$ is extended to a non-principal ultrafilter on X. Conversely, every non-principal ultrafilter on X contains the family $\mathcal{F}$.*

In the rest of this appendix we review some basic properties of the space βX. A more complete treatment of this material can be found in [**HS, B2, To**].

D.2. The topology of βX. Let X be a nonempty set. For every $A \subseteq X$ let

$$(A)_{\beta X} = \{\mathcal{U} \in \beta X : A \in \mathcal{U}\}. \tag{D.1}$$

Also let $e_X : X \to \beta X$ be defined by $e_X(x) = \{A \subseteq X : x \in A\}$. We have the following proposition.

Proposition D.3. *Let X be a nonempty set and set $\mathcal{B} = \{(A)_{\beta X} : A \subseteq X\}$. Then $\mathcal{B}$ is a basis for a Hausdorff topology on βX with the following properties.*

(a) *For every $A \subseteq X$ the set $e_X(A)$ is a dense open subset of $(A)_{\beta X}$. In particular, $e_X(X)$ is a dense open subset of βX.*
(b) *The topological space βX is compact.*

PROOF. Notice that $(X)_{\beta X} = \beta X$. Moreover, $\beta X \setminus (A)_{\beta X} = (X \setminus A)_{\beta X}$ and $(A)_{\beta X} \cap (B)_{\beta X} = (A \cap B)_{\beta X}$ for every $A, B \subseteq X$. It follows, in particular, that the family $\mathcal{B}$ is a basis for a topology on βX consisting of clopen sets. To see that this topology is Hausdorff let $\mathcal{U}, \mathcal{W} \in \beta X$ with $\mathcal{U} \neq \mathcal{W}$. By the maximality of $\mathcal{U}$, there exists $A \in \mathcal{U} \setminus \mathcal{W}$. Hence, $\mathcal{U} \in (A)_{\beta X}$ and $\mathcal{W} \in (X \setminus A)_{\beta X}$.

Next observe that for every $x \in X$ we have $(\{x\})_{\beta X} = \{e_X(x)\}$. This implies that the singleton $\{e_X(x)\}$ is a basic open set, and as a consequence, the set $e_X(A) = \bigcup_{x\in A}\{e_X(x)\}$ is open for every $A \subseteq X$. We will show that the set $e_X(A)$ is dense in $(A)_{\beta X}$. Indeed, let $\mathcal{U} \in (A)_{\beta X}$ be arbitrary. Also let $(B)_{\beta X}$ be a basic open neighborhood of $\mathcal{U}$. Then $A, B \in \mathcal{U}$ and so $A\cap B \in \mathcal{U}$. In particular, $A\cap B \neq \emptyset$ which implies that $(B)_{\beta X} \cap e_X(A) = e_X(A \cap B) \neq \emptyset$.

It remains to show that the space βX is compact. To this end it is enough to show that every cover of βX by a family of basic sets has a finite subcover. So let $\mathcal{G}$ be a family of subsets of X such that $\beta X = \bigcup_{A\in\mathcal{G}}(A)_{\beta X}$. Assume, towards a contradiction, that there is no finite subfamily $\mathcal{H}$ of $\mathcal{G}$ such that $\beta X = \bigcup_{A\in\mathcal{H}}(A)_{\beta X}$. Taking complements and using the identity $\beta X \setminus (A)_{\beta X} = (X \setminus A)_{\beta X}$, we see that the family $\mathcal{F} = \{X \setminus A : A \in \mathcal{G}\}$ has the finite intersection property. By Fact D.2, there exists an ultrafilter $\mathcal{U}$ on X such that $\mathcal{F} \subseteq \mathcal{U}$. It follows that $\mathcal{U} \notin \bigcup_{A\in\mathcal{G}}(A)_{\beta X}$, a contradiction. The proof of Proposition D.3 is thus completed. $\square$

In what follows all topological properties of the space βX will refer to the topology described in Proposition D.3. Moreover, we will identify X with the set of all principal ultrafilters on X via the map $X \ni x \mapsto e_X(x) \in \beta X$. Having this identification in mind, for every $A \subseteq X$ we will write $\mathrm{C}\ell_{\beta X}(A)$ to denote the closure of the set $e_X(A)$ in βX. Notice that, by Proposition D.3,

$$\mathrm{C}\ell_{\beta X}(A) = (A)_{\beta X} \tag{D.2}$$

for every $A \subseteq X$.

It turns out that the space βX is homeomorphic to the Stone–Čech compactification of the set X equipped with the discrete topology. In particular, the space βX satisfies the following universal property.

PROPOSITION D.4. *Let X be a nonempty set. Also let Y be a nonempty set and $f\colon X \to Y$. Then the function f has a unique continuous extension $\widetilde{f}\colon \beta X \to \beta Y$ which is defined by the rule*

$$\widetilde{f}(\mathcal{U}) = \{B \subseteq Y : f^{-1}(B) \in \mathcal{U}\} \tag{D.3}$$

for every $\mathcal{U} \in \beta X$.

PROOF. It is easy to see that for every $\mathcal{U} \in \beta X$ the set $\widetilde{f}(\mathcal{U})$ is an ultrafilter on Y. Therefore the map $\widetilde{f}$ is well defined. Now let $B \subseteq Y$ and observe that

$$\begin{aligned} \widetilde{f}^{-1}\big((B)_{\beta Y}\big) &= \{\mathcal{U} \in \beta X : \widetilde{f}(\mathcal{U}) \in (B)_{\beta Y}\} = \{\mathcal{U} \in \beta X : B \in \widetilde{f}(\mathcal{U})\} \\ &= \{\mathcal{U} \in \beta X : f^{-1}(B) \in \mathcal{U}\} = (f^{-1}(B))_{\beta X}. \end{aligned}$$

Hence, the map $\widetilde{f}\colon \beta X \to \beta Y$ is continuous. Also notice that

$$\widetilde{f}(x) = \widetilde{f}(e_X(x)) = \{B \subseteq Y : f(x) \in B\} = e_Y(f(x)) = f(x)$$

which implies, of course, that $\widetilde{f}$ is indeed an extension of f. Finally, observe that the uniqueness of $\widetilde{f}$ follows by the fact that the set X is dense in βX. The proof of Proposition D.4 is completed. $\square$

For notational simplicity, for every function $f\colon X \to Y$ we will still denote by f the unique extension obtained by Proposition D.4. We will also write $f(\mathcal{U})$ instead of $\widetilde{f}(\mathcal{U})$ for every $\mathcal{U} \in \beta X$.

D.3. Ultrafilters as quantifiers. There is an alternative description of ultrafilters as quantifiers which is extremely convenient from a combinatorial perspective. Specifically, with every ultrafilter $\mathcal{U}$ on a nonempty set X we associate a quantifier $(\mathcal{U}x)$ as follows. If $P(x)$ is a property of elements $x \in X$, then we write

$$(\mathcal{U}x)\ P(x) \Leftrightarrow \{x \in X : P(x)\} \in \mathcal{U}. \tag{D.4}$$

That is, the formula $(\mathcal{U}x)\ P(x)$ is satisfied if and only if the set of all $x \in X$ which satisfy P is "large" in the sense that it belongs to the ultrafilter $\mathcal{U}$. For example, if A is a subset of X and $P(x)$ is the statement "$x \in A$", then

$$(\mathcal{U}x)\ [x \in A] \Leftrightarrow A \in \mathcal{U}. \tag{D.5}$$

Using basic properties of ultrafilters it is easy to see the quantifier $(\mathcal{U}x)$ commutes with conjunction and negation. Namely, if $P(x)$ and $Q(x)$ are properties of elements of X, then we have

$$\big((\mathcal{U}x)\ P(x)\big) \wedge \big((\mathcal{U}x)\ Q(x)\big) \Leftrightarrow (\mathcal{U}x)\ [P(x) \wedge Q(x)] \tag{D.6}$$

and

$$\neg\big((\mathcal{U}x)\ P(x)\big) \Leftrightarrow (\mathcal{U}x)\ [\neg P(x)]. \tag{D.7}$$

Notice that if $f\colon X \to Y$ and $\mathcal{U} \in \beta X$, then the quantifier associated with the ultrafilter $f(\mathcal{U})$ satisfies

$$\big(f(\mathcal{U})y\big)\ P(y) \Leftrightarrow (\mathcal{U}x)\ P\big(f(x)\big) \tag{D.8}$$

for every property $P(y)$ of elements of the set Y. In particular, we have

$$B \in f(\mathcal{U}) \Leftrightarrow \big(f(\mathcal{U})y\big)\ [y \in B] \Leftrightarrow (\mathcal{U}x)\ [f(x) \in B] \tag{D.9}$$

for every $B \subseteq Y$.

D.4. Algebraic properties of βX. Let $(X, *)$ be a *semigroup*, that is, a nonempty set X equipped with an associative binary relation $*$ on X. Our goal is to extend the semigroup structure of X on βX. To this end, for every $\mathcal{V}, \mathcal{W} \in \beta X$ we define

$$\mathcal{V} * \mathcal{W} = \big\{A \subseteq X : (\mathcal{V}x)(\mathcal{W}y)\ [x * y \in A]\big\}. \tag{D.10}$$

Setting

$$A_x = \{y \in X : x * y \in A\} \tag{D.11}$$

for every $A \subseteq X$ and every $x \in X$, we see that

$$\begin{aligned} A \in \mathcal{V} * \mathcal{W} \quad &\Leftrightarrow \quad \big\{x \in X : \{y \in X : x * y \in A\} \in \mathcal{W}\big\} \in \mathcal{V} \\ &\Leftrightarrow \quad \{x \in X : A_x \in \mathcal{W}\} \in \mathcal{V}. \end{aligned} \tag{D.12}$$

Recall that a *compact semigroup* is a semigroup $(S, *)$ together with a topology τ on S such that: (i) the topological space (S, τ) is compact and Hausdorff, and (ii) for every $s \in S$ the map

$$S \ni t \mapsto t * s \in S$$

is continuous. We have the following proposition.

Proposition D.5. *If $(X, *)$ is a semigroup, then the space $(\beta X, *)$ is a compact semigroup. Moreover, the binary operation $*$ on βX is an extension of the semigroup operation $*$ on X.*

Proof. The proof is based on a series of claims. First we will show that the binary operation $*$ on βX is well defined.

Claim D.6. *For every $\mathcal{V}, \mathcal{W} \in \beta X$ we have $\mathcal{V} * \mathcal{W} \in \beta X$.*

Proof of Claim D.6. We fix $\mathcal{V}, \mathcal{W} \in \beta X$. It is easy to see that $\emptyset \notin \mathcal{V} * \mathcal{W}$. Hence, property (U1) in Definition D.1 is satisfied. To see that property (U2) is satisfied, let $A \subseteq B \subseteq X$. Notice that $A_x \subseteq B_x$ for every $x \in X$. Therefore,

$$\begin{aligned} A \in \mathcal{V} * \mathcal{W} \quad &\Leftrightarrow \quad \{x \in X : A_x \in \mathcal{W}\} \in \mathcal{V} \\ &\Rightarrow \quad \{x \in X : B_x \in \mathcal{W}\} \in \mathcal{V} \Leftrightarrow B \in \mathcal{V} * \mathcal{W}. \end{aligned}$$

We proceed to show that property (U3) is satisfied. Let $A, B \in \mathcal{V} * \mathcal{W}$. Observe that $(A \cap B)_x = A_x \cap B_x$ for every $x \in X$, and so,

$$\begin{aligned} \{x \in X : (A \cap B)_x \in \mathcal{W}\} \quad &= \quad \{x \in X : A_x \cap B_x \in \mathcal{W}\} \\ &= \quad \{x \in X : A_x \in \mathcal{W}\} \cap \{x \in X : B_x \in \mathcal{W}\}. \end{aligned}$$

Hence,

$$\begin{aligned} A, B \in \mathcal{V} * \mathcal{W} \quad &\Leftrightarrow \quad \{x \in X : A_x \in \mathcal{W}\} \in \mathcal{V} \text{ and } \{x \in X : B_x \in \mathcal{W}\} \in \mathcal{V} \\ &\Leftrightarrow \quad \big(\{x \in X : A_x \in \mathcal{W}\} \cap \{x \in X : B_x \in \mathcal{W}\}\big) \in \mathcal{V} \\ &\Leftrightarrow \quad \{x \in X : (A \cap B)_x \in \mathcal{W}\} \in \mathcal{V} \Leftrightarrow A \cap B \in \mathcal{V} * \mathcal{W}. \end{aligned}$$

It remains to verify that $\mathcal{V} * \mathcal{W}$ satisfies property (U4). To this end let $A \subseteq X$ be arbitrary. Notice that $X \setminus A_x = (X \setminus A)_x$ for every $x \in X$. Thus,

$$\begin{aligned} A \notin \mathcal{V} * \mathcal{W} \quad &\Leftrightarrow \quad \{x \in X : A_x \in \mathcal{W}\} \notin \mathcal{V} \Leftrightarrow \{x \in X : A_x \notin \mathcal{W}\} \in \mathcal{V} \\ &\Leftrightarrow \quad \{x \in X : X \setminus A_x \in \mathcal{W}\} \in \mathcal{V} \\ &\Leftrightarrow \quad \{x \in X : (X \setminus A)_x \in \mathcal{W}\} \in \mathcal{V} \Leftrightarrow X \setminus A \in \mathcal{V} * \mathcal{W}. \end{aligned}$$

The proof of Claim D.6 is completed. □

Next we will show that the binary operation $*$ on βX is associative.

Claim D.7. *For every $\mathcal{U}, \mathcal{V}, \mathcal{W} \in \beta X$ we have $\mathcal{U} * (\mathcal{V} * \mathcal{W}) = (\mathcal{U} * \mathcal{V}) * \mathcal{W}$.*

Proof of Claim D.7. Fix $\mathcal{U}, \mathcal{V}, \mathcal{W} \in \beta X$. Let A be an arbitrary subset of X. Notice that $(A_x)_y = A_{x*y}$ for every $x, y \in X$. Hence,

$$\begin{aligned} A \in \mathcal{U} * (\mathcal{V} * \mathcal{W}) &\Leftrightarrow \{x \in X : A_x \in \mathcal{V} * \mathcal{W}\} \in \mathcal{U} \\ &\Leftrightarrow \big\{x \in X : \{y \in Y : (A_x)_y \in \mathcal{W}\} \in \mathcal{V}\big\} \in \mathcal{U} \\ &\Leftrightarrow \big\{x \in X : \{y \in Y : A_{x*y} \in \mathcal{W}\} \in \mathcal{V}\big\} \in \mathcal{U}. \end{aligned} \tag{D.13}$$

On the other hand,

$$\begin{aligned} A \in (\mathcal{U} * \mathcal{V}) * \mathcal{W} &\Leftrightarrow \{z \in X : A_z \in \mathcal{W}\} \in \mathcal{U} * \mathcal{V} \\ &\Leftrightarrow \big\{x \in X : \{z \in X : A_z \in \mathcal{W}\}_x \in \mathcal{V}\big\} \in \mathcal{U}. \end{aligned} \tag{D.14}$$

Notice that for every $x \in X$ we have

$$\{z \in X : A_z \in \mathcal{W}\}_x = \{y \in X : A_{x*y} \in \mathcal{W}\}. \tag{D.15}$$

Indeed,

$$\begin{aligned} w \in \{z \in X : A_z \in \mathcal{W}\}_x &\Leftrightarrow x * w \in \{z \in X : A_z \in \mathcal{W}\} \\ &\Leftrightarrow A_{x*w} \in \mathcal{W} \Leftrightarrow w \in \{y \in X : A_{x*y} \in \mathcal{W}\}. \end{aligned}$$

Therefore, by (D.14) and (D.15), we obtain that

$$A \in (\mathcal{U} * \mathcal{V}) * \mathcal{W} \Leftrightarrow \big\{x \in X : \{y \in Y : A_{x*y} \in \mathcal{W}\} \in \mathcal{V}\big\} \in \mathcal{U}. \tag{D.16}$$

Summing up, by (D.13) and (D.16), we conclude that $A \in \mathcal{U} * (\mathcal{V} * \mathcal{W})$ if and only if $A \in (\mathcal{U} * \mathcal{V}) * \mathcal{W}$. Since A was arbitrary, this shows that $\mathcal{U} * (\mathcal{V} * \mathcal{W}) = (\mathcal{U} * \mathcal{V}) * \mathcal{W}$ and the proof of Claim D.7 is completed. □

We proceed with the following claim.

Claim D.8. *For every $\mathcal{W} \in \beta X$ the map $\mathcal{V} \mapsto \mathcal{V} * \mathcal{W}$ is continuous.*

Proof of Claim D.8. Fix $\mathcal{W} \in \beta X$ and let $\phi\colon \beta X \to \beta X$ be defined by the rule $\phi(\mathcal{V}) = \mathcal{V} * \mathcal{W}$. Also let $A \subseteq X$ be arbitrary and observe that

$$\begin{aligned} \phi^{-1}\big((A)_{\beta X}\big) &= \{\mathcal{V} \in \beta X : \phi(\mathcal{V}) \in (A)_{\beta X}\} = \{\mathcal{V} \in \beta X : A \in \phi(\mathcal{V})\} \\ &= \{\mathcal{V} \in \beta X : A \in \mathcal{V} * \mathcal{W}\} = \big\{\mathcal{V} \in \beta X : \{x \in X : A_x \in \mathcal{W}\} \in \mathcal{V}\big\}. \end{aligned}$$

Therefore, setting $B = \{x \in X : A_x \in \mathcal{W}\}$, we see that $\phi^{-1}\big((A)_{\beta X}\big) = (B)_{\beta X}$. This implies, of course, that the map ϕ is continuous. The proof of Claim D.8 is completed. □

We are now ready to complete the proof of the proposition. Notice first that, by Proposition D.3 and Claims D.6, D.7 and D.8, the space $(\beta X, *)$ is a compact topological semigroup. So, it remains to show that for every $y, z \in X$ we have $e_X(y) * e_X(z) = e_X(y * z)$. Indeed, let $A \subseteq X$ be arbitrary and observe that

$$\begin{aligned} A \in e_X(y) * e_X(z) &\Leftrightarrow y \in \{x \in X : z \in A_x\} \\ &\Leftrightarrow z \in A_y \Leftrightarrow y * z \in A \Leftrightarrow A \in e_X(y * z). \end{aligned}$$

The proof of Proposition D.5 is thus completed. □

We close this subsection with the following proposition.

Proposition D.9. *Let $(X,*)$ be a semigroup. Also let $(Y,\cdot)$ be a semigroup and $T\colon (X,*) \to (Y,\cdot)$ a semigroup homomorphism, that is, $T(x_1 * x_2) = T(x_1)\cdot T(x_2)$ for every $x_1, x_2 \in X$. Then the unique extension $T\colon \beta X \to \beta Y$ is also a semigroup homomorphism.*

Proof. Let $\mathcal{V}, \mathcal{W} \in \beta X$ and fix $B \subseteq Y$. Since $T\colon X \to Y$ is a semigroup homomorphism, we see that $(T^{-1}(B))_x = T^{-1}\big(B_{T(x)}\big)$ for every $x \in X$. Therefore,

$$\begin{aligned} B \in T(\mathcal{V} * \mathcal{W}) \quad &\Leftrightarrow \quad T^{-1}(B) \in \mathcal{V} * \mathcal{W} \Leftrightarrow \{x \in X : (T^{-1}(B))_x \in \mathcal{W}\} \in \mathcal{V} \\ &\Leftrightarrow \quad \{x \in X : T^{-1}(B_{T(x)}) \in \mathcal{W}\} \in \mathcal{V} \\ &\Leftrightarrow \quad \{x \in X : B_{T(x)} \in T(\mathcal{W})\} \in \mathcal{V} \\ &\Leftrightarrow \quad T^{-1}\big(\{y \in Y : B_y \in T(\mathcal{W})\}\big) \in \mathcal{V} \\ &\Leftrightarrow \quad \{y \in Y : B_y \in T(\mathcal{W})\} \in T(\mathcal{V}) \Leftrightarrow B \in T(\mathcal{V}) \cdot T(\mathcal{W}) \end{aligned}$$

and the proof of Proposition D.9 is completed. □

D.5. Compact semigroups. In this section we present some basic properties of compact semigroups. This material is somewhat more general and is not intrinsically related to ultrafilters. However, it is conceptually quite close to the context of this appendix.

Let $(S,*)$ be a compact semigroup. Recall that a subset J of S is said to be a *left ideal* (respectively, *right ideal*) provided that $S*J \subseteq J$ (respectively, $J*S \subseteq J$). A *minimal left ideal* is a left ideal J of S not containing any left ideal of S other than itself. Finally, a subset I of S is said to be a *two-sided ideal* if I is both left and right ideal of S. We gather, below, some basic properties of left ideals (and related structures) of compact semigroups.

Proposition D.10. *Let $(S,*)$ be a compact semigroup.*

(a) *Every left ideal of S contains a minimal left ideal.*
(b) *Every minimal left ideal is closed.*
(c) *Let J be a minimal left ideal of S and $s \in S$. Then $J * s$ is a minimal left ideal. In particular, we have that $J * s = J$ if $s \in J$.*
(d) *Every minimal left ideal is contained in every two-sided ideal of S.*

Proof. (a) Fix a left ideal J of S and set

$$\mathcal{M} = \{I \subseteq J : I \text{ is closed left ideal of } S\}.$$

Since $S * s \in \mathcal{M}$ for every $s \in J$, we see that $\mathcal{M} \neq \emptyset$. By Zorn's lemma, there exists a minimal (with respect to inclusion) element I_0 of $\mathcal{M}$. We claim that I_0 is a minimal left ideal. Indeed, let $I \subseteq I_0$ be a left ideal, and let $s \in I$ be arbitrary. Notice that $S * s \subseteq I \subseteq I_0$. Since $S * s$ is a closed left ideal of S, the minimality of I_0 yields that $S * s = I_0$ which implies, of course, that $I = I_0$.

(b) Let J be a minimal left ideal of S and fix $s_0 \in J$. Notice that $S * s_0$ is a left ideal of S which is contained in J. The minimality of J yields, in particular, that $S * s_0 = J$. Invoking the continuity of the map $t \mapsto t * s_0$, we conclude that J is compact and hence closed.

(c) Notice, first, that $J*s$ is a left ideal of S since $S*(J*s) = (S*J)*s \subseteq J*s$. So we only need to show that the left ideal $J*s$ is actually minimal. To this end, let I be a minimal left ideal contained in $J*s$ (it exists by part (a) above). Fix $t \in I$. Then $t \in J*s$ and so $t = w*s$ for some $w \in J$. Observe that $S*t = I$ and $S*w = J$, since I and J are both minimal left ideals of S. Therefore,

$$I = S*t = S*(w*s) = (S*w)*s = J*s$$

which implies that is $J*s$ is a minimal left ideal.

Finally, let $s \in J$ and notice that $J*s$ is a left ideal of S with $J*s \subseteq J$. Invoking the minimality of J, we see that $J*s = J$.

(d) Let J be a minimal left ideal of S. Also let I be a two-sided ideal of S. In particular, I is a left ideal of S and so $S*(I*J) \subseteq I*J$. Hence, $I*J$ is a left ideal. On the other hand, J is a left ideal S which implies that $I*J \subseteq J$. Summing up, we see that $I*J$ is a left ideal of S which is contained in J. Invoking the minimality of J we get that $I*J = J$. Since I is also a right ideal, we conclude that $J = I*J \subseteq I$. The proof of Proposition D.10 is completed. □

An element p of a compact semigroup S is said to be an *idempotent* if $p*p = p$. The existence of idempotents in arbitrary compact semigroups is a fundamental result which is known as *Ellis's lemma*.

Lemma D.11. *Every compact semigroup contains an idempotent.*

Proof. Let $(S,*)$ be a compact semigroup and let $\mathcal{C}$ be the family of all compact subsemigroups of S ordered by inclusion. By Zorn's lemma, there exists a minimal element T_0 of $\mathcal{C}$. Let $p \in T_0$ be arbitrary. We will show that p is an idempotent. To this end, notice that $T_0 * p$ is a compact subsemigroup of T_0. By the minimality of T_0, we see that $T_0 * p = T_0$. It follows, in particular, that the set $T = \{t \in T_0 : t*p = p\}$ is nonempty. Next observe that T is a compact subsemigroup of T_0. Invoking the minimality of T_0 once again, we get that $T = T_0$. Therefore, $p \in T$ which implies, of course, that $p*p = p$. The proof of Lemma D.11 is completed. □

Now let S be a compact semigroup. On the set of all idempotents of S we define a (partial) binary relation $\preccurlyeq$ by the rule

$$p \preccurlyeq q \Leftrightarrow p*q = q*p = p. \tag{D.17}$$

It is easy to see that the relation $\preccurlyeq$ is a partial order. An idempotent p of S is said to be *minimal* if for every idempotent $q \in S$ with $q \preccurlyeq p$ we have $q = p$. We have the following proposition.

Proposition D.12. *Let $(S,*)$ be a compact semigroup.*

(a) *For every idempotent q of S there is a minimal idempotent p with $p \preccurlyeq q$.*
(b) *An idempotent of S is minimal if and only if it belongs to some minimal left ideal of S.*

Proof. We start with the following claim.

Claim D.13. *Let J be a closed left ideal of S. Also let q be an idempotent of S. Then $J * q$ contains an idempotent p with $p \preccurlyeq q$.*

Proof of Claim D.13. Notice that $(J*q)*(J*q) \subseteq J*q$ and so $J*q$ is a subsemigroup of S. Moreover, by the continuity of the map $t \mapsto t*q$, we have that $J*q$ is a compact subsemigroup of S. By Lemma D.11, there exists an idempotent r contained in $J*q$. Let $t \in J$ be such that $r = t*q$ and set $p = q*t*q = q*r$. We will show that p is as desired. Indeed, $p = q*t*q \in S*J*q \subseteq J*q$ and so $p \in J*q$. Moreover, $p*p = (q*t*q)*(q*t*q) = q*(t*q)*(t*q) = q*r*r = q*r = p$. Therefore, p is an idempotent. Finally, notice that $p*q = q*t*q*q = q*t*q = p$ and $q*p = q*q*t*q = q*t*q = p$. Hence, $p \preccurlyeq q$ and the proof of Claim D.13 is completed. □

We proceed with the following claim.

Claim D.14. *Every idempotent of a minimal left ideal is minimal.*

Proof of Claim D.14. Let J be a minimal left ideal. Also let $p \in J$ be an idempotent and fix an idempotent $q \in S$ with $q \preccurlyeq p$. Since $q = q*p \in S*J \subseteq J$ we see that $q \in J$. By Proposition D.10, we have $J*q = J$ and so there exists $r \in J$ such that $r*q = p$. Therefore, $q = p*q = r*q*q = r*q = p$. The proof of Claim D.14 is completed. □

We are now ready to complete the proof of the proposition. To this end, let q be an idempotent of S. Also let J be an arbitrary minimal left ideal of S. By part (b) of Proposition D.10 and Claim D.13, there exists an idempotent $p \in J*q$ with $p \preccurlyeq q$. By part (c) of Proposition D.10, we have that $J*q$ is also a minimal left ideal. Hence, by Claim D.14, the idempotent p is minimal. Summing up, we see that for every idempotent q the idempotent p selected above satisfies $p \preccurlyeq q$ and is minimal; in particular, the first part of the proposition is satisfied. If, in addition, the idempotent q is minimal, then $q = p$ which implies that q belongs to the minimal left ideal $J*q$. On the other hand, by Claim D.14, every idempotent of a minimal left ideal is minimal. This shows that part (b) is also satisfied. The proof of Proposition D.12 is thus completed. □

We close this appendix with the following result. It is an immediate consequence of Propositions D.10 and D.12.

Corollary D.15. *Let $(S,*)$ be a compact semigroup. Then every two-sided ideal contains all minimal idempotents of S. In particular, if I is a two-sided ideal and q is an idempotent of S, then there is a minimal idempotent $p \in I$ with $p \preccurlyeq q$.*

APPENDIX E

Probabilistic background

Let (Ω, Σ, μ) be a probability space. Also let $A \in \Sigma$ with $\mu(A) > 0$. For every $B \in \Sigma$ the *conditional probability of B given A* is the quantity

$$\mu(B \,|\, A) = \frac{\mu(B \cap A)}{\mu(A)}. \tag{E.1}$$

The *conditional probability measure of μ relative to A* is the probability measure μ_A on (Ω, Σ) defined by the rule

$$\mu_A(B) = \mu(B \,|\, A) \tag{E.2}$$

for every $B \in \Sigma$. More generally, let $f \colon \Omega \to \mathbb{R}$ be an integrable random variable and let $\mathbb{E}(f)$ denote the *expected value* of f, that is,

$$\mathbb{E}(f) = \int f \, d\mu. \tag{E.3}$$

The *conditional expectation of f with respect to A* is defined by

$$\mathbb{E}(f \,|\, A) = \frac{\int_A f \, d\mu}{\mu(A)}. \tag{E.4}$$

Notice that the conditional expectation of f with respect to Ω coincides with the expected value of f and observe that $\mathbb{E}(f \,|\, A) = \int f \, d\mu_A$ and $\mathbb{E}(\mathbf{1}_B \,|\, A) = \mu(B \,|\, A)$ for every $B \in \Sigma$. By convention we set $\mathbb{E}(f \,|\, A) = 0$ if $\mu(A) = 0$.

E.1. Main probabilistic inequalities. The following basic inequality relates the distribution of a non-negative random variable with its expected value.

Markov's inequality. *Let (Ω, Σ, μ) be a probability space. Then for every non-negative random variable f and every $\lambda > 0$ we have*

$$\mu\big(\{\omega \in \Omega : f(\omega) \geqslant \lambda\}\big) \leqslant \frac{\mathbb{E}(f)}{\lambda}. \tag{E.5}$$

Markov's inequality can be used to control the order of magnitude of a given random variable f. Note, however, that this control is insufficient when one needs to know whether f does not deviate significantly from its expected value. This information can be obtained from the higher moments of f. Specifically, for every random variable f let $\mathrm{Var}(f)$ denote the *variance* of f, that is,

$$\mathrm{Var}(f) = \int |f - \mathbb{E}(f)|^2 \, d\mu. \tag{E.6}$$

We have the following general large deviation inequality.

Chebyshev's inequality. *Let* (Ω, Σ, μ) *be a probability space. Then for every random variable* f *and every* $\lambda > 0$ *we have*

$$\mu\big(\{\omega \in \Omega : |f(\omega) - \mathbb{E}(f)| \geqslant \lambda\}\big) \leqslant \frac{\mathrm{Var}(f)}{\lambda^2}. \tag{E.7}$$

E.2. The L_p spaces. Let (Ω, Σ, μ) be a probability space and $1 \leqslant p < +\infty$. By $L_p(\Omega, \Sigma, \mu)$ we denote the vector space of all random variables $f\colon \Omega \to \mathbb{R}$ for which the quantity $\int |f|^p \, d\mu$ is finite (modulo, of course, μ-a.e. equality). For every $f \in L_p(\Omega, \Sigma, \mu)$ the L_p *norm* of f is the quantity

$$\|f\|_{L_p} = \Big(\int |f|^p \, d\mu \Big)^{1/p}. \tag{E.8}$$

(If $f \notin L_p(\Omega, \Sigma, \mu)$, then we set $\|f\|_{L_p} = +\infty$.) The vector space $L_p(\Omega, \Sigma, \mu)$ equipped with the L_p norm is a Banach space. Of particular importance is the space $L_2(\Omega, \Sigma, \mu)$ which is a Hilbert space.

Many structural properties of the spaces $L_p(\Omega, \Sigma, \mu)$ follow from the following fundamental inequality.

Hölder's inequality. *Let* $1 < p \leqslant q < +\infty$ *with* $1/p + 1/q = 1$. *Then for every pair* f, g *of random variables on a probability space* (Ω, Σ, μ) *we have*

$$\|fg\|_{L_1} \leqslant \|f\|_{L_p} \cdot \|g\|_{L_q}. \tag{E.9}$$

Notice that the case "$p = q = 2$" in (E.9) is the *Cauchy–Schwarz inequality.* Another important consequence of Hölder's inequality is the monotonicity of the L_p norms. More precisely, observe that for every $1 \leqslant p \leqslant q < +\infty$ and every random variable f we have

$$\|f\|_{L_p} \leqslant \|f\|_{L_q}. \tag{E.10}$$

This fact also follows from the following powerful inequality.

Jensen's inequality. *Let* I *be an interval of* $\mathbb{R}$ *and* $\phi\colon I \to \mathbb{R}$ *a convex function. Then for every* I*-valued integrable random variable* f *on a probability space* (Ω, Σ, μ) *we have*

$$\phi\Big(\int f \, d\mu \Big) \leqslant \int (\phi \circ f) \, d\mu. \tag{E.11}$$

E.3. Algebras and conditional expectation. Let Ω be a nonempty set and let $\mathcal{A}$ be an algebra of subsets of Ω. A set $A \in \mathcal{A}$ is said to be an *atom* of $\mathcal{A}$ if for every nonempty $B \in \mathcal{A}$ with $B \subseteq A$ we have that $B = A$. The set of all nonempty atoms of $\mathcal{A}$ will be denoted by $\mathrm{Atoms}(\mathcal{A})$. Although an infinite algebra may be atomless, note that every finite algebra has plenty of atoms. Specifically, for every finite algebra $\mathcal{A}$ on Ω the set $\mathrm{Atoms}(\mathcal{A})$ is a finite partition of Ω. Conversely, let $\mathcal{P}$ be a finite partition of Ω and denote by $\mathcal{A}_{\mathcal{P}}$ the algebra generated by $\mathcal{P}$. Notice that the algebra $\mathcal{A}_{\mathcal{P}}$ is finite and coincides with the set of all finite (possibly empty) unions of elements of $\mathcal{P}$. Also observe that $\mathrm{Atoms}(\mathcal{A}_{\mathcal{P}}) = \mathcal{P}$.

Now let (Ω, Σ, μ) be a probability space and Σ' a σ-algebra on Ω with $\Sigma' \subseteq \Sigma$. For every $f \in L_1(\Omega, \Sigma, \mu)$ by $\mathbb{E}(f \,|\, \Sigma')$ we shall denote the *conditional expectation* *of f relative to* Σ'. If $\mathcal{A}$ is a finite algebra on Ω with $\mathcal{A} \subseteq \Sigma$, then the conditional expectation $\mathbb{E}(f \,|\, \mathcal{A})$ has a particularly simple description, namely

$$\mathbb{E}(f \,|\, \mathcal{A}) = \sum_{A \in \mathrm{Atoms}(\mathcal{A})} \mathbb{E}(f \,|\, A) \cdot \mathbf{1}_A. \tag{E.12}$$

In the following proposition we recall some basic properties of the conditional expectation. For a detailed presentation of this material see, e.g., [**Bi**].

PROPOSITION E.1. *Let (Ω, Σ, μ) be a probability space and Σ' a σ-algebra on Ω with $\Sigma' \subseteq \Sigma$. Then the following hold.*

(a) *Let $1 \leqslant p < +\infty$. Also let $f, g \in L_p(\Omega, \Sigma, \mu)$ and $a, b \in \mathbb{R}$. Then we have*

(i) $\mathbb{E}(af + bg \,|\, \Sigma') = a\,\mathbb{E}(f \,|\, \Sigma') + b\,\mathbb{E}(g \,|\, \Sigma')$,

(ii) $\|\mathbb{E}(f \,|\, \Sigma')\|_{L_p} \leqslant \|f\|_{L_p}$,

(iii) $\mathbb{E}(\mathbb{E}(f \,|\, \Sigma') \,|\, \Sigma') = \mathbb{E}(f \,|\, \Sigma')$, *and*

(iv) *if $f \in L_p(\Omega, \Sigma', \mu)$, then $\mathbb{E}(f \,|\, \Sigma') = f$.*

Hence, the map $f \mapsto \mathbb{E}(f \,|\, \Sigma')$ is a linear, norm-one projection from $L_p(\Omega, \Sigma, \mu)$ onto $L_p(\Omega, \Sigma', \mu)$.

(b) *For every $f, g \in L_2(\Omega, \Sigma, \mu)$ we have*

$$\int f \cdot \mathbb{E}(g \,|\, \Sigma')\, d\mu = \int \mathbb{E}(f \,|\, \Sigma') \cdot g\, d\mu.$$

That is, the projection $f \mapsto \mathbb{E}(f \,|\, \Sigma')$ is self-adjoint on $L_2(\Omega, \Sigma, \mu)$. In particular, we have $\|f\|^2_{L_2} = \|\mathbb{E}(f \,|\, \Sigma')\|^2_{L_2} + \|f - \mathbb{E}(f \,|\, \Sigma')\|^2_{L_2}$ for every $f \in L_2(\Omega, \Sigma, \mu)$.

(c) *If Σ'' is a σ-algebra on Ω with $\Sigma' \subseteq \Sigma'' \subseteq \Sigma$, then for every $f \in L_1(\Omega, \Sigma, \mu)$ we have $\mathbb{E}(\mathbb{E}(f \,|\, \Sigma'') \,|\, \Sigma') = \mathbb{E}(f \,|\, \Sigma')$. Therefore, for every $1 \leqslant p < +\infty$ the projections $f \mapsto \mathbb{E}(f \,|\, \Sigma')$ and $f \mapsto \mathbb{E}(f \,|\, \Sigma'')$ in $L_p(\Omega, \Sigma, \mu)$ are commuting.*

E.4. Products of probability spaces. Let $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ be two probability spaces. We endow the Cartesian product $\Omega_1 \times \Omega_2$ with the *tensor product σ-algebra* $\Sigma_1 \otimes \Sigma_2$ of Σ_1 and Σ_2, that is, the σ-algebra on $\Omega_1 \times \Omega_2$ generated by the sets

$$\{A_1 \times A_2 : A_1 \in \Sigma_1 \text{ and } A_2 \in \Sigma_2\}.$$

The *product measure* $\mu_1 \times \mu_2$ of μ_1 and μ_2 is the unique probability measure on the measurable space $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2)$ satisfying

$$(\mu_1 \times \mu_2)(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2)$$

for every $A_1 \in \Sigma_1$ and every $A_2 \in \Sigma_2$. Finally, the *product* of the spaces $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ is the probability space $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, \mu_1 \times \mu_2)$. The product of an arbitrary nonempty finite family of probability spaces is constructed by iterating this basic operation.

Now let $f \colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ be a function. Given $x \in \Omega_1$ we define $f_x \colon \Omega_2 \to \mathbb{R}$ by the rule $f_x(y) = f(x, y)$. Respectively, for every $y \in \Omega_2$ we define $f_y \colon \Omega_1 \to \mathbb{R}$ by $f_y(x) = f(x, y)$. The following result is known as *Fubini's theorem* and is a fundamental property of product spaces (see, e.g., [**Bi**]).

THEOREM E.2. *Let $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ be two probability spaces and $f \in L_1(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, \mu_1 \times \mu_2)$. Then the following hold.*

(a) *We have $f_x \in L_1(\Omega_2, \Sigma_2, \mu_2)$ for μ_1-almost all $x \in \Omega_1$.*
(b) *We have $f_y \in L_1(\Omega_1, \Sigma_1, \mu_1)$ for μ_2-almost all $y \in \Omega_2$.*
(c) *The random variables $x \mapsto \mathbb{E}(f_x)$ and $y \mapsto \mathbb{E}(f_y)$ are integrable and*

$$\int f \, d(\mu_1 \times \mu_2) = \int \mathbb{E}(f_x) \, d\mu_1 = \int \mathbb{E}(f_y) \, d\mu_2. \tag{E.13}$$

E.5. General lemmas. We close this appendix by presenting some basic facts, of probabilistic nature, which are used throughout this book. We start with the following variant of Markov's inequality.

LEMMA E.3. *Let $0 < \varepsilon < \delta \leqslant 1$ and $(\delta_i)_{i=1}^n$ a nonempty finite sequence in $[0,1]$. Assume that $\mathbb{E}_{i \in [n]} \delta_i \geqslant \delta$ and $|\{i \in [n] : \delta_i \geqslant \delta + \varepsilon^2\}| \leqslant \varepsilon^3 n$. Then we have*

$$|\{i \in [n] : \delta_i \geqslant \delta - \varepsilon\}| \geqslant (1 - \varepsilon) n. \tag{E.14}$$

PROOF. We set $I = \{i \in [n] : \delta_i \geqslant \delta - \varepsilon\}$ and $J = [n] \setminus I$. Moreover, let $I_1 = \{i \in [n] : \delta - \varepsilon \leqslant \delta_i < \delta + \varepsilon^2\}$ and $I_2 = \{i \in [n] : \delta + \varepsilon^2 \leqslant \delta_i \leqslant 1\}$. Notice that $I = I_1 \cup I_2$ and $|I_2| \leqslant \varepsilon^3 n$. Therefore,

$$\begin{aligned} \delta n \leqslant \sum_{i=1}^n \delta_i &= \sum_{i \in I_1} \delta_i + \sum_{i \in I_2} \delta_i + \sum_{i \in J} \delta_i \\ &\leqslant (\delta + \varepsilon^2)|I_1| + |I_2| + (\delta - \varepsilon)(n - |I|) \\ &\leqslant (\delta + \varepsilon^2)|I| + \varepsilon^3 n + (\delta - \varepsilon)(n - |I|) \end{aligned}$$

which implies that $|I| \geqslant (1 - \varepsilon) n$. □

We proceed with the following lemma.

LEMMA E.4. *Let (Ω, Σ, μ) be a probability space. Also let $0 < \delta \leqslant 1$ and $(A_i)_{i=1}^n$ a nonempty finite sequence of measurable events in (Ω, Σ, μ) such that $\mu(A_i) \geqslant \delta$ for every $i \in [n]$. For every $\omega \in \Omega$ we set $L_\omega = \{i \in [n] : \omega \in A_i\}$. Then there exists $\omega_0 \in \Omega$ such that $|L_{\omega_0}| \geqslant \delta n$. Moreover,*

$$\mu\big(\{\omega \in \Omega : |L_\omega| \geqslant (\delta/2) n\}\big) \geqslant \delta/2. \tag{E.15}$$

PROOF. Set $f = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{A_i}$ and notice that $|L_\omega| = f(\omega) \cdot n$ for every $\omega \in \Omega$. Since $\mathbb{E}(f) \geqslant \delta$ there exists $\omega_0 \in \Omega$ with $f(\omega_0) \geqslant \delta$ which is equivalent to saying that $|L_{\omega_0}| \geqslant \delta n$. Moreover, the random variable f takes values in $[0,1]$ and so

$$\delta \leqslant \mathbb{E}(f) \leqslant \mu\big(\{\omega \in \Omega : f(\omega) \geqslant (\delta/2) n\}\big) + \frac{\delta}{2}.$$

The proof of Lemma E.4 is completed. □

The next result asserts that any sufficiently large family of measurable events in a probability space contains two events which are at least as correlated as if they were independent.

Lemma E.5. *Let $0 < \theta < \varepsilon \leqslant 1$ and $n \in \mathbb{N}$ with $n \geqslant (\varepsilon^2 - \theta^2)^{-1}$. If $(A_i)_{i=1}^n$ is a finite sequence of measurable events in a probability space (Ω, Σ, μ) satisfying $\mu(A_i) \geqslant \varepsilon$ for every $i \in [n]$, then there exist $i, j \in [n]$ with $i \neq j$ and $\mu(A_i \cap A_j) \geqslant \theta^2$.*

Proof. Let $f = \sum_{i=1}^n \mathbf{1}_{A_i}$. Then $\mathbb{E}(f) \geqslant \varepsilon n$ and so, by Jensen's inequality,

$$\sum_{i \in [n]} \sum_{j \in [n] \setminus \{i\}} \mu(A_i \cap A_j) = \mathbb{E}(f^2 - f) \geqslant \varepsilon^2 n^2 - \varepsilon n.$$

Therefore, there exist $i, j \in [n]$ with $i \neq j$ such that $\mu(A_i \cap A_j) \geqslant \theta^2$. □

The following lemma concerns, essentially, the distribution of a measurable event in the sets of a partition of the sample space, though the precise statement is somewhat more general. This more general form will be needed in Chapter 9.

Lemma E.6. *Let (Ω, Σ, μ) be a probability space and $0 < \lambda, \beta, \varepsilon \leqslant 1$. Let A and B be two measurable events in (Ω, Σ, μ) with $A \subseteq B$ and such that $\mu(A) \geqslant \lambda \mu(B)$ and $\mu(B) \geqslant \beta$. Suppose that $\mathcal{Q} = \{Q_1, \dots, Q_n\}$ is a nonempty finite family of pairwise disjoint measurable events in (Ω, Σ, μ) such that $\mu(B \setminus \cup \mathcal{Q}) \leqslant \varepsilon\beta/2$ and $\mu(Q_i) > 0$ for every $i \in [n]$. Then, setting*

$$I = \big\{ i \in [n] : \mu_{Q_i}(A) \geqslant (\lambda - \varepsilon)\mu_{Q_i}(B) \text{ and } \mu_{Q_i}(B) \geqslant \beta\varepsilon/4 \big\}, \tag{E.16}$$

we have

$$\sum_{i \in I} \mu(Q_i) \geqslant \beta\varepsilon/4. \tag{E.17}$$

In particular, if $\mu(Q_i) = \mu(Q_j)$ for every $i, j \in [n]$, then $|I| \geqslant (\beta\varepsilon/4)n$.

Proof. Notice, first, that $\mu(A \setminus \cup \mathcal{Q}) \leqslant \varepsilon\beta/2$. This is easily seen to imply that

$$\sum_{i=1}^n \frac{\mu(A \cap Q_i)}{\mu(B)} \geqslant \lambda - \varepsilon/2. \tag{E.18}$$

For every $i \in [n]$ let $a_i = \mu_{Q_i}(A)/\mu_{Q_i}(B)$, $b_i = \mu_{Q_i}(B)$ and $c_i = \mu(Q_i)/\mu(B)$ with the convention that $a_i = 0$ if $\mu(B \cap Q_i) = 0$. Then inequality (E.18) can be reformulated as

$$\sum_{i=1}^n a_i b_i c_i \geqslant \lambda - \varepsilon/2. \tag{E.19}$$

Notice that

$$\sum_{i=1}^n b_i c_i \leqslant 1 \quad \text{and} \quad \sum_{i=1}^n c_i \leqslant \frac{1}{\beta}. \tag{E.20}$$

Also observe that $I = \{i \in [n] : a_i \geqslant \lambda - \varepsilon \text{ and } b_i \geqslant \beta\varepsilon/4\}$. Since $0 \leqslant a_i, b_i \leqslant 1$ for every $i \in [n]$, by (E.19) and (E.20) and the previous remarks, we conclude that $\sum_{i \in I} c_i \geqslant \varepsilon/4$. The proof of Lemma E.6 is completed. □

The last result is a classical estimate for the tail of the binomial distribution. Specifically, let $H \colon [0,1] \to \mathbb{R}$ be the *binary entropy function* (see, e.g., [**Re**]). Recall that $H(0) = H(1) = 0$ and

$$H(x) = -x \log_2(x) - (1-x)\log_2(1-x) \tag{E.21}$$

for every $0 < x < 1$. Observe that $H(1/2 - z) = H(1/2 + z)$ for all $0 \leqslant z \leqslant 1/2$. Also notice that H is continuous and its restriction on the interval $[0, 1/2]$ is strictly increasing and onto $[0, 1]$. We have the following lemma.

LEMMA E.7. *Let $0 < \varepsilon \leqslant 1/2$. Then for every positive integer n we have*

$$\sum_{i=0}^{\lfloor \varepsilon \cdot n \rfloor} \binom{n}{i} < 2^{H(\varepsilon)\cdot n}. \tag{E.22}$$

PROOF. Since $2^{H(\varepsilon)\cdot n} = \varepsilon^{-\varepsilon\cdot n}(1-\varepsilon)^{-(1-\varepsilon)\cdot n}$ it is enough to show that

$$\sum_{i=0}^{\lfloor \varepsilon \cdot n \rfloor} \binom{n}{i} \varepsilon^{\varepsilon\cdot n}(1-\varepsilon)^{(1-\varepsilon)\cdot n} < 1.$$

Using the fact that $0 < \varepsilon \leqslant 1/2$ we see that $\varepsilon^{\varepsilon\cdot n}(1-\varepsilon)^{(1-\varepsilon)\cdot n} \leqslant \varepsilon^{i}(1-\varepsilon)^{n-i}$ for every $0 \leqslant i \leqslant \lfloor \varepsilon \cdot n \rfloor$. Therefore,

$$\begin{aligned} \sum_{i=0}^{\lfloor \varepsilon \cdot n \rfloor} \binom{n}{i} \varepsilon^{\varepsilon\cdot n}(1-\varepsilon)^{(1-\varepsilon)\cdot n} &\leqslant \sum_{i=0}^{\lfloor \varepsilon \cdot n \rfloor} \binom{n}{i} \varepsilon^{i}(1-\varepsilon)^{n-i} \\ &< \sum_{i=0}^{n} \binom{n}{i} \varepsilon^{i}(1-\varepsilon)^{n-i} = \big(\varepsilon + (1-\varepsilon)\big)^{n} = 1 \end{aligned}$$

and the proof of Lemma E.7 is completed. $\square$

APPENDIX F

Open problems

F.1. Hales–Jewett numbers and related problems. We start with the following classical problem in Ramsey theory.

Problem 1. *Which is the asymptotic behavior of the Hales–Jewett numbers?*

There is no reasonable conjecture in this direction, partly because it is very difficult to predict the growth of the numbers $\mathrm{HJ}(k, r)$. As we have pointed out in Section 2.4, any significant improvement on Shelah's bound would be of fundamental importance.

The understanding of the growth of the density Hales–Jewett numbers is even less satisfactory. Indeed, the best known upper bounds for the numbers $\mathrm{DHJ}(k, \delta)$ have an Ackermann-type dependence with respect to k.

Problem 2. *Which is the asymptotic behavior of the density Hales–Jewett numbers? In particular, is it true that the numbers* $\mathrm{DHJ}(k, \delta)$ *are upper bounded by a primitive recursive function?*

It is quite likely that the second part of Problem 2 has an affirmative answer. In fact, it is natural to expect stronger results in this direction.

Conjecture 3. *The density Carlson–Simpson numbers,* $\mathrm{DCS}(k, m, \delta)$*, are upper bounded by a primitive recursive function.*

Note that, by Proposition 9.59, an affirmative answer to Conjecture 3 also yields an affirmative answer to the second part of Problem 2.

F.2. Carlson's theorem. As we have mentioned in Section 4.1, all known proofs of Carlson's theorem rely on the use of ultrafilters and/or methods from topological dynamics. (Note, however, that most of its consequences can be proved by combinatorial means.)

Problem 4. *Find a purely combinatorial proof of Carlson's theorem.*

F.3. Extensions of the Furstenberg–Weiss theorem. The following problem asks whether a multidimensional version of the Furstenberg–Weiss theorem (Theorem 9.65) holds true.

Problem 5. *Let* $\ell \geqslant 3$ *and* $0 < \delta \leqslant 1$*. Also let* $\mathbf{T} = (T_1, \dots, T_d)$ *be a vector homogeneous tree of finite height and* D *a subset of the level product of* $\mathbf{T}$ *satisfying*

$$\mathbb{E}_{n \in \{0,\dots,h(\mathbf{T})-1\}} \frac{|D \cap \big(T_1(n) \times \dots \times T_d(n)\big)|}{|T_1(n) \times \dots \times T_d(n)|} \geqslant \delta.$$

If the height $h(\mathbf{T})$ of $\mathbf{T}$ is sufficiently large (depending on ℓ, δ, d and the branching numbers of $T_1, \dots, T_d$), then does there exist a vector strong subtree $\mathbf{S}$ of $\mathbf{T}$ of height ℓ with $\otimes\mathbf{S} \subseteq D$ and whose level set $L_{\mathbf{T}}(\mathbf{S})$ is an arithmetic progression?

There is a stronger (and in a sense more complete) version of Problem 5 which refers to Carlson–Simpson spaces.

Problem 6. *Let $m \geqslant 3$ and $0 < \delta \leqslant 1$, and let A be a finite alphabet with $|A| \geqslant 2$. Also let N be a positive integer and D a subset of $A^{<N+1}$ satisfying*

$$\mathbb{E}_{n\in\{0,\dots,N\}} \frac{|D \cap A^n|}{|A^n|} \geqslant \delta.$$

If N is sufficiently large (depending on m, δ and the cardinality of A), then does there exist an m-dimensional Carlson–Simpson subspace W of $A^{<N+1}$ which is contained in D and whose level set $L(W)$ is an arithmetic progression?

Not much is known in this direction. In fact, even the following coloring version of Problem 6 is open, and is already quite interesting.

Problem 6′. *Let $m \geqslant 3$ and $r \geqslant 2$, and let A be a finite alphabet with $|A| \geqslant 2$. Also let N be a positive integer and $c\colon A^{<N+1} \to [r]$ a coloring. If N is sufficiently large (depending on m, r and the cardinality of A), then does there exist a monochromatic m-dimensional Carlson–Simpson subspace W of $A^{<N+1}$ whose level set $L(W)$ is an arithmetic progression?*

F.4. Bounds for the hypergraph removal lemma. As we have already pointed out in Section 7.6, all known effective proofs of the hypergraph removal lemma yield lower bounds for the constant $\varrho(n, r, \varepsilon)$ in Theorem 7.16 which have an Ackermann-type dependence with respect to r.

Problem 7. *Which is the asymptotic behavior of the constants $\varrho(n, r, \varepsilon)$? In particular, is it true that there exist primitive recursive bounds for the hypergraph removal lemma?*

We notice that Problem 7 and instances thereof have been asked by several authors (see, e.g., [**Tao1**]).

* * *

We close this appendix with a brief discussion on three long-standing open problems in Ramsey theory. Although these problems are somewhat distinct from the main theme of this book, they are certainly in line with the general context of this appendix.

F.5. Diagonal Ramsey numbers. For every $k \in \mathbb{N}$ with $k \geqslant 2$ let $\mathrm{R}(k)$ be the *k-th diagonal Ramsey number*, that is, the least integer $n \geqslant k$ such that for every n-element set X and every 2-coloring of $\binom{X}{2}$ there exists $Z \in \binom{X}{k}$ such that the set $\binom{Z}{2}$ is monochromatic. The existence of these numbers follows, of course, from the work of Ramsey [**Ra**], but the standard upper bound, $\mathrm{R}(k) \leqslant \binom{2k-2}{k-1}$, is due to Erdős and Szekeres [**ErdS**]. On the other hand, the first non-trivial lower bound was obtained by Erdős [**Erd1**] who showed that $2^{k/2} \leqslant \mathrm{R}(k)$. Combining these classical estimates, we see that

$$\sqrt{2} \leqslant \mathrm{R}(k)^{1/k} \leqslant 4. \tag{F.1}$$

Currently, the best known general estimates for the diagonal Ramsey numbers are

$$\big(1+o(1)\big)\frac{\sqrt{2}}{e}\, k\, 2^{k/2} \leqslant R(k) \leqslant (k-1)^{-C\frac{\log(k-1)}{\log\log(k-1)}}\binom{2k-2}{k-1} \tag{F.2}$$

where $C>0$ is an absolute constant. The lower bound in (F.2) is due to Spencer [**Spe**], while the upper bound is due to Conlon [**Co**] who improved previous work of Thomason [**Th**]. Although (F.2) is an important advance, it has a minor impact on (F.1) since it yields that

$$\sqrt{2} \leqslant \liminf_{k\to\infty} \mathrm{R}(k)^{1/k} \leqslant \limsup_{k\to\infty} \mathrm{R}(k)^{1/k} \leqslant 4.$$

The question of determining the exact asymptotic behavior of the sequence $\langle \mathrm{R}(k)^{1/k} : k \geqslant 2\rangle$ is of fundamental importance in Ramsey theory and has been asked by several authors (see, e.g., [**Erd2, GRo, GRS**]). This is the content of the following problem due to Erdős.

Problem 8. *Is it true that the sequence $\langle \mathrm{R}(k)^{1/k} : k \geqslant 2\rangle$ converges? And if yes, then what is its limit?*

F.6. Bounds for Szemerédi's theorem. For every pair N,k of positive integers with $N \geqslant k \geqslant 3$ let $r_k(N)$ be the cardinality of the largest subset of $[N]$ not containing an arithmetic progression of length k. Note that Szemerédi's theorem is equivalent to saying that

$$\lim_{N\to\infty}\frac{r_k(N)}{N} = 0$$

for every integer $k \geqslant 3$.

Problem 9. *Which is the asymptotic behavior of $r_k(N)$?*

Problem 9 is discussed in detail in [**Go6**]. The case "$k=3$" is among the most heavily investigated questions in combinatorial number theory. A classical lower bound for $r_3(N)$ is due to Behrend [**Beh**], while the first upper bound was obtained by Roth [**Ro**]. Currently, the best known estimates for $r_3(N)$ are[1]

$$Ne^{-\sqrt{8\log N}}(\log N)^{1/4} \ll r_3(N) \ll \frac{N}{(\log N)^{1-o(1)}} \tag{F.3}$$

due to Elkin [**Elk**] and T. Sanders [**Sa2**] respectively (see also [**Blo**]). The estimation of $r_k(N)$ becomes harder as k increases, and as such, progress for $k \geqslant 4$ has been much slower. In particular, the best known general estimates are

$$Ne^{-c_k(\log N)^{1/\lceil \log_2 k\rceil}}(\log N)^{1/2\lceil \log_2 k\rceil} \ll r_k(N) \ll N(\log\log N)^{-1/2^{2^{k+9}}}. \tag{F.4}$$

[1]We write $f(N) \ll g(N)$ to denote that there exists an absolute constant $C>0$ such that $f(N) \leqslant Cg(N)$ for all sufficiently large $N \in \mathbb{N}$.

The lower bound in (F.4) is due to O'Bryant [**OBr**]. The upper bound is due to Gowers [**Go3**] and is the only "reasonable" upper bound for $r_k(N)$.

Problem 9 is closely related to the following famous conjecture of Erdős.

CONJECTURE 10 (Erdős' conjecture on arithmetic progressions). *Let A be a subset of $\mathbb{N}$ such that $\sum_{n\in A} n^{-1} = \infty$. Then A contains arbitrarily long arithmetic progressions.*

It is not difficult to see that to prove Erdős' conjecture it is sufficient to show that $r_k(N) \ll N(\log N)^{-1}(\log\log N)^{-2}$ for every $k \geqslant 3$. Also note that an affirmative answer to Erdős's conjecture would imply the celebrated result of Green and Tao [**GT**] that the set of primes contains arbitrarily long arithmetic progressions.

F.7. The density polynomial Hales–Jewett conjecture. Let A be a finite alphabet with $|A| \geqslant 2$, and fix a letter x not belonging to A which we view as a variable. For every pair n, d of positive integers let $A^{[n]^d}$ be the set of all maps from the d-fold Cartesian product $[n]^d$ to A. A *polynomial variable word of* $A^{[n]^d}$ is a map $v\colon [n]^d \to A \cup \{x\}$ such that $v^{-1}(\{x\}) = X^d$ for some nonempty subset X of $[n]$. If v is a polynomial word of $A^{[n]^d}$ and $a \in A$, then $v(a)$ denotes the unique element of $A^{[n]^d}$ obtained by substituting in v all appearances of the variable x with a. A *polynomial line of* $A^{[n]^d}$ is a set of the form $\{v(a) : a \in A\}$ where v is a polynomial variable word of $A^{[n]^d}$.

The following result is known as the *polynomial Hales–Jewett theorem* and is due to Bergelson and Leibman [**BL**].

THEOREM F.1. *For every triple k, d, r of positive integers with $k \geqslant 2$ there exists a positive integer N with the following property. If $n \geqslant N$, then for every alphabet A with $|A| = k$ and every r-coloring of $A^{[n]^d}$ there exists a polynomial variable word v of $A^{[n]^d}$ such that the set $\{v(a) : a \in A\}$ is monochromatic. The least positive integer with this property will be denoted by* $\mathrm{PHJ}(k, d, r)$.

The original proof of Theorem F.1 was based on tools from topological dynamics, but soon after its discovery a combinatorial proof was given in [**W**]. The best known upper bounds for the numbers $\mathrm{PHJ}(k, d, r)$ were obtained slightly later by Shelah [**Sh2**].

The polynomial Hales–Jewett theorem has a number of beautiful consequences in Ramsey theory, several of which are discussed in detail in [**McC1**]. However, it is not known whether there exists a density version of the polynomial Hales–Jewett theorem. This is the content of the following conjecture of Bergelson [**Ber**].

CONJECTURE 11. *For every pair k, d of positive integers with $k \geqslant 2$ and every $0 < \delta \leqslant 1$ there exists a positive integer N with the following property. If $n \geqslant N$ and A is an alphabet with $|A| = k$, then every subset of $A^{[n]^d}$ with cardinality at least δk^{n^d} contains a polynomial line of $A^{[n]^d}$.*

Note that the case "$d = 1$" of Conjecture 11 is just the density Hales–Jewett theorem, but even the simplest higher-dimensional case, "$k = d = 2$", is open. This particular case is equivalent to a conjectural two-dimensional extension of Sperner's theorem and is very interesting on its own.

Problem 12. *For every* $0 < \delta \leqslant 1$ *there exists a positive integer* N *with the following property. If* $n \geqslant N$ *and* $\mathcal{D}$ *is a family of subsets of* $[n]^2$ *with cardinality at least* $\delta 2^{n^2}$*, then there exist* $A, B \in \mathcal{D}$ *with* $A \subseteq B$ *such that* $B \setminus A$ *is of the form* $X \times X$ *for some nonempty subset* X *of* $[n]$.

Bibliography

[ASz] M. Ajtai and E. Szemerédi, *Sets of lattice points that form no squares*, Stud. Sci. Math. Hungar. **9** (1974), 9–11.

[ADLRY] N. Alon, R. A. Duke, H. Lefmann, V. Rödl and R. Yuster, *The algorithmic aspects of the regularity lemma*, J. Algorithms **16** (1994), 80–109.

[AFKS] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, *Efficient testing of large graphs*, Combinatorica **20** (2000), 451–476.

[ASp] N. Alon and J. Spencer, *The Probabilistic Method*, 2nd edition, John Wiley & Sons, 2000.

[AFK] S. A. Argyros, V. Felouzis and V. Kanellopoulos, *A proof of Halpern–Läuchli partition theorem*, Eur. J. Comb. **23** (2002), 1–10.

[A] T. Austin, *Deducing the density Hales–Jewett theorem from an infinitary removal lemma*, J. Theor. Probab. **24** (2011), 615–633.

[AT] T. Austin and T. Tao, *On the testability and repair of hereditary hypergraph properties*, Random Struct. Algorithms **36** (2010), 373–463.

[Bau] J. E. Baumgartner, *A short proof of Hindman's theorem*, J. Comb. Theory, Ser. A **17** (1974), 384–386.

[Beh] F. A. Behrend, *On sets of integers which contain no three in arithmetic progression*, Proc. Natl. Acad. Sci. USA **23** (1946), 331–332.

[Ber] V. Bergelson, *Ergodic Ramsey theory—an update*, in "Ergodic Theory of $\mathbb{Z}^d$-Actions", London Mathematical Society Lecture Note Series, Vol. 228, Cambridge University Press, 1996, 1–61.

[BBH] V. Bergelson, A. Blass and N. Hindman, *Partition theorems for spaces of variable words*, Proc. London Math. Soc. **68** (1994), 449–476.

[BL] V. Bergelson and A. Leibman, *Set-polynomials and polynomial extension of the Hales–Jewett theorem*, Ann. Math. **150** (1999), 33–75.

[B1] A. Blass, *A partition theorem for perfect sets*, Proc. Amer. Math. Soc. **82** (1981), 271–277.

[B2] A. Blass, *Ultrafilters: where topological dynamics = algebra = combinatorics*, Topology Proc. **18** (1993), 33–56.

[Blo] T. F. Bloom, *A quantitative improvement for Roth's theorem on arithmetic progressions*, J. London Math. Soc. **93** (2016), 643–663.

[BV] R. Bicker and B. Voigt, *Density theorems for finitistic trees*, Combinatorica **3** (1983), 305–313.

[Bi] P. Billingsley, *Probability and Measure*, 3rd edition, John Wiley & Sons, 1995.

[BN] B. Bollobás and V. Nikiforov, *An abstract regularity lemma*, in "Building Bridges", Bolyai Society Mathematical Studies, Vol. 19, Springer, 2008, 219–240.

[C] T. J. Carlson, *Some unifying principles in Ramsey theory*, Discr. Math. **68** (1988), 117–169.

[CS] T. J. Carlson and S. G. Simpson, *A dual form of Ramsey's theorem*, Adv. Math. **53** (1984), 265–290.

[Co] D. Conlon, *A new upper bound for diagonal Ramsey numbers*, Ann. Math. **170** (2009), 941–960.

[DKK] P. Dodos, V. Kanellopoulos and N. Karagiannis, *A density version of the Halpern–Läuchli theorem*, Adv. Math. **244** (2013), 955–978.

[DKKa] P. Dodos, V. Kanellopoulos and Th. Karageorgos, *Szemerédi's regularity lemma via martingales*, Electron. J. Comb. **23** (2016), Research Paper P3.11, 1–24.

[DKT1] P. Dodos, V. Kanellopoulos and K. Tyros, *Dense subsets of products of finite trees*, Int. Math. Res. Not. **4** (2013), 924–970.

[DKT2] P. Dodos, V. Kanellopoulos and K. Tyros, *A simple proof of the density Hales–Jewett theorem*, Int. Math. Res. Not. **12** (2014), 3340–3352.

[DKT3] P. Dodos, V. Kanellopoulos and K. Tyros, *A density version of the Carlson–Simpson theorem*, J. Eur. Math. Soc. **16** (2014), 2097–2164.

[DKT4] P. Dodos, V. Kanellopoulos and K. Tyros, *Measurable events indexed by words*, J. Comb. Theory, Ser. A **127** (2014), 176–223.

[DKT5] P. Dodos, V. Kanellopoulos and K. Tyros, *A concentration inequality for product spaces*, J. Funct. Anal. **270** (2016), 609–620.

[ES] G. Elek and B. Szegedy, *A measure-theoretic approach to the theory of dense hypergraphs*, Adv. Math. **231** (2012), 1731–1772.

[Elk] M. Elkin, *An improved construction of progression-free sets*, Israel J. Math. **184** (2011), 93–128.

[E] E. Ellentuck, *A new proof that analytic sets are Ramsey*, J. Symb. Logic **39** (1974), 163–165.

[Erd1] P. Erdős, *Some remarks on the theory of graphs*, Bull. Amer. Math. Soc. **53** (1947), 292–294.

[Erd2] P. Erdős, *Problems and results on graphs and hypergraphs: similarities and differences*, in "Mathematics of Ramsey Theory", Algorithms and Combinatorics, Vol. 5, Springer, 1990, 12–28.

[EFR] P. Erdős, P. Frankl and V. Rödl, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs Comb. **2** (1986), 113–121.

[EH] P. Erdős and A. Hajnal, *Some remarks on set theory, IX. Combinatorial problems in measure theory and set theory*, Mich. Math. Journal **11** (1964), 107–127.

[ER] P. Erdős and R. Rado, *Combinatorial theorems on classifications of subsets of a given set*, Proc. London Math. Soc. **2** (1952), 417–439.

[ErdS] P. Erdős and G. Szekeres, *A combinatorial problem in geometry*, Compos. Math. **2** (1935), 463–470.

[FR1] P. Frankl and V. Rödl, *The uniformity lemma for hypergraphs*, Graphs Comb. **8** (1992), 309–312.

[FR2] P. Frankl and V. Rödl, *Extremal problems on set systems*, Random Struct. Algorithms **20** (2002), 131–164.

[FrK] A. Frieze and R. Kannan, *Quick approximation to matrices and applications*, Combinatorica **19** (1999), 175–220.

[Fu] Z. Füredi, *Extremal hypergraphs and combinatorial geometry*, in "Proceedings of the International Congress of Mathematicians", Vol. 2, Birkhäuser, 1995, 1343–1352.

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

[FK1] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Anal. Math. **31** (1978), 275–291.

[FK2] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for IP-systems and combinatorial theory*, J. Anal. Math. **45** (1985), 117–168.

[FK3] H. Furstenberg and Y. Katznelson, *Idempotents in compact semigroups and Ramsey theory*, Israel J. Math. **68** (1989), 257–270.

[FK4] H. Furstenberg and Y. Katznelson, *A density version of the Hales–Jewett theorem*, J. Anal. Math. **57** (1991), 64–119.

[FW] H. Furstenberg and B. Weiss, *Markov processes and Ramsey theory for trees*, Comb. Prob. Comput. **12** (2003), 547–563.

[Ga1] F. Galvin, *A generalization of Ramsey's theorem*, Notices Amer. Math. Soc. **15** (1968), 548.

[Ga2] F. Galvin, *Partition theorems for the real line*, Notices Amer. Math. Soc. **15** (1968), 660.

[GP] F Galvin and K. Prikry, *Borel sets and Ramsey's theorem*, J. Symb. Logic **38** (1973), 193–198.

[Go1] W. T. Gowers, *Lipschitz functions on classical spaces*, Eur. J. Comb. **13** (1992), 141–151.

[Go2] W. T. Gowers, *Lower bounds of tower type for Szemerédi's uniformity lemma*, Geom. Funct. Anal. **7** (1997), 322–337.

[Go3] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465–588.

[Go4] W. T. Gowers, *Quasirandomness, counting and regularity for 3-uniform hypergraphs*, Comb. Prob. Comput. **15** (2006), 143–184.

[Go5] W. T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. Math. **166** (2007), 897–946.

[Go6] W. T. Gowers, *Erdős and arithmetic progressions*, in "Erdős Centennial", Bolyai Society Mathematical Studies, Vol. 25, Springer, 2013, 265–287.

[GLR] R. L. Graham, K. Leeb and B. L. Rothschild, *Ramsey's theorem for a class of categories*, Adv. Math. **8** (1972), 417–433.

[GRo] R. L. Graham and V. Rödl, *Numbers in Ramsey theory*, in "Surveys in Combinatorics, 1987", London Mathematical Society Lecture Note Series, Vol. 123, Cambridge University Press, 1987, 111–153.

[GR] R. L. Graham and B. L. Rothschild, *Ramsey's theorem for n-parameter sets*, Trans. Amer. Math. Soc. **159** (1971), 257–292.

[GRS] R. L. Graham, B. L. Rothschild and J. H. Spencer, *Ramsey Theory*, 2nd edition, John Wiley & Sons, 1980.

[GT] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. Math. **167** (2008), 481–547.

[Grz] A. Grzegorczyk, *Some classes of recursive functions*, Diss. Math. **4** (1953), 1–45.

[HJ] A. H. Hales and R. I. Jewett, *Regularity and positional games*, Trans. Amer. Math. Soc. **106** (1963), 222–229.

[HL] J. D. Halpern and H. Läuchli, *A partition theorem*, Trans. Amer. Math. Soc. **124** (1966), 360–367.

[HLe] J. D. Halpern and A. Lévy, *The boolean prime ideal theorem does not imply the axiom of choice*, Proceedings of the Symposium in Pure Mathematics, Vol. 13, part I, American Mathematical Society, Providence, R. I., 1967, 83–134.

[H] N. Hindman, *Finite sums from sequences within cells of a partition of* $\mathbb{N}$, J. Comb. Theory, Ser. A **17** (1974), 1–11.

[HM1] N. Hindman and R. McCutcheon, *Partition theorems for left and right variable words*, Combinatorica **24** (2004), 271–286.

[HM2] N. Hindman and R. McCutcheon, *One sided ideals and Carlson's theorem*, Proc. Amer. Math. Soc. **130** (2002), 2559–2567.

[HS] N. Hindman and D. Strauss, *Algebra in the Stone–Čech compactification: theory and applications*, Walter de Gruyter, Berlin, 1998.

[Ka1] V. Kanellopoulos, *A proof of W. T. Gowers' c_0 theorem*, Proc. Amer. Math. Soc. **132** (2004), 3231–3242.

[Ka2] V. Kanellopoulos, *Ramsey families of subtrees of the dyadic tree*, Trans. Amer. Math. Soc. **357** (2005), 3865–3886.

[Ke] A. S. Kechris, *Classical Descriptive Set Theory*, Graduate Texts in Mathematics, Vol. 156, Springer, 1995.

[KSSS] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi, *The regularity lemma and its applications in graph theory*, in "Theoretical Aspects of Computer Science", Lecture Notes in Computer Science, Vol. 2292, Springer, 2002, 84–112.

[KS] J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, in "Combinatorics: Paul Erdős is Eighty", Vol. 2 (Keszthely, 1993), Bolyai Society Mathematical Studies, Vol. 2, János Bolyai Mathematical Society, 1996, 295–352.

[L] R. Laver, *Products of infinitely many perfect trees*, J. London Math. Soc. **29** (1984), 385–396.

[LSV] A. Louveau, S. Shelah and B. Veličković, *Borel partitions of infinite trees of a perfect tree*, Ann. Pure App. Logic **63** (1993), 271–281.

[LS] L. Lovász and B. Szegedy, *Szemerédi's lemma for the analyst*, Geom. Funct. Anal. **17** (2007), 252–270.

[Lu] D. Lubell, *A short proof of Sperner's lemma*, J. Comb. Theory **1** (1966), 299.

[McC1] R. McCutcheon, *Elemental Methods in Ergodic Ramsey Theory*, Lecture Notes in Mathematics, Vol. 1722, Springer, 1999.

[McC2] R. McCutcheon, *Two new extensions of the Hales–Jewett theorem*, Electron. J. Comb. **7** (2000), Research Paper R49, 1–18.

[M1] K. Milliken, *Ramsey's theorem with sums or unions*, J. Comb. Theory, Ser. A **18** (1975), 276–290.

[M2] K. Milliken, *A Ramsey theorem for trees*, J. Comb. Theory, Ser. A **26** (1979), 215–237.

[M3] K. Milliken, *A partition theorem for the infinite subtrees of a tree*, Trans. Amer. Math. Soc. **263** (1981), 137–148.

[NRS] B. Nagle, V. Rödl and M. Schacht, *The counting lemma for regular k-uniform hypergraphs*, Random Struct. Algorithms **28** (2006), 113–179.

[NW] C. St. J. A. Nash-Williams, *On well-quasi-ordering transfinite sequences*, Proc. Camb. Philos. Soc. **61** (1965), 33–39.

[OBr] K. O'Bryant, *Sets of integers that do not contain long arithmetic progressions*, Electron. J. Comb. **18** (2011), Research Paper P59, 1–15.

[PST] J. Pach, J. Solymosi and G. Tardos, *Remarks on a Ramsey theory for trees*, Combinatorica **32** (2012), 473–482.

[Pat] A. L. T. Paterson, *Amenability*, Mathematical Surveys and Monographs, Vol. 29, Amererican Mathematical Society, 1988.

[Paw] J. Pawlikowski, *Parameterized Ellentuck theorem*, Topology Appl. **37** (1990), 65–73.

[P] D. H. J. Polymath, *A new proof of the density Hales–Jewett theorem*, Ann. Math. **175** (2012), 1283–1327.

[Rado1] R. Rado, *Studien zur Kombinatorik*, Math Z. **36** (1933), 424–480.

[Rado2] R. Rado, *Note on combinatorial analysis*, Proc. London Math. Soc. **48** (1943), 122–160.

[Ra] F. P. Ramsey, *On a problem of formal logic*, Proc. London Math. Soc. **30** (1930), 264–286.

[Re] A. Rényi, *Foundations of Probability*, Holden-Day Series in Probability and Statistics, 1970.

[RSc1] V. Rödl and M. Schacht, *Regular partitions of hypergraphs: regularity lemmas*, Combin. Probab. Comput. **16** (2007), 833–885.

[RSc2] V. Rödl and M. Schacht, *Regular partitions of hypergraphs: counting lemmas*, Combin. Probab. Comput. **16** (2007), 887–901.

[RSc3] V. Rödl and M. Schacht, *Generalizations of the removal lemma*, Combinatorica **29** (2009), 467–501.

[RSTT] V. Rödl, M. Schacht, E. Tengan and N. Tokushige, *Density theorems and extremal hypergraph problems*, Israel J. Math. **152** (2006), 371–380.

[RSk] V. Rödl and J. Skokan, *Regularity lemma for k-uniform hypergraphs*, Random Struct. Algorithms **25** (2004), 1–42.

[Ros] H. E. Rose, *Subrecursion: functions and hierarchies*, Oxford Logic Guide 9, Oxford University Press, Oxford, 1984.

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

[Ru] W. Rudin, *Principles of Mathematical Analysis*, 3rd edition, International Series in Pure and Applied Mathematics, McGraw–Hill, 1976.

[RS] I. Z. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles*, in "Combinatorics" (Keszthely, 1976), Colloquia Mathematics Societies János Bolyai, Vol. 18, 1978, 939–945.

[Sa1] J. H. Sanders, *A generalization of Schur's theorem*, Thesis, Yale University, 1968.

[Sa2] T. Sanders, *On Roth's theorem on progressions*, Ann. Math. **174** (2011), 619–636.

[Sh1] S. Shelah, *Primitive recursive bounds for van der Waerden numbers*, J. Amer. Math. Soc. **1** (1988), 683–697.

[Sh2] S. Shelah, *A partition theorem*, Sci. Math. Jpn. **56** (2002), 413–438.

[So] J. Solymosi, *A note on a question of Erdős and Graham*, Combin. Probab. Comput. **13** (2004), 263–267.

[Spe] J. Spencer, *Ramsey's theorem—A new lower bound*, J. Comb. Theory, Ser. A **18** (1975), 108115.

[Sp] E. Sperner, *Ein Satz über Untermengen einer endlichen Menge*, Math. Z. **27** (1928) 544–548.

[St] J. Stern, *A Ramsey theorem for trees with an application to Banach spaces*, Israel J. Math. **29** (1978), 179–188.

[Sz1] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

[Sz2] E. Szemerédi, *Regular partitions of graphs*, in "Problèmes combinatoires et théorie des graphes" (Colloq. Internat. du CNRS, Univ. Orsay, Orsay, 1976), Colloques Internationaux du CNRS, Vol. 260, CNRS, 1978, 399–401.

[Tao1] T. Tao, *A variant of the hypergraph removal lemma*, J. Comb. Theory, Ser. A **113** (2006), 1257–1280.

[Tao2] T. Tao, *Szemerédi's regularity lemma revisited*, Contrib. Discrete Math. **1** (2006), 8–28.

[Tao3] T. Tao, *A correspondence principle between (hyper)graph theory and probability theory, and the (hyper)graph removal lemma*, J. Anal. Math. **103** (2007), 1–45.

[Tao4] T. Tao, *An Epsilon of Room, II: Pages from Year Three of a Mathematical Blog*, American Mathematical Society, 2011.

[TV] T. Tao and V. Vu, *Additive Combinatorics*, Cambridge Studies in Advanced Mathematics, Vol. 105, Cambridge University Press, 2006.

[Tay1] A. D. Taylor, *A canonical partition relation for the finite subsets of ω*, J. Comb. Theory, Ser. A **21** (1976), 137–146.

[Tay2] A. D. Taylor, *Bounds for the disjoint unions theorem*, J. Comb. Theory, Ser. A **30** (1981), 339–344.

[Th] A. Thomason, *An upper bound for some Ramsey numbers*, J. Graph Theory **12** (1988), 509–517.

[To] S. Todorcevic, *Introduction to Ramsey Spaces*, Annals of Mathematics Studies, No. 174, Princeton University Press, 2010.

[Ty] K. Tyros, *On colorings of variable words*, Discr. Math. **338** (2015), 1025–1028.

[vdW] B. L. van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw. Arch. Wisk. **15** (1927), 212–216.

[W] M. Walters, *Combinatorial proofs of the polynomial van der Waerden and the polynomial Hales–Jewett theorem*, J. London Math. Soc. **61** (2000), 1–12.

Index