Mauro Di Nasso, Isaac Goldbring, and Martino Lupini

Nonstandard Methods in Ramsey Theory and Combinatorial Number Theory

August 7, 2017

Springer

To...

Acknowledgements

To be filled in later...

Introduction

Generally speaking, Ramsey theory studies which combinatorial configurations of a structure are can always be found in one of the pieces of a given finite partition. More generally, it considers the problem of which combinatorial configurations can be found in sets that are "large" in some suitable sense. Dating back to the foundational results of van der Waerden, Ramsey, Erdős, Turán and others from the 1920s and 1930s, Ramsey theory has since then had an extraordinary development. On the one had, many applications of Ramsey theory have been found to numerous other areas of mathematics, ranging from functional analysis, topology, and dynamics, to set theory, model theory, and computer science. On the other hand, results and methods from other areas of mathematics have been successfully applied to establish new results in Ramsey theory. For instance, ergodic theory and the theory of recurrence in measurable dynamics has had a huge impact on Ramsey theory, to the point of giving rise to the research area of "ergodic Ramsey theory." Perhaps the best known achievement of this approach is the ergodic-theoretic proof of Szemerédi's theorem due to Furstenberg in the 1980s. In a different (but intimately related) direction, the theory of ultrafilters has been an important source of methods and ideas for Ramsey theory. In particular, the study of topological and algebraic properties of the space of ultrafilters has been used to give short and elegant proofs of deep combinatorial pigeonhole principles. Paradigmatic in this direction is the Galvin–Glazer ultrafilter proof of Hindman's theorem on sets of finite sums, previously established by Hindman in 1974 via a delicate, purely combinatorial argument.

Recently, a new thread of research has emerged, where problems in Ramsey theory are studied from the perspective of nonstandard analysis and nonstandard methods. Developed by Abraham Robinson in the 1960s and based on first order logic and model theory, nonstandard analysis provided a formal and rigorous treatment of calculus and classical analysis via infinitesimals, an approach more similar in spirit to the approach originally taken in the development of calculus in the 17th and 18th century, and avoids the epsilon-delta arguments that are inherent in its later formalization due to Weierstrass. While this is perhaps its most well known application, nonstandard analysis is actually much more versatile. The foundations of nonstandard analysis provide an approach, which we shall call the nonstandard method, that is applicable to virtually any area of mathematics. The nonstandard method has thus far been used in numerous areas of mathematics, including functional analysis, measure theory, ergodic theory, differential equations, and stochastic analysis, just to name a few such areas.

In a nutshell, the nonstandard method allows one to extend the given mathematical universe and thus regard it as contained in a much richer nonstandard universe. Such a nonstandard universe satisfies strong saturation properties which in particular allow one to consider limiting objects which do not exist in the standard universe. This procedure is similar to passing to an ultrapower, and in fact the nonstandard method can also be seen as a way to axiomatize the ultrapower construction in a way that distillates its essential features and benefits, but avoids being bogged down by the irrelevant details of its concrete implementation. This limiting process allows one to reformulate a given problem involving finite (but arbitrarily large) structures or configurations into a problem involving a single structure or configuration which is infinite but for all purposes behaves as though it were finite (in the precise sense that it is hyperfinite in the nonstandard universe). This reformulation can then be tackled directly using finitary methods, ranging from combinatorial counting arguments to recurrence theorems for measurable dynamics, recast in the nonstandard universe.

In the setting of Ramsey theory and combinatorics, the application of nonstandard methods had been pioneered by the work of Keisler, Leth, and Jin from the 1980s and 1990s. These applications had focused on density problems in combinatorial number theory. The general goal in this area is to establish the existence of combinatorial configurations in sets that are large in that sense that they have positive asymptotic density. For example, the aforementioned celebrated theorem of Szmerédi from 1970 asserts that a set of integers of positive density contains arbitrarily long finite arithmetic progressions. One of the contributions of the nonstandard approach is to translate the notion of asymptotic density on the integers, which does not satisfies all the properties of a measure, into an actual measure in the nonstandard universe. This translation then makes methods from measure theory and ergodic theory, such as the ergodic theorem or other recurrence theorems, available for the study of density problems. In a sense, this can be seen as a version of Furstenberg's correspondence (between sets of integers and measurable sets in a dynamical system), with the extra feature that the dynamical system obtained perfectly reflects *all* the combinatorial properties of the set that one started with. The achievements of the nonstandard approach in this area include the work of Leth on arithmetic progressions in sparse sets, Jin's theorem on sumsets, as well as Jin's Freiman-type results on inverse problems for sumsets. More recently, these methods have also been used by Jin, Leth, Mahlburg, and the present authors to tackle a conjecture of Erdős concerning sums of infinite sets (the so-called $B+C$ conjecture).

Nonstandard methods are also tightly connected with ultrafilter methods. This has been made precise and successfully applied in recent work of Di Nasso, where he observed that there is a perfect correspondence between ultrafilters and elements of the nonstandard universe up to a natural notion of equivalence. One the one hand, this allows one to manipulate ultrafilters as nonstandard points, and to use ultrafilter methods to prove the existence of certain combinatorial configurations in the nonstandard universe. One the other hand, this gives an intuitive and direct way to infer, from the existence of certain ultrafilter configurations, the existence of corresponding standard combinatorial configuration via the fundamental principle of transfer in the nonstandard method. This perspective has successfully been applied by Di Nasso and co-authors to the study of partition regularity problems for Diophantine equations over the integers, providing in particular a far-reaching generalization of the classical theorem of Rado on partition regularity of systems of linear equations. Unlike Rado's theorem, this recent generalization also includes equations that are *not* linear.

Finally, it is worth mentioning that many other results in combinatorics can be seen, directly or indirectly, as applications of the nonstandard method. For instance, the groundbreaking work of Hrushovski and Breuillard–Green–Tao on approximate groups, although not originally presented in this way, admit a natural nonstandard treatment. The same applies to the work of Bergelson and Tao on recurrence in quasirandom groups.

The goal of this present manuscript is to introduce the uninitiated reader to the nonstandard method and to provide an overview of its most prominent applications in Ramsey theory and combinatorial number theory. In particular, no previous knowledge of nonstandard analysis will be assumed. Instead, we will provide a complete and self-contained introduction to the nonstandard method in the first part of this book. Novel to our introduction is a treatment of the topic of iterated hyperextensions, which is crucial for some applications and has thus far appeared only in specialized research articles. The intended audience for this book include researchers in combinatorics that desire to get acquainted with the nonstandard approach, as well as experts of nonstandard analysis who have been working in this or other areas of research. The list of applications of the nonstandard method to combinatorics and Ramsey theory presented here is quite extensive, including cornerstone results of Ramsey theory such as Ramsey's theorem, Hindman's theorem on sets of finite sums, the Hales–Jewett theorem on variable words, and Gowers' theorem on FIN_k. It then proceeds with results on partition regularity of diophantine equations and with density problems in combinatorial number theory. A nonstandard treatment of the triangle removal lemma, the Szemerédi regularity lemma, and of the already mentioned work of Hrushovski and Breuillard–Green–Tao on approximate groups conclude the book. We hope that such a complete list of examples will help the reader unfamiliar with the nonstandard method get a good grasp on how the approach works and can be applied. At the same time, we believe that collecting these results together, and providing a unified presentation and approach, will provide a useful reference for researchers in the field and will further stimulate the research in this currently very active area.

Pisa, Italy
Irvine, California
Pasadena, California

Notation and Conventions

We set $\mathbb{N} := \{1,2,3,\ldots\}$ to denote the set of *positive* natural numbers and $\mathbb{N}_0 := \{0,1,2,3,\ldots\}$ to denote the set of natural numbers.

We use the following conventions for elements of particular sets:

- m and n always range over $\mathbb{N}$;
- k and l always range over $\mathbb{Z}$;
- H,K,M, and N always range over infinite elements of ${}^*\mathbb{N}$;
- δ and ε always denote (small) positive real numbers, while ε denotes a positive infinitesimal element of ${}^*\mathbb{R}$;
- Given any set S, we let α, β, and γ denote arbitrary (possibly standard) elements of ${}^*S$;

$\mathrm{Fin}(X) = \{F \subseteq X \mid X \text{ is finite}\}$.

For any n, we write $[n] := \{1,\ldots,n\}$. Similarly, we write $[N] := \{1,\ldots,N\}$.

Given any nonempty finite set I and any set A, we write $\delta(A,I) := \frac{|A\cap I|}{|I|}$. We extend this to the nonstandard situation: if I is a nonempty hyperfinite set and A is an internal set, we set $\delta(A,I) := \frac{|A\cap |I|}{|I|}$. We also write $\delta(A,n) := \delta(A,[n])$ and $\delta(A,N) := \delta(A,[N])$.

Given a hyperfinite set X, we let $\mathscr{L}_X$ denote the σ-algebra of Loeb measurable subsets of X and we let μ_X denote the Loeb measure on $\mathscr{L}_X$ that extends the normalized counting measure on X. (See Chapter 6.) When $X = [1,N]$, we write $\mathscr{L}_N$ and μ_N instead of $\mathscr{L}_X$ and μ_X. If A is also internal, we write $\mu_X(A) := \mu_X(A\cap X)$.

Suppose that $A \subseteq \mathbb{Z}$ and $k \in \mathbb{N}$. We write

$$\Sigma_k(A) := \{x_1 + \cdots + x_k \ : \ x_1,\ldots,x_k \in A\}$$

and

$$kA := \{kx \ : \ x \in A\}.$$

Of course $k \cdot A \subseteq \Sigma_k(A)$.

Throughout, log always denotes log base 2.

Contents

Part I
Preliminaries

Chapter 1
Ultrafilters

1.1 Basics on ultrafilters

Throughout this chapter, we let S denote an infinite set.

Definition 1.1. A (proper) *filter* on S is a set $\mathscr{F}$ of subsets of S (that is, $\mathscr{F} \subseteq \mathscr{P}(S)$) such that:

- $\emptyset \notin \mathscr{F}$, $S \in \mathscr{F}$;
- if $A, B \in \mathscr{F}$, then $A \cap B \in \mathscr{F}$;
- if $A \in \mathscr{F}$ and $A \subseteq B$, then $B \in \mathscr{F}$.

We think of elements of $\mathscr{F}$ as "big" sets (because that is what filters do, they catch the big objects). The first and third axioms are (hopefully) intuitive properties of big sets. Perhaps the second axiom is not as intuitive, but if one thinks of the complement of a big set as a "small" set, then the second axiom asserts that the union of two small sets is small (which is hopefully more intuitive).

Exercise 1.2. Set $\mathscr{F} := \{A \subseteq S \mid S \setminus A \text{ is finite}\}$. Prove that $\mathscr{F}$ is a filter on S, called the *Frechet* or *cofinite* filter on S.

Exercise 1.3. Suppose that $\mathscr{D}$ is a set of subsets of S with the *finite intersection property*: whenever $D_1, \ldots, D_n \in \mathscr{D}$, we have $D_1 \cap \cdots \cap D_n \neq \emptyset$. Set

$$\langle \mathscr{D} \rangle := \{E \subseteq S \mid D_1 \cap \cdots D_n \subseteq E \text{ for some } D_1, \ldots, D_n \in \mathscr{D}\}.$$

Show that $\langle \mathscr{D} \rangle$ is the smallest filter on S containing $\mathscr{D}$, called the *filter generated by* $\mathscr{D}$.

If $\mathscr{F}$ is a filter on S, then a subset of S cannot be simultaneously big and small (that is, both it and its complement belong to $\mathscr{F}$), but there is no requirement that it be one of the two. It will be desirable (for reasons that will become clear in a moment) to add this as an additional property:

Definition 1.4. If $\mathscr{F}$ is a filter on S, then $\mathscr{F}$ is an *ultrafilter* if, for any $A \subseteq S$, either $A \in \mathscr{F}$ or $S \setminus A \in \mathscr{F}$ (but not both!).

Ultrafilters are usually denoted by $\mathscr{U}$. Observe that the Frechet filter on S is not an ultrafilter since there are sets $A \subseteq S$ such that A and $S \setminus A$ are both infinite.

The following exercise illustrates one of the most important properties of ultrafilters .

Exercise 1.5. Suppose that $\mathscr{U}$ is an ultrafilter on S and $A_1, \ldots, A_n$ are pairwise disjoint subsets of S such that $A_1 \cup \cdots \cup A_n \in \mathscr{U}$. Prove that there is a unique $i \in \{1, \ldots, n\}$ such that $A_i \in \mathscr{U}$.

We have yet to see an example of an ultrafilter. Here is a "trivial" source of ultrafilters:

Definition 1.6. Given $s \in S$, set $\mathscr{F}_s := \{A \subseteq S \mid s \in A\}$.

Exercise 1.7. For $s \in S$, prove that $\mathscr{F}_s$ is an ultrafilter on S, called the *principal ultrafilter generated by* s.

We say that an ultrafilter $\mathscr{U}$ on S is *principal* if $\mathscr{U} = \mathscr{F}_s$ for some $s \in S$. Although principal ultrafilters settle the question of the existence of ultrafilters, they will turn out to be useless for our purposes, as we will soon see. From a philosophical viewpoint, principal ultrafilters failure to capture the idea that sets belonging to the ultrafilter are large, for $\{s\}$ belongs to the ultrafilter $\mathscr{F}_s$ and yet hardly anyone would dare say that the set $\{s\}$ is large!

Exercise 1.8. Prove that an ultrafilter $\mathscr{U}$ on S is principal if and only if there is a *finite* set $A \subseteq S$ such that $A \in \mathscr{U}$.

We now would like to prove the existence of nonprincipal ultrafilters. The following exercise will be the key to doing this.

Exercise 1.9. Suppose that $\mathscr{F}$ is a filter on S. Then $\mathscr{F}$ is an ultrafilter on S if and only if it is a maximal filter, that is, if and only if, whenever $\mathscr{F}'$ is a filter on S such that $\mathscr{F} \subseteq \mathscr{F}'$, we have $\mathscr{F} = \mathscr{F}'$.

Since it is readily verified that the union of an increasing chain of filters on S containing the Frechet filter on S is once again a filter on S containing the Frechet filter on S, the previous exercise and Zorn's lemma yields the following:

Corollary 1.10. *Nonprincipal ultrafilters on S exist. In fact, given any filter $\mathscr{F}$ on S extending the Frechet filter on S, there is a nonprincipal ultrafilter on S containing $\mathscr{F}$.*

Exercise 1.11. Suppoes that $f : S \to T$ is a function between sets. Then given any ultrafilter $\mathscr{U}$ on S, the set

$$f(\mathscr{U}) := \{A \subseteq T \ : \ f^{-1}(A) \in \mathscr{U}\}$$

is an ultrafilter on T, called the *image ultrafilter of $\mathscr{U}$ under f*.

1.2 The space of ultrafilters βS

In this section, S continues to denote an infinite set. Since topological matters are the subject of this subsection, we will also treat S as a topological space equipped with the discrete topology.

The set of ultrafilters on S is denoted βS. There is a natural topology on βS by declaring, for $A \subseteq S$, the sets $U_A := \{\mathscr{U} \in \beta S \ : \ A \in \mathscr{U}\}$ as basic open sets. (Note that the U_A's are indeed a base for a topology as $U_A \cap U_B = U_{A\cap B}$.) Since the complement of U_A in βS is $U_{S\setminus A}$, we see that the basic open sets are in fact clopen. Note also that βS is Hausdorff: if $\mathscr{U}, \mathscr{V} \in \beta S$ are distinct, take $A \subseteq S$ with $A \in \mathscr{U}$ and $S \setminus A \in \mathscr{V}$; then $\mathscr{U} \in U_A$ and $\mathscr{V} \in U_{S\setminus A}$ and clearly U_A and $U_{S\setminus A}$ are disjoint.

Theorem 1.12. *βS is a compact space.*

Proof. It is enough to show that every covering of βS by basic open sets has a finite subcover. Let (A_i) be a family of subsets of S such that (U_{A_i}) covers βS. Suppose, towards a contradiction, that this cover of βS has no finite subcover. We claim then that $(S \setminus A_i)$ has the finite intersection property. Indeed, given $J \subseteq I$ finite, there is $\mathscr{U} \in \beta S \setminus \bigcup_{i\in J} U_{A_i}$, whence $S \setminus A_i \in \mathscr{U}$ for each $i \in J$, and hence $\bigcap_{i\in J}(S \setminus A_i) \neq \emptyset$. It follows that there is a $\mathscr{U} \in \beta S$ such that $S \setminus A_i \in \mathscr{U}$ for all $i \in I$, contradicting the fact that $\mathscr{U} \in U_{A_i}$ for some $i \in I$.

We identify S with the set of principal ultrafilters on S; under this identification, S is dense in βS: if $A \subseteq S$ is nonemtpy and $s \in A$, then $\mathscr{U}_s \in U_A$. Thus, βS is a compactification of S. In fact, we have:

Theorem 1.13. *βS is the Stone-Cech compactification of S.*

We reminder the reader that the Stone-Cech compactification of S is the unique compactifiction X of S with the following property: any function $f : S \to Y$ with Y compact Hausdorff has a unique extension $\tilde{f} : X \to Y$. In order to prove the previous theorem, we will first need the following lemma, which is important in its own right:

Lemma 1.14. *Suppose that Y is a compact Hausdorff space and $(y_s)_{s\in S}$ is a familiy of elements of Y indexed by S. Then for any $\mathscr{U} \in \beta S$, there is a unique element $y \in Y$ with the property that, for any open neighborhood U of y, we have $\{s \in S \ : \ y_s \in U\} \in \mathscr{U}$.*

Proof. Suppose, towards a contradiction, that no such y exists. Then for every $y \in Y$, there is an open neighborhood U_y of y such that $\{s \in S \,:\, y_s \in U_y\} \notin \mathscr{U}$. By compactness, there are $y_1, \ldots, y_n \in Y$ such that $Y = U_{y_1} \cup \cdots \cup U_{y_n}$. There is then a unique $i \in \{1, \ldots, n\}$ such that $\{s \in S \,:\, y_s \in U_{y_i}\} \in \mathscr{U}$, yielding the desired contradiction.

The uniqueness of y follows from the fact that Y is Hausdorff together with the fact that $\mathscr{U}$ does not contain two disjoint sets.

Definition 1.15. In the context of the previous lemma, we call the unique y the *ultralimit of* (y_s) *with respect to* $\mathscr{U}$, denoted $\lim_{s,\mathscr{U}} y_s$ or simply just $\lim_{\mathscr{U}} y_s$.

Proof (of Theorem 1.13). Suppose that $f : S \to Y$ is a function into a compact Hausdorff space. Define $\tilde{f} : \beta S \to Y$ by $\tilde{f}(\mathscr{U}) := \lim_{\mathscr{U}} f(s)$, which exists by Lemma 1.14. It is clear that $\tilde{f}(\mathscr{U}_s) = f(s)$, so $\tilde{f}$ extends f. We must show that $\tilde{f}$ is continuous. Fix $\mathscr{U} \in \beta S$ and let U be an open neighborhood of $\tilde{f}(\mathscr{U})$ in Y. Let $V \subseteq U$ be an open neighborhood of $\tilde{f}(\mathscr{U})$ in Y such that $\overline{V} \subseteq U$. Take $A \in \mathscr{U}$ such that $f(s) \in V$ for $s \in A$. Suppose $\mathscr{V} \in U_A$, so $A \in \mathscr{V}$; then $\lim_{\mathscr{V}} f(s) \in \overline{V} \subseteq U$, so $U_A \subseteq \tilde{f}^{-1}(U)$.

We leave it to the reader to verify that if $g : \beta S \to Y$ agrees with f on S, then $g = \tilde{f}$.

Now that we have shown that βS is the Stone-Cech compactification of S, given $f : S \to Y$ where Y is a compact Hausdorff space, we will let $\beta f : \beta S \to Y$ denote the unique continuous extension of f.

Definition 1.16. Fix $k \in \mathbb{N}$. Let $m_k : \mathbb{N} \to \mathbb{N}$ be defined by $m_k(n) := kn$. Then for $\mathscr{U} \in \beta\mathbb{N}$, we set $k\mathscr{U} := (\beta m_k)(\mathscr{U})$.

The ultrafilters $k\mathscr{U}$ will play an important role in Chapter 10.

Exercise 1.17. Given $A \subseteq S$, show that $\overline{A} = U_A$, where $\overline{A}$ denotes the closure of A in βS.

Let $B(S)$ denote the space of bounded real-valued functions on S. Given $f \in B(S)$, take $M \in \mathbb{R}^{>0}$ such that $f(S) \subseteq [-M, M]$, whence we may consider $\beta f : \beta S \to [-M, M]$. Note that the function βf does not depend on the choice of M. The following exercise will be useful in Chapter 13.

Exercise 1.18. The function $f \mapsto \beta f$ is an isomorphism between $B(S)$ and $C(\beta S)$ as Banach spaces.

1.3 The case of a semigroup

We now suppose that S is the underlying set of a semigroup $(S, \cdot)$. Then one can extend the semigroup operation $\cdot$ to an operation $\odot$ on βS by declaring, for $\mathscr{U}, \mathscr{V} \in \beta S$ and $A \subseteq S$, that

$$A \in \mathscr{U} \odot \mathscr{V} \Leftrightarrow \{s \in S \,:\, s^{-1}A \in \mathscr{V}\} \in \mathscr{U}.$$

Here, $s^{-1}A := \{t \in S \,:\, st \in A\}$. In other words, $\mathscr{U} \odot \mathscr{V} = \lim_{s,\mathscr{U}}(\lim_{t,\mathscr{V}} s \cdot t)$, where these limits are taken in the compact space βS. In particular, note that $\mathscr{U}_s \odot \mathscr{U}_t = \mathscr{U}_{s \cdot t}$, so this operation on βS does indeed extend the original operationon S. It is also important to note that, in general, ultralimits do not commutate and thus, in general, $\mathscr{U} \odot \mathscr{V} \neq \mathscr{V} \odot \mathscr{U}$, even if $(S, \cdot)$ is commutative. (See Chapter 3 for more on this lack of commutativity.)

The following theorem is the key to many applications of ultrafilter/nonstandard methods in Ramsey theory.

Theorem 1.19. $(\beta S, \odot)$ *is a compact,* semitopological semigroup*, that is, for each $\mathscr{V} \in \beta S$, the map $\mathscr{U} \mapsto \mathscr{U} \odot \mathscr{V} : \beta S \to \beta S$ is continuous.*

Proof. Fix $\mathscr{V} \in \beta S$ and let $\rho_{\mathscr{V}} : \beta S \to \beta S$ be defined by $\rho_{\mathscr{V}}(\mathscr{U}) := \mathscr{U} \odot \mathscr{V}$. We need to show that $\rho_{\mathscr{V}}$ is continuous. Towards this end, fix $A \subseteq S$; we must show that $\rho_{\mathscr{V}}^{-1}(U_A)$ is open. Let $B := \{s \in S \,:\, s^{-1}A \in \mathscr{V}\}$. It remains to note that $\rho_{\mathscr{V}}^{-1}(U_A) = U_B$.

In the case that the semigroup operation is commutative, there is some notation that is often helpful. Indeed, let us suppose that the semigroup operation is commutative and denoted by $+$. Given $A \subseteq S$ and $\mathscr{U} \in \beta S$, let $A - \mathscr{U} := \{s \in S \,:\, A - s \in \mathscr{U}\}$, where, as a reminder, recall that $A - s := \{t \in S \,:\, s + t \in A\}$. Notice that this notation is consistent with our identification of S as the principal ultrafilters on S. Indeed,

$$A - \mathscr{U}_t = \{s \in S \,:\, A - s \in \mathscr{U}_t\} = \{s \in S \,:\, t \in A - s\} = A - t.$$

The operation $\oplus$ on βS can thus be defined by the intuitive formula

$$A \in \mathscr{U} \oplus \mathscr{V} \Leftrightarrow A - \mathscr{V} \in \mathscr{U}.$$

Chapter 2
Nonstandard analysis

If one wants to present the methods of nonstandard analysis in their full generality and with full rigor, then notions and tools from mathematical logic such as "first-order formula" or "elementary extension" are definitely needed. However, we believe that a gentle introduction to the basics of nonstandard methods and their use in combinatorics does not directly require any technical machinery from logic. Only at a later stage, when advanced nonstandard techniques are applied and their use must be put on firm foundations, logic will play its unavoidable role.

As a first preliminary step, one should become familiar with the primary features of the nonstandard versions of the natural, integer, rational, and real numbers, that will be named by adding the prefix "hyper". To this end, "partial" definitions, to be completed further on, will suffice. A second step is to introduce the fundamental principle of nonstandard analysis, namely the *transfer principle* of the *star map*, in a semiformal way, which even so can give a sufficiently clear picture of the basic notions and tools as used in practice.

In the appendix, we give sound and rigorous foundations to nonstandard analysis in its full generality by introducing the formal language of first order logic.

2.1 Warming-up

To begin with, let us recall the following notions, which are at the very base of nonstandard analysis. We note that they make sense in any ordered field (actually, in any ordered ring).

Definition 2.1. A number ε is *infinitesimal* (or *infinitely small*) if $|\varepsilon| < \frac{1}{n}$ for every $n \in \mathbb{N}$. A number Ω is *infinite* if $|\Omega| > n$ for every $n \in \mathbb{N}$.

Clearly, a nonzero number is infinite if and only if its reciprocal is infinitesimal. We say that a number is *finite* or *bounded* if it is not infinite.

Exercise 2.2.

1. If ξ and ζ are finite, then $\xi + \zeta$ and $\xi \cdot \zeta$ are finite.
2. If ξ and ζ are infinitesimal, then $\xi + \zeta$ is infinitesimal.
3. If ξ is infinitesimal and ζ is finite, then $\xi \cdot \zeta$ is infinitesimal.
4. If ξ is infinite and ζ is not infinitesimal, then $\xi \cdot \zeta$ is infinite.
5. If $\xi \neq 0$ is infinitesimal and ζ is not infinitesimal, then ξ / ζ is infinitesimal.
6. If ξ is infinite and ζ is finite, then ξ / ζ is infinite.

Recall that an ordered field $\mathbb{F}$ is *Archimedean* if for every positive $x \in \mathbb{F}$ there exists $n \in \mathbb{N}$ such that $nx > 1$.

Exercise 2.3. The following properties are equivalent for an ordered field $\mathbb{F}$:

1. $\mathbb{F}$ is non-Archimedean;

2. There are nonzero infinitesimal numbers in $\mathbb{F}$;
3. The set of natural numbers is bounded in $\mathbb{F}$.

We are now ready to introduce the nonstandard reals.

Definition 2.4. The *hyperreal numbers* $^*\mathbb{R}$ are a proper extension of the ordered field $\mathbb{R}$ that satisfies additional properties (to be specified further on).

By just using the above incomplete definition, the following is proved.

Proposition 2.5. *The hyperreal field $^*\mathbb{R}$ is non-Archimedean, and hence it contains nonzero infinitesimals and infinite numbers.*

Proof. Since $^*\mathbb{R}$ is a proper extension of the real field, we can pick a number $\xi \in {}^*\mathbb{R} \setminus \mathbb{R}$. Without loss of generality, let us assume $\xi > 0$. If ξ is infinite, then we are done. Otherwise, by the completeness property of $\mathbb{R}$, we can consider the number $r = \inf\{x \in \mathbb{R} \mid x > \xi\}$. (Notice that it may be $r < \xi$.) It is readily checked that $\xi - r$ is a nonzero infinitesimal number.

We remark that, as a non-Archimedean field, $^*\mathbb{R}$ is *not* complete (e.g., the set of infinitesimals is bounded but has no least upper bound). The nonstandard counterpart of completeness is given by the following property. We say that two real numbers are *infinitely close* if their difference is infinitesimal.

Theorem 2.6 (Standard Part). *Every finite hyperreal number $\xi \in {}^*\mathbb{R}$ is infinitely close to a unique real number $r \in \mathbb{R}$, called the* standard part *of ξ. In this case, we use the notation $r = \mathrm{st}(\xi)$.*

Proof. By the completeness of $\mathbb{R}$, we can set $\mathrm{st}(\xi) := \inf\{x \in \mathbb{R} \mid x > \xi\} = \sup\{y \in \mathbb{R} \mid y < r\}$. By the supremum (or infimum) property, it directly follows that $\mathrm{st}(\xi)$ is infinitely close to ξ. Moreover, $\mathrm{st}(\xi)$ is the unique real number with that property, since infinitely close real numbers are necessarily equal.

It follows that every finite hyperreal number ξ has a unique representation in the form $\xi = r + \varepsilon$ where $r = \mathrm{st}(\xi) \in \mathbb{R}$ and $\varepsilon = \xi - \mathrm{st}(\xi)$ is infinitesimal.

The following are the counterparts in the nonstandard setting of the familiar properties of limits of real sequences.

Exercise 2.7. For all finite hyperreal numbers ξ, ζ:

1. $\mathrm{st}(\xi) < \mathrm{st}(\zeta) \Rightarrow \xi < \zeta \Rightarrow \mathrm{st}(\xi) \leq \mathrm{st}(\zeta)$;
2. $\mathrm{st}(\xi + \zeta) = \mathrm{st}(\xi) + \mathrm{st}(\zeta)$;
3. $\mathrm{st}(\xi \cdot \zeta) = \mathrm{st}(\xi) \cdot \mathrm{st}(\zeta)$;
4. $\mathrm{st}(\frac{\xi}{\zeta}) = \frac{\mathrm{st}(\xi)}{\mathrm{st}(\zeta)}$ whenever ζ is not infinitesimal.

Definition 2.8. The *hyperinteger numbers* $^*\mathbb{Z}$ are an unbounded discretely ordered subring of $^*\mathbb{R}$ that satisfies special properties (to be specified further on), including the following:

- For every $\xi \in {}^*\mathbb{R}$ there exists $\zeta \in {}^*\mathbb{Z}$ with $\zeta \leq \xi < \zeta + 1$. Such a ζ is called the *hyperinteger part* of ξ, denoted $\zeta = \lfloor \xi \rfloor$.

Since $^*\mathbb{Z}$ is discretely ordered, notice that its finite part coincides with $\mathbb{Z}$. This means that for every $z \in \mathbb{Z}$ there are *no* hyperintegers $\zeta \in {}^*\mathbb{Z}$ such that $z < \zeta < z + 1$.

Definition 2.9. The *hypernatural numbers* $^*\mathbb{N}$ are the positive part of $^*\mathbb{Z}$; thus $^*\mathbb{Z} = -{}^*\mathbb{N} \cup \{0\} \cup {}^*\mathbb{N}$, where $-{}^*\mathbb{N} = \{-\xi \mid \xi \in {}^*\mathbb{N}\}$ are the negative hyperintegers.

Definition 2.10. The field of *hyperrational numbers* $^*\mathbb{Q}$ is the quotient field of $^*\mathbb{Z}$; thus hyperrational numbers $\zeta \in {}^*\mathbb{Q}$ can be represented as ratios $\zeta = \frac{\xi}{\nu}$ where $\xi \in {}^*\mathbb{Z}$ and $\nu \in {}^*\mathbb{N}$.

Exercise 2.11. The hyperrational numbers $^*\mathbb{Q}$ are dense in $^*\mathbb{R}$, that is, for every pair $\xi < \xi'$ in $^*\mathbb{R}$ there exists $\eta \in {}^*\mathbb{Q}$ such that $\xi < \eta < \xi'$.

We remark that, although still incomplete, our definitions suffice to get a clear picture of the order-structure of the two main nonstandard objects that we will consider here, namely the hypernatural numbers ${}^*\mathbb{N}$ and the hyperreal line ${}^*\mathbb{R}$. In particular, let us focus on the nonstandard natural numbers. One possible way (but certainly not the only possible way) to visualize them is the following:

- The *hypernatural numbers* ${}^*\mathbb{N}$ are the extended version of the natural numbers that is obtained by allowing the use of a "mental telescope" to also see infinite numbers beyond the finite ones.

So, beyond the usual finite numbers $\mathbb{N} = \{1,2,3,\dots\}$, one finds infinite numbers $\xi > n$ for all $n \in \mathbb{N}$. Every $\xi \in {}^*\mathbb{N}$ has a successor $\xi + 1$, and every non-zero $\xi \in {}^*\mathbb{N}$ has a predecessor $\xi - 1$.

$$ {}^*\mathbb{N} \;=\; \{\underbrace{1,2,3,\dots,n,\dots}_{\text{finite numbers}} \quad \underbrace{\dots,N-2,N-1,N,N+1,N+2,\dots}_{\text{infinite numbers}}\} $$

Thus the set of finite numbers $\mathbb{N}$ does not have a greatest element and the set of infinite numbers ${}^*\mathbb{N} \setminus \mathbb{N}$ does not have a least element, whence ${}^*\mathbb{N}$ is *not* well-ordered.

Exercise 2.12. Consider the equivalence relation $\sim_f$ on ${}^*\mathbb{N}$ defined by setting $\xi \sim_f \zeta$ if $\xi - \zeta$ is finite. Show that the quotient set ${}^*\mathbb{N}/{\sim_f}$ of the corresponding equivalence classes, called *galaxies*, inherits the order structure of a dense set with a least element $[1] = \mathbb{N}$ and with no greatest element.

2.2 The star map and the transfer principle

As we have seen in the previous section, corresponding to each of the numerical sets $\mathbb{N},\mathbb{Z},\mathbb{Q},\mathbb{R}$, one has a *nonstandard extension*, namely the sets ${}^*\mathbb{N},{}^*\mathbb{Z},{}^*\mathbb{Q},{}^*\mathbb{R}$, respectively. A defining feature of nonstandard analysis is that one has a canonical way of extending *every* mathematical object A under study to an object ${}^*A$ which inherits all "elementary" properties of the initial object.

Definition 2.13. The *star map* is a function that associates to each "mathematical object" A under study its *hyper-extension* (or *nonstandard extension*) ${}^*A$ in such a way that the following holds:

- *Transfer principle:* Let $P(A_1,\dots,A_n)$ be an "elementary property" of the mathematical objects $A_1,\dots,A_n$. Then $P(A_1,\dots,A_n)$ is true if and only if $P({}^*A_1,\dots,{}^*A_n)$ is true:

$$ P(A_1,\dots,A_n) \iff P({}^*A_1,\dots,{}^*A_n). $$

One can think of hyper-extensions as a sort of weakly isomorphic copy of the initial objects; indeed, by the *transfer principle*, an object A and its hyper-extension ${}^*A$ are indistinguishable as far as their "elementary properties" are concerned. Of course, the crucial point here is to see precisely which properties are "elementary" and which are not.

Let us remark that the above definition is indeed incomplete in that the notions of "mathematical object" and of "elementary property" are still to be made precise and rigorous. As anticipated in the introduction, we will do this gradually.

To begin with, it will be enough to include in our considered "mathematical objects" the following:

1. Real numbers and tuples of real numbers;
2. All sets $A \subseteq \mathbb{R}^k$ of real tuples, and all functions $f : A \to B$ between them;
3. All sets made up of objects in (1) and (2), including, *e.g.*, the families $\mathscr{F} \subseteq \bigcup_k \mathscr{P}(\mathbb{R}^k)$ of sets of real k-tuples, and the families of functions $\mathscr{G} \subseteq \mathrm{Fun}(\mathbb{R}^k,\mathbb{R}^h)$.

More generally, every structure under study could be safely taken as one of our "mathematical objects".[1]

[1] According to the usual set-theoretic foundational framework, every mathematical object is identified with a set (see the remarks in Section XXX). However, here we will stick to the common perception that considers numbers, ordered pairs, relations, functions, and sets as mathematical objects of distinct nature.

As for the notion of "elementary property", we will start working with a semi-formal definition. Although not fully rigorous from a logical point of view, it may nevertheless looks perfectly fine to many, and we believe that it can be safely adopted to get introduced to nonstandard analysis and to familiarize oneself with its basic notions and tools.

Definition 2.14. A property P is *elementary* if it can be expressed by an *elementary formula*, that is, by a formula where:

1. Besides the usual logical connectives ("not", "and", "or", "if ... then", "if and only if") and the quantifiers ("there exists", "for every") only the basic notions of equality, membership, set, ordered k-tuple, k-ary relation, domain, range, function, value of a function at a given point, are involved;
2. The scope of every quantifier is *bounded*, that is, quantifiers always occur in the form "there exists $x \in X$" or "for every $y \in Y$" for specified sets X,Y. More generally, also nested quantifiers "$Qx_1 \in x_2$ and $Qx_2 \in x_3$ ... and $Qx_n \in X$" are allowed, where Q is either "there exists" or "for every", $x_1,\dots,x_n$ are variables, and X is a specified set.

An immediate consequence of the *transfer* principle is that all fundamental mathematical constructions are preserved under the star map, with the only two relevant exceptions being *powersets* and *function sets* (see Example XXX). Below we give three comprehensive lists in three distinct propositions, the first one about sets and ordered tuples, the second one about relations, and the third one about functions. As long as the notion of "elementary property" has not been made rigidly precise, the suspicious reader who does not accept the proofs may directly take those properties as axioms of the star map.

Proposition 2.15.

1. $a = b \Leftrightarrow {}^*a = {}^*b$.
2. $a \in A \Leftrightarrow {}^*a \in {}^*A$.
3. *A is a set if and only if ${}^*A$ is a set.*
4. ${}^*\emptyset = \emptyset$.

If $A, A_1, \dots, A_k, B$ are sets:

5. $A \subseteq B \Leftrightarrow {}^*A \subseteq {}^*B$.
6. ${}^*(A \cup B) = {}^*A \cup {}^*B$.
7. ${}^*(A \cap B) = {}^*A \cap {}^*B$.
8. ${}^*(A \setminus B) = {}^*A \setminus {}^*B$.
9. ${}^*\{a_1,\dots,a_k\} = \{{}^*a_1,\dots,{}^*a_k\}$.
10. ${}^*(a_1,\dots,a_k) = ({}^*a_1,\dots,{}^*a_k)$.
11. ${}^*(A_1 \times \dots \times A_k) = {}^*A_1 \times \dots \times {}^*A_k$.
12. ${}^*\{(a,a) \mid a \in A\} = \{(\xi,\xi) \mid \xi \in {}^*A\}$.

If $\mathscr{F}$ is a family of sets:

13. ${}^*\{(x,y) \mid x \in y \in \mathscr{F}\} = \{(\xi,\zeta) \mid \xi \in \zeta \in {}^*\mathscr{F}\}$.
14. ${}^*(\bigcup_{F\in\mathscr{F}} F) = \bigcup_{G\in{}^*\mathscr{F}} G$.

Proof. Recall that by our definition, the notions of equality, membership, set, and ordered k-tuple are elementary; thus by direct applications of *transfer* one obtains (1), (2), (3), and (10), respectively. All other properties are easily proved by considering suitable elementary formulas. As examples, we will consider here only three of them.

(8). The property "$C = A \setminus B$" is elementary, because it is formalized by the elementary formula:

$$\text{“}\forall x \in C\ (x \in A \text{ or } x \notin B) \text{ and } \forall x \in A\ (x \notin B \Rightarrow x \in C)\text{”}.$$

So, by *transfer*, we have that $C = A \setminus B$ holds if and only if

$$\text{“}\forall x \in {}^*C\ (x \in {}^*A \text{ or } x \notin {}^*B) \text{ and } \forall x \in {}^*A\ (x \notin {}^*B \Rightarrow x \in {}^*C)\text{”},$$

that is, if and only if ${}^*C = {}^*A \setminus {}^*B$.

(9) The property "$C = \{a_1,\dots,a_k\}$" is formalized by the elementary formula: "$a_1 \in C$ and ...... and $a_k \in C$ and $\forall x \in C\ (x = a_1$ or ... or $x = a_k)$". So, we can apply *transfer* and obtain that ${}^*C = \{{}^*a_1,\dots,{}^*a_k\}$.

(14). The property "$A = \bigcup_{F\in\mathscr{F}} F$" is formalized by the elementary formula: "$\forall x \in A\ (\exists y \in \mathscr{F}$ with $x \in y)$ and $\forall y \in \mathscr{F}\ \forall x \in y\ (x \in A)$." Then by *transfer* one gets "${}^*A = \bigcup_{y\in{}^*\mathscr{F}} y$."

Proposition 2.16.

1. *R is a k-ary relation if and only if* $^*R$ *is a k-ary relation.*

   If R is a binary relation:

2. $^*\{a \mid \exists b\, R(a,b)\} = \{\xi \mid \exists\zeta\, {}^*R(\xi,\zeta)\}$. *that is,* $^*domain(R) = domain(^*R)$.
3. $^*\{b \mid \exists a\, R(a,b)\} = \{\zeta \mid \exists\xi\, {}^*R(\xi,\zeta)\}$. *that is,* $^*range(R) = range(^*R)$.
4. $^*\{(a,b) \mid R(b,a)\} = \{(\xi,\zeta) \mid {}^*R(\zeta,\xi)\}$.

   If S is a ternary relation:

5. $^*\{(a,b,c) \mid S(c,a,b)\} = \{(\xi,\zeta,\eta) \mid {}^*S(\xi,\eta,\zeta)\}$.
6. $^*\{(a,b,c) \mid S(a,c,b)\} = \{(\xi,\zeta,\eta) \mid {}^*S(\xi,\eta,\zeta)\}$.

Proof. (1), (2), and (3) are proved by direct applications of *transfer*, because the notions of k-ary relation, domain, and range are elementary by definition.

(4). The property "$C = \{(a,b) \mid R(b,a)\}$" is formalized by the the conjunction of the elementary formula "$\forall z \in C\ \exists x \in \text{domain}(R)\ \exists y \in \text{range}(R)$ s.t. $R(x,y)$ and $z = (y,x)$" and the elementary formula "$\forall x \in \text{domain}(R)\ \forall y \in \text{range}(R)\ (y,x) \in C$". Thus *transfer* applies and one obtains $^*C = \{(\xi,\zeta) \mid (\zeta,\xi) \in {}^*R\}$.

(5) and (6) are proved by considering similar elementary formulas as in (4).

Proposition 2.17.

1. f *is a function if and only if* $^*f$ *is a function.*

   If f,g *are functions and* A,B *are sets:*

2. $^*domain(f) = domain(^*f)$.
3. $^*range(f) = range(^*f)$.
4. $f : A \to B$ *if and only if* $^*f : {}^*A \to {}^*B$.[2]
5. $^*graph(f) = graph(^*f)$.
6. $^*(f(a)) = (^*f)(^*a)$ *for every* $a \in domain(f)$.
7. *If* $f : A \to A$ *is the identity, then* $^*f : {}^*A \to {}^*A$ *is the identity, that is* $^*(1_A) = 1_{^*A}$.
8. $^*\{f(a) \mid a \in A\} = \{^*f(\xi) \mid \xi \in {}^*A\}$. *that is* $^*(f(A)) = {}^*f(^*A)$.
9. $^*\{a \mid f(a) \in B\} = \{\xi \mid {}^*f(\xi) \in {}^*B\}$. *that is* $^*(f^{-1}(B)) = (^*f)^{-1}(^*B)$.
10. $^*(f \circ g) = {}^*f \circ {}^*g$.
11. $^*\{(a,b) \in A \times B \mid f(a) = g(b)\} = \{(\xi,\zeta) \in {}^*A \times {}^*B \mid {}^*f(\xi) = {}^*g(\zeta)\}$.

Proof. (1), (2), (3), and (6) are proved by direct applications of *transfer*, because the notions of function, value of a function at a given point, domain, and range, are elementary by definition. (4) is a direct corollary of the previous properties. We only prove two of the remaining properties as all of the proofs are similar to one another.

(5). The property "$C = \text{graph}(f)$" is formalized by the elementary formula obtained as the conjunction of the formula "$\forall z \in C\ \exists x \in \text{domain}(f)\ \exists y \in \text{range}(f)$ s.t. $y = f(x)$ and $(x,y) \in C$" with the formula "$\forall x \in \text{domain}(f)\ \forall y \in \text{range}(f)\ (y = f(x) \Rightarrow (x,y) \in C)$". The desired equality follows by *transfer* and by the previous properties.

(10). If $f : A \to B$ and $g : B \to C$, then the property "$h = g \circ f$" is formalized by the formula "$h : A \to C$ and $\forall x \in A\ \forall y \in C\ (h(x) = y \Leftrightarrow \exists z \in B\ f(x) = z \text{ and } g(z) = y)$".

Exercise 2.18. Prove that a function $f : A \to B$ is 1-1 if and only if $^*f : {}^*A \to {}^*B$ is 1-1.

We now discuss a general result about the star map that is really useful in practice (and, in fact, several particular cases have already been included in the previous propositions): If a set is defined by means of an elementary property, then its hyper-extension is defined by the same property where one puts stars in front of the parameters.

---

[2] Recall that notation $f : A \to B$ means that f is a function with $\text{domain}(f) = A$ and $\text{range}(f) \subseteq B$.

Proposition 2.19. *Let $\varphi(x,y_1,\dots,y_n)$ be an elementary formula. For all objects $B,A_1,\dots,A_n$ one has*

$$^*\{x\in B \mid \varphi(x,A_1,\dots,A_n)\} \;=\; \{x\in {}^*B \mid \varphi(x,{}^*A_1,\dots,{}^*A_n)\}.$$

Proof. Let us denote by $C=\{x\in B\mid \varphi(x,A_1,\dots,A_n)\}$. Then the following property holds:

$$P(A_1,\dots,A_n,B,C):\;\forall x\,(x\in C\Leftrightarrow(x\in B \text{ and } \varphi(x,A_1,\dots,A_n))\,.$$

The above formula is elementary; indeed, it is an abbreviation for the conjuction of the two formulas: "$\forall x\in C\ (x\in B$ and $\varphi(x,A_1,\dots,A_n)$" and "$\forall x\in B\ (\varphi(x,A_1,\dots,A_n)\Rightarrow x\in C)$", where all quantifiers are bounded, and where φ is elementary by hypothesis. Then we can apply *transfer* and obtain the validity of $P({}^*A_1,\dots,{}^*A_n,{}^*B,{}^*C)$, that is ${}^*C=\{x\in{}^*B\mid \varphi(x,{}^*A_1,\dots,{}^*A_n)\}$.

An immediate corollary is the following.

Proposition 2.20. *If $(a,b)=\{x\in\mathbb{R}\mid a<x<b\}$ is an open interval of real numbers then ${}^*(a,b)=\{\xi\in{}^*\mathbb{R}\mid a<\xi<b\}$, and similarly for intervals of the form $[a,b)$, $(a,b]$, (a,b), $(-\infty,b]$ and $[a,+\infty)$. Analogous properties hold for intervals of natural, integer, or rational numbers.*

We reveal in advance to the suspicious reader that star maps satisfying the *transfer principle*, and hence all the properties itemized in this section, do actually exist; indeed, they can be easily constructed by means of ultrafilters, or, equivalently, by means of maximal ideals of rings of functions (see Section 2.4).

2.2.1 Additional assumptions

By property Proposition 2.15 (1) and (2), the hyper-extension ${}^*A$ of a set A contains a copy of A given by the hyper-extensions of its elements

$$^{\sigma}A \;=\; \{{}^*a\mid a\in A\}\;\subseteq{}^*A.$$

Notice that, by *transfer*, an hyper-extension ${}^*x$ belongs to ${}^*A$ if and only if $x\in A$. In consequence, ${}^*A\cap{}^{\sigma}B={}^{\sigma}(A\cap B)$ for all sets A,B.

Following the common use in nonstandard analysis, to simplify matters we will assume that ${}^*r=r$ for all $r\in\mathbb{R}$, and more generally, that ${}^*(r_1,\dots,r_k)=(r_1,\dots,r_k)$ for all tuples of real numbers. This means that ${}^{\sigma}(\mathbb{R}^k)=\mathbb{R}^k$ and, in consequence, hyper-extensions of real sets and functions are actual extensions:

- $A\subseteq{}^*A$ for every $A\subseteq\mathbb{R}^k$,
- If $f:A\to B$ where $A\subseteq\mathbb{R}^k$ and $B\subseteq\mathbb{R}^h$, then ${}^*f$ is an extension of f, that is, ${}^*f(a)=f(a)$ for every $a\in A$.

To avoid trivialities, in nonstandard analysis it is always assumed that the star map satisfies the following

- *Properness condition:* ${}^*\mathbb{N}\neq\mathbb{N}$.

Proposition 2.21. *If the properness condition ${}^*\mathbb{N}\neq\mathbb{N}$ holds then ${}^{\sigma}A\neq{}^*A$ for every infinite A.*

Proof. Given an infinite set A, pick a surjective map $f:A\to\mathbb{N}$; then also the hyper-extension ${}^*f:{}^*A\to{}^*\mathbb{N}$ is surjective, and

$${}^*\mathbb{N}\;=\;\{{}^*f(\alpha)\mid\alpha\in{}^*A\}\;=\;\{{}^*f({}^*a)\mid a\in A\}\;=\;\{{}^*(f(a))\mid a\in A\}\;=\;\{{}^*n\mid n\in\mathbb{N}\}\;=\;\mathbb{N}.$$

As a first consequence of the properness condition, one gets a nonstandard characterization of finite sets as those sets that are not "extended" by hyper-extensions.

Proposition 2.22. *For every set A one has the equivalence: "A is finite if and only if ${}^*A={}^{\sigma}A$". (When $A\subseteq\mathbb{R}^k$, this is the same as "A is finite if and only if ${}^*A=A$".)*

Proof. If $A = \{a_1, \dots, a_k\}$ is finite, we already saw in Proposition 2.15 (9) that ${}^*A = \{{}^*a_1, \dots, {}^*a_k\} = \{{}^*a \mid a \in A\}$. Conversely, if A is infinite, we can pick a surjective function $f : A \to \mathbb{N}$. Then also ${}^*f : {}^*A \to {}^*\mathbb{N}$ is onto. Now notice that for every $a \in A$, one has that $({}^*f)({}^*a) = {}^*(f(a)) \in \mathbb{N}$ (recall that ${}^*n = n$ for every $n \in \mathbb{N}$). Then if $\xi \in {}^*\mathbb{N} \setminus \mathbb{N}$ there exists $\alpha \in {}^*A \setminus \{{}^*a \mid a \in A\}$ with ${}^*f(\alpha) = \xi$.

One can safely extend the simplifying assumption ${}^*r = r$ from real numbers r to elements of any given mathematical object X under study.

- *Unless explicitly mentioned otherwise, when studying a specific mathematical object X by nonstandard analysis, we will assume that ${}^*x = x$ for all $x \in X$, so that $X = {}^\sigma X \subseteq {}^*X$.*

2.3 The transfer principle, in practice

As we already pointed out, a correct application of *transfer* needs a precise understanding of the notion of elementary property. Basically, a property is elementary if it talks about the elements of some given structures and *not* about their subsets or the functions between them.[3] Indeed, in order to correctly apply the *transfer principle*, one must always point out the range of quantifiers, and formulate them in the forms "$\forall x \in X \dots$" and "$\exists y \in Y \dots$" for suitable specified sets X, Y. With respect to this, the following remark is particularly relevant.

Remark 2.23. Before applying *transfer*, all quantifications on subsets "$\forall x \subseteq X \dots$" or "$\exists x \subseteq X \dots$" must be reformulated as "$\forall x \in \mathscr{P}(X) \dots$" and "$\exists x \in \mathscr{P}(X) \dots$", respectively, where $\mathscr{P}(X) = \{A \mid A \subseteq X\}$ is the *powerset* of X. Similarly, all quantifications on functions $f : A \to B$ must be bounded by $\mathrm{Fun}(A, B)$, the *set of functions* from A to B. We stress that these instructions are pivotal because in general ${}^*\mathscr{P}(X) \neq \mathscr{P}({}^*X)$ and ${}^*\mathrm{Fun}(A, B) \neq \mathrm{Fun}({}^*A, {}^*B)$, as we will show in Proposition 2.49.

Example 2.24. Consider the property: "$<$ is a linear ordering on the set A". Notice first that $<$ is a binary relation on A, and hence its hyper-extension ${}^*{<}$ is a binary relation on ${}^*A$. By definition, $<$ is a linear ordering if and only if the following are satisfied:

(a) $\forall x \in A\ (x \not< x)$,
(b) $\forall x, y, z \in A\ (x < y \text{ and } y < z) \Rightarrow x < z$,
(c) $\forall x, y \in A\ (x < y \text{ or } y < x \text{ or } x = y)$.

Notice that the three formulas above are elementary. Then we can apply *transfer* and conclude that: "${}^*{<}$ is a linear ordering on ${}^*A$."

Whenever confusion is unlikely, some asterisks will be omitted. So, for instance, we will write $+$ to denote both the sum operation on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$, and the corresponding operations on ${}^*\mathbb{N}$, ${}^*\mathbb{Z}$, ${}^*\mathbb{Q}$ and ${}^*\mathbb{R}$, respectively, as given by the hyper-extension ${}^*{+}$.

Similarly as in the example above, it is readily verified that the properties of a discretely ordered ring, as well as the properties of a real-closed ordered field, are elementary because they just talk about the *elements* of the given structures. Thus, by a direct application of *transfer*, one obtain the following results, which are coherent with what already seen in the "warming-up" section §2.2.1.

Theorem 2.25.

1. *${}^*\mathbb{R}$, endowed with the hyper-extensions of the sum, product, and order on $\mathbb{R}$, is a real-closed ordered field.*[4]
2. *${}^*\mathbb{Z}$ is an unbounded discretely ordered subring of ${}^*\mathbb{R}$, whose positive part is ${}^*\mathbb{N}$.*
3. *The ordered subfield ${}^*\mathbb{Q} \subset {}^*\mathbb{R}$ is the quotient field of ${}^*\mathbb{Z}$.*

[3] In logic, properties that talks about elements of a given structure are called *first-order properties*; properties about subsets of the given structure are called *second-order*; properties about subsets of subsets of the given structure are called *third-order*; and so forth.

[4] Recall that an ordered field is *real closed* if every positive element is a square, and every polynomial of odd degree has a root.

Again by direct applications of *transfer*, one also obtains the following properties.

Proposition 2.26.

1. *Every non-zero* $\nu \in {}^*\mathbb{N}$ *has a successor* $\nu + 1$ *and a predecessor* $\nu - 1$.[5]
2. *For every positive* $\xi \in {}^*\mathbb{R}$ *there exists a unique* $\nu \in {}^*\mathbb{N}$ *with* $\nu \leq \xi < \nu + 1$*; as a result,* ${}^*\mathbb{N}$ *is unbounded in* ${}^*\mathbb{R}$.
3. *The hyperrational numbers* ${}^*\mathbb{Q}$*, as well as the hyperirrational numbers* ${}^*(\mathbb{R} \setminus \mathbb{Q}) = {}^*\mathbb{R} \setminus {}^*\mathbb{Q}$*, are dense in* ${}^*\mathbb{R}$.

Proposition 2.27. $(\mathbb{N}, \leq)$ *is an initial segment of* $({}^*\mathbb{N}, \leq)$*, that is, if* $\nu \in {}^*\mathbb{N} \setminus \mathbb{N}$*, then* $\nu > n$ *for all* $n \in \mathbb{N}$*,*

Proof. For every $n \in \mathbb{N}$, by *transfer* one obtains the validity of the following elementary formula: "$\forall x \in {}^*\mathbb{N}$ $(x \neq 1$ and $\ldots$ and $x \neq n) \Rightarrow x > n$", and hence the proposition holds.

To get a clearer picture of the situation, examples of *non*-elementary properties that are not preserved under hyper-extensions, are now in order.

Example 2.28. The property of well-ordering (that is, every nonempty subset has a least element) is *not* elementary; indeed, it is about the *subsets* of the given ordered set. Similarly, the property of completeness of an ordered set is *not* elementary, because it is about the *subsets* of the given structure. Notice that those properties are not preserved by hyper-extensions; indeed, $\mathbb{N}$ is well-ordered but ${}^*\mathbb{N}$ is not (*e.g.*, the set of infinite hyper-natural numbers has no least element); and the real line $\mathbb{R}$ is complete but ${}^*\mathbb{R}$ is not (*e.g.*, the set of infinitesimal numbers is bounded with no least upper bound).

Remark 2.29. *Transfer* applies also to the the well-ordering property of $\mathbb{N}$, provided one formalizes it as: "*Every nonempty element of* $\mathscr{P}(\mathbb{N})$ *has a least element*". (The property "X has a least element" is elementary: "there exists $x \in X$ such that for every $y \in X$, $x \leq y$.") In this way, one gets: "*Every nonempty element of* ${}^*\mathscr{P}(\mathbb{N})$ *has a least element*". The crucial point here is that ${}^*\mathscr{P}(\mathbb{N})$ is a only a proper subfamily of $\mathscr{P}({}^*\mathbb{N})$ (see Proposition 2.49 below). So, the well-ordering property is *not* an elementary property of $\mathbb{N}$, but it is actually an elementary property of $\mathscr{P}(\mathbb{N})$. Much the same observations can be made about the completeness property. Indeed, virtually all properties of mathematical objects can be formalized by elementary formulas, provided one uses the appropriate parameters.

A much more slippery example of a *non*-elementary property is the following.

Example 2.30. The Archimedean property of an ordered field $\mathbb{F}$ is *not* elementary. Notice that to formulate it, one needs to also consider the substructure $\mathbb{N} \subset \mathbb{F}$:

"For all positive $x \in \mathbb{F}$ *there exists* $n \in \mathbb{N}$ *such that* $nx > 1$*."*

While the above is an elementary property of the pair $(\mathbb{F}, \mathbb{N})$ since it talks about the elements of $\mathbb{F}$ and $\mathbb{N}$ combined, it is *not* an elementary property of the ordered field $\mathbb{F}$ alone. In regard to this, we remark that the following expression:

"For all positive $x \in \mathbb{F}$ *it is* $x > 1$ *or* $2x > 1$ *or* $3x > 1$ *or* $\ldots$ *or* $nx > 1$ *or* $\ldots$*."*

is *not* a formula, because it would consist in an infinitely long string of symbols if written in full. Notice that the Archimedean property is not preserved by hyper-extensions; for instance, $\mathbb{R}$ is an Archimedean, but the hyperreal line ${}^*\mathbb{R}$ is not, in that an ordered field that properly extends $\mathbb{R}$ (see Proposition 2.5).

Similarly, also the properties of being infinitesimal, finite, or infinite are *not* elementary properties of elements in a given ordered field $\mathbb{F}$, because to formulate them one needs to also consider the substructure $\mathbb{N} \subset \mathbb{F}$ as a parameter.

---

[5] An element η is the *successor* of ξ (or ξ is the *predecessor* of η) if $\xi < \eta$ and there are no elements ζ with $\xi < \zeta < \eta$.

2.4 The ultrapower model

It is now time to justify what we have seen in the previous sections and show that star maps that satisfy the *transfer* principle do actually exist. Many researchers using nonstandard methods, especially those who do not have a strong background in logic, feel more comfortable in directly working with a model; however we remark that this is not necessary. Rather, it is worth stressing that all that one needs in practice is a good understanding of the *transfer* principle and its use, whereas the underlying construction of a specific star map does not play any role.[6] The situation is similar to what happens when working in real analysis: what really matters are the properties of a complete ordered field, along with the fact that a complete ordered field does actually exist; whereas the specific construction of the real line (*e.g.*, by means of Dedekind cuts or by a suitable quotient of the set of Cauchy sequences) is irrelevant when developing the theory.

2.4.1 The ultrapower construction

The ultrapower construction relies on ultrafilters and so, to begin with, let us fix an ultrafilter $\mathscr{U}$ on a set of indices I.

For simplicity, in the following we will focus on the real numbers; however, the same construction can be carried out by starting with any mathematical structure.

Definition 2.31. The *ultrapower* of $\mathbb{R}$ modulo the ultrafilter $\mathscr{U}$, denoted $\mathbb{R}^I/\mathscr{U}$, is the quotient of the family of real I-sequences $\mathbb{R}^I = \mathrm{Fun}(I,\mathbb{R}) = \{\sigma \mid \sigma : I \to \mathbb{R}\}$ modulo the equivalence relation $\equiv_{\mathscr{U}}$ defined by setting:

$$\sigma \equiv_{\mathscr{U}} \tau \;\Leftrightarrow\; \{i \in I \mid \sigma(i) = \tau(i)\} \in \mathscr{U}.$$

Notice that the properties of being a filter on $\mathscr{U}$ guarantee that $\equiv_{\mathscr{U}}$ is actually an equivalence relation. Equivalence classes are denoted by using square brackets: $[\sigma] = \{\tau \in \mathrm{Fun}(I,\mathbb{R}) \mid \tau \equiv_{\mathscr{U}} \sigma\}$. The pointwise sum and product operations on the ring $\mathrm{Fun}(I,\mathbb{R})$ are inherited by the ultrapower; indeed, it is easily verified that the following definitions are well-posed:

$$[\sigma]+[\tau] = [\sigma+\tau] \quad \text{and} \quad [\sigma]\cdot[\tau] = [\sigma\cdot\tau].$$

The order relation $<$ on the ultrapower is defined by putting:

$$[\sigma] < [\tau] \;\Leftrightarrow\; \{i \in I \mid \sigma(i) < \tau(i)\} \in \mathscr{U}.$$

Proposition 2.32. *The ultrapower* $(\mathbb{R}^I/\mathscr{U}, +, \cdot, <, \mathbf{0}, \mathbf{1})$ *is an ordered field.*

Proof. All properties of an ordered field are directly proved by using the properties of an ultrafilter. For example, to prove that $<$ is a total ordering, one considers the partition $I = I_1 \cup I_2 \cup I_3$ where $I_1 = \{i \in I \mid \sigma(i) < \tau(i)\}$, $I_2 = \{i \in I \mid \sigma(i) = \tau(i)\}$ and $I_3 = \{i \in I \mid \sigma(i) > \tau(i)\}$: exactly one out of the three sets belongs to $\mathscr{U}$, and hence exactly one out of $[\sigma] < [\tau]$, $[\sigma] = [\tau]$, or $[\sigma] > [\tau]$ holds. As another example, let us show that every $[\sigma] \neq 0$ has a multiplicative inverse. By assumption, $A = \{i \in I \mid \sigma(i) = 0\} \notin \mathscr{U}$, and so the complement $A^c = \{i \in I \mid \sigma(i) \neq 0\} \in \mathscr{U}$. Now pick any I-sequence τ such that $\tau(i) = 1/\sigma(i)$ whenever $i \in A^c$. Then $A^c \subseteq \{i \in I \mid \sigma(i)\cdot\tau(i) = 1\} \in \mathscr{U}$, and hence $[\sigma]\cdot[\tau] = 1$.

There is a canonical way of embedding $\mathbb{R}$ into its ultrapower.

Definition 2.33. The *diagonal embedding* $d : \mathbb{R} \to \mathbb{R}^I/\mathscr{U}$ is the function that associates to every real number r the equivalence class of the corresponding constant I-sequence $[c_r]$.

It is readily seen that d is a 1-1 map that preserves sums, products and the order relation. As a result, without loss of generality, we can identify every $r \in \mathbb{R}$ with its diagonal image $d(r) = [c_r]$, and assume that $\mathbb{R} \subseteq \mathbb{R}^I/\mathscr{U}$ is an ordered subfield.

Notice that if $\mathscr{U} = \mathscr{U}_j$ is principal then the corresponding ultrapower $\mathbb{R}^I/\mathscr{U}_j = \mathbb{R}$ is trivial. Indeed, in this case one has $\sigma \equiv_{\mathscr{U}_j} \tau \Leftrightarrow \sigma(j) = \tau(j)$; so, every sequence is equivalent to the constant I-sequence with value $\sigma(j)$, and the diagonal embedding $d : \mathbb{R} \to \mathbb{R}^I/\mathscr{U}_j$ is onto.

[6] Actually, there are a few exceptions to this statement, but we will never see them in the combinatorial applications presented in this book.

Remark 2.34. Under the *continuum hypothesis*, one can show that for every pair $\mathscr{U}, \mathscr{V}$ of non-principal ultrafilters on $\mathbb{N}$, the hyperreal numbers given by the corresponding ultrapower models $\mathbb{R}^{\mathbb{N}}/\mathscr{U} \cong \mathbb{R}^{\mathbb{N}}/\mathscr{V}$ are isomorphic as ordered fields. (This is because they are are $\aleph_1$-saturated models of cardinality $\aleph_1$ in a finite language.)

2.4.2 Hyper-extensions in the ultrapower model

In this section we will see how the ultrapower $\mathbb{R}^I/\mathscr{U}$ can be made a model of the hyperreal numbers of nonstandard analysis. Let us start by denoting

$$^*\mathbb{R} \;=\; \mathbb{R}^I/\mathscr{U}.$$

We now have to show that the ordered field $^*\mathbb{R}$ has all the special features that make it a set of hyperreal numbers; to this end, we will define a *star map* on the family of all sets of tuples of real numbers and of all real functions, in such a way that the *transfer* principle holds.

Definition 2.35. Let $A \subseteq \mathbb{R}$. Then its *hyper-extension* $^*A \subseteq {}^*\mathbb{R}$ is defined as the family of all equivalence classes of I-sequences that take values in A, that is:

$$^*A \;=\; A^I/\mathscr{U} \;=\; \{[\sigma] \mid \sigma : I \to A\} \;\subseteq\; {}^*\mathbb{R}.$$

Similarly, if $A \subseteq \mathbb{R}^k$ is a set of real k-tuples, then its *hyper-extension* is defined as

$$^*A \;=\; \{([\sigma_1], \ldots, [\sigma_k]) \mid \sigma_1, \ldots, \sigma_k : I \to A\} \;\subseteq\; {}^*\mathbb{R}^k.$$

Notice that, by the properties of ultrafilter, for every $\sigma_1, \ldots, \sigma_k, \tau_1, \ldots, \tau_k : I \to \mathbb{R}$, one has

$$\{i \in I \mid (\sigma_1(i), \ldots, \sigma_k(i)) = (\tau_1(i), \ldots, \tau_k(i))\} \in \mathscr{U} \iff \sigma_s \equiv_{\mathscr{U}} \tau_s \text{ for every } s = 1, \ldots, k.$$

In consequence, the above definition is well-posed.

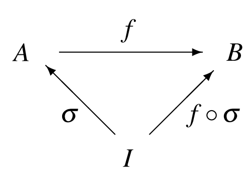
We also define the star map on real tuples by setting

$$^*(r_1, \ldots, r_k) = (r_1, \ldots, r_k).$$

Recall that we identified every $r \in \mathbb{R}$ with the equivalence class $[c_r]$ of the corresponding constant sequence and so, by letting $^*r = r = [c_r]$, we have that $A \subseteq {}^*A$ for every $A \subseteq \mathbb{R}^k$.

We have already seen that $^*\mathbb{R}$ is an ordered field that extends the real line; as a result, every rational function $f : \mathbb{R} \to \mathbb{R}$ is naturally extended to a function $^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$. However, here we are interested in extending *all* real functions $f : A \to B$ where A and B are set of real tuples, to functions $^*f : {}^*A \to {}^*B$; with ultrapowers, this can be done in a natural way.

Definition 2.36. Let $f : A \to B$ where $A, B \subseteq \mathbb{R}$. Then the *hyper-extension* of f is the function $^*f : {}^*A \to {}^*B$ defined by setting $^*f([\sigma]) = [f \circ \sigma]$ for every $\sigma : I \to A$.



If $f : A \to B$ is a function of several variables where $A \subseteq \mathbb{R}^k$ and $B \subseteq \mathbb{R}$, then $^*f : {}^*A \to {}^*B$ is defined by setting for every $\sigma_1, \ldots, \sigma_k : I \to \mathbb{R}$:

$$^*f([\sigma_1], \ldots, [\sigma_k]) \;=\; [\langle f(\sigma_1(i), \ldots, \sigma_k(i)) \mid i \in I\rangle].$$

Similarly as for hyper-extensions of sets of tuples, it is routine to check that the properties of a filter guarantee that the above definition is well-posed.

Let us now see that the ultrapower model has all the desired properties.

Theorem 2.37. *The hyper-extensions of real tuples, sets of real tuples and real functions, as defined above, satisfy all the properties itemized in Propositions 2.15, 2.16, and 2.17:*[7]

1. $a = b \Leftrightarrow {}^*a = {}^*b$.
2. A *is a set if and only if* ${}^*A$ *is a set.*
3. ${}^*\emptyset = \emptyset$.
4. $A \subseteq B \Leftrightarrow {}^*A \subseteq {}^*B$.
5. ${}^*(A \cup B) = {}^*A \cup {}^*B$.
6. ${}^*(A \cap B) = {}^*A \cap {}^*B$.
7. ${}^*(A \setminus B) = {}^*A \setminus {}^*B$.
8. ${}^*\{a_1,\dots,a_k\} = \{a_1,\dots,a_k\}$.
9. ${}^*(a_1,\dots,a_k) = (a_1,\dots,a_k)$.
10. ${}^*(A_1 \times \ldots \times A_k) = {}^*A_1 \times \ldots \times {}^*A_k$.
11. ${}^*\{(a,a) \mid a \in A\} = \{(\xi,\xi) \mid \xi \in A\}$.
12. R *is a k-ary relation if and only if* ${}^*R$ *is a k-ary relation.*
13. ${}^*\{a \mid \exists b\, R(a,b)\} = \{\xi \mid \exists \zeta\, {}^*R(\xi,\zeta)\}$, *that is,* ${}^*domain(R) = domain({}^*R)$.
14. ${}^*\{b \mid \exists a\, R(a,b)\} = \{\zeta \mid \exists \xi\, {}^*R(\xi,\zeta)\}$, *that is,* ${}^*range(R) = range({}^*R)$.
15. ${}^*\{(a,b) \mid R(b,a)\} = \{(\xi,\zeta) \mid {}^*R(\zeta,\xi)\}$.
16. ${}^*\{(a,b,c) \mid S(c,a,b)\} = \{(\xi,\zeta,\eta) \mid {}^*S(\xi,\eta,\zeta)\}$.
17. ${}^*\{(a,b,c) \mid S(a,c,b)\} = \{(\xi,\zeta,\eta) \mid {}^*S(\xi,\eta,\zeta)\}$.
18. f *is a function if and only if* ${}^*f$ *is a function.*
19. ${}^*domain(f) = domain({}^*f)$.
20. ${}^*range(f) = range({}^*f)$.
21. $f : A \to B$ *if and only if* ${}^*f : {}^*A \to {}^*B$.
22. ${}^*graph(f) = graph({}^*f)$.
23. $({}^*f)(a) = f(a)$ *for every* $a \in domain(f)$.
24. *If* $f : A \to A$ *is the identity, then* ${}^*f : {}^*A \to {}^*A$ *is the identity, that is* ${}^*(1_A) = 1_{{}^*A}$.
25. ${}^*\{f(a) \mid a \in A\} = \{{}^*f(\xi) \mid \xi \in {}^*A\}$, *that is* ${}^*(f(A)) = {}^*f({}^*A)$.
26. ${}^*\{a \mid f(a) \in B\} = \{\xi \mid {}^*f(\xi) \in {}^*B\}$, *that is* ${}^*(f^{-1}(B)) = ({}^*f)^{-1}({}^*B)$.
27. ${}^*(f \circ g) = {}^*f \circ {}^*g$.
28. ${}^*\{(a,b) \in A \times B \mid f(a) = g(b)\} = \{(\xi,\zeta) \in {}^*A \times {}^*B \mid {}^*f(\xi) = {}^*g(\zeta)\}$.

Proof. ……

We disclose that the above result essentially states that our defined star map satisfies the *transfer* principle. Indeed, once the notion of elementary property will be made fully rigorous, one can show that *transfer* is actually equivalent to the validity of the properties listed above.

Remark 2.38. A "strong isomorphism" between two sets of hyperreals ${}^*\mathbb{R}$ and ${}^\star\mathbb{R}$ is defined as a bijection $\psi : {}^*\mathbb{R} \to {}^\star\mathbb{R}$ that it coherent with hyper-extensions, that is, $(\xi_1,\dots,\xi_k) \in {}^*A \Leftrightarrow (\Psi(\xi_1),\dots,\Psi(\xi_k)) \in {}^\star A$ for every $A \subseteq \mathbb{R}^k$ and for every $\xi_1,\dots,\xi_k \in {}^*\mathbb{R}$, and ${}^*f(\xi_1,\dots,\xi_k) = \eta \Leftrightarrow {}^\star f(\Psi(\xi_1),\dots,\Psi(\xi_k)) = \Psi(\eta)$ for every $f : \mathbb{R}^k \to \mathbb{R}$ and for every $\xi_1,\dots,\xi_k,\eta \in {}^*\mathbb{R}$. Then one can show that two ultrapower models $\mathbb{R}^{\mathbb{N}}/\mathscr{U}$ and $\mathbb{R}^{\mathbb{N}}/\mathscr{V}$ are "strongly isomorphic" if and only if the ultrafilters $\mathscr{U} \cong \mathscr{V}$ are isomorphic, that is, there exists a permutation $\sigma : \mathbb{N} \to \mathbb{N}$ such that $A \in \mathscr{U} \Leftrightarrow \sigma(A) \in \mathscr{V}$ for every $A \subseteq \mathbb{N}$. We remark that there exist plently of non-isomorphic ultrafilters (indeed, one can show that there are $2^{\mathfrak{c}}$-many distinct classes of isomorphic ultrafilters on $\mathbb{N}$). This is to be contrasted with the previous Remark 2.34, where the notion of isomorphism between sets of hyperreals was limited to the structure of ordered field.

[7] Since hyper-extensions of families of sets have not been defined, properties (13) and (14) of Proposition 2.15 are not included in the list. Clearly, (4), (5), (6), (7) only applies when A,B are sets of real tuples; (8) and (9) only applies when objects a_i are real tuples; and so forth.

2.4.3 The properness condition in the ultrapower model

In the previous section, we observed that principal ultrafilters generate trivial ultrapowers. Below, we precisely isolate the class of those ultrafilters that produce models where the properness condition $\mathbb{N} \neq {}^*\mathbb{N}$ (as well as $\mathbb{R} \neq {}^*\mathbb{R}$) holds.

Recall that an ultrafilter $\mathscr{U}$ is called *countably incomplete* if it is not closed under countable intersections, that is, if there exists a countable family $\{I_n\}_{n\in\mathbb{N}} \subseteq \mathscr{U}$ such that $\bigcap_{n\in\mathbb{N}} I_n \notin \mathscr{U}$. We remark that all non-principal ultrafilters on $\mathbb{N}$ or on $\mathbb{R}$ are countably incomplete.[8]

Exercise 2.39. An ultrafilter $\mathscr{U}$ on I is countably incomplete if and only if there exists a countable partition $I = \bigcup_{n\in\mathbb{N}} J_n$ where $J_n \notin \mathscr{U}$ for every n.

Proposition 2.40. *In the ultrapower model modulo the ultrafilter $\mathscr{U}$ on I the following properties are equivalent:*

1. *Properness condition: ${}^*\mathbb{N} \neq \mathbb{N}$;*
2. *$\mathscr{U}$ is countably incomplete.*

Proof. Assume first that ${}^*\mathbb{N} \neq \mathbb{N}$. Pick a sequence $\sigma : I \to \mathbb{N}$ such that $[\sigma] \notin \mathbb{N}$. Then $I_n = \{i \in I \mid \sigma(i) \neq n\} \in \mathscr{U}$ for every $n \in \mathbb{N}$, but $\bigcap_n I_n = \emptyset \notin \mathscr{U}$. Conversely, if $\mathscr{U}$ is countably incomplete, pick a countable partition $I = \bigcup J_n$ where $J_n \notin \mathscr{U}$ for every n, and pick the sequence $\sigma : I \to \mathbb{N}$ where $\sigma(i) = n$ for $i \in J_n$. Then $[\sigma] \in {}^*\mathbb{N}$ but $[\sigma] \neq d(n)$ for every n.

Proposition 2.41. $\mathbb{R} \neq \mathbb{R}^I/\mathscr{U}$ *if and only if the ultrafilter $\mathscr{U}$ is countably incomplete.*

Proof. Assume first that ${}^*\mathbb{N} \neq \mathbb{N}$. Pick a sequence $\sigma : I \to \mathbb{N}$ such that $[\sigma] \notin \mathbb{N}$. Then $I_n = \{i \in I \mid \sigma(i) \neq n\} \in \mathscr{U}$ for every $n \in \mathbb{N}$, but $\bigcap_n I_n = \emptyset \notin \mathscr{U}$. Conversely, if $\mathscr{U}$ is countably incomplete, pick a family $\{I_n\}_{n\in\mathbb{N}}$ of subsets of I such that $I_n \in \mathscr{U}$ for every n but $\bigcap_n I_n \notin \mathscr{U}$. By replacing every I_k with $(\bigcap_{j=1}^k I_k) \setminus (\bigcap_n I_n)$, we can assume that $\{I_n\}_{n\in\mathbb{N}}$ is non-increasing and that $\bigcap_{n\in\mathbb{N}} I_n = \emptyset$. Let $\sigma : I \to \mathbb{N}$ be the sequence such that $\sigma(i) = \min\{k \mid i \notin I_k\}$. Then for every n one has that $\{i \mid \sigma(i) = n\} = I_{n-1} \setminus I_n \notin \mathscr{U}$ (where we let $I_0 = \mathbb{N}$), and so $[\sigma] \notin \mathrm{range}(d)$.

In the sequel we will always assume that ultrapower models are constructed by using ultrafilters $\mathscr{U}$ that are countably incomplete.

2.4.4 An algebraic presentation

The ultrapower model can be presented in an alternative, but equivalent, purely algebraic fashion where only the notion of quotient field of a ring modulo a maximal ideal is assumed.[9] Here are the steps of the construction.

- Consider $\mathrm{Fun}(I,\mathbb{R})$, the ring of real valued sequences where the sum and product operations are defined pointwise.
- Let $\mathfrak{i}$ be the ideal of those sequences with *finite support*:

$$\mathfrak{i} \;=\; \{\sigma \in \mathrm{Fun}(I,\mathbb{R}) \mid \sigma(i) = 0 \text{ for all but at most finitely many } i\}.$$

- Extend $\mathfrak{i}$ to a maximal ideal $\mathfrak{m}$, and define the hyperreal numbers as the quotient field:

$${}^*\mathbb{R} \;=\; \mathrm{Fun}(I,\mathbb{R})/\mathfrak{m}.$$

- For every subset $A \subseteq \mathbb{R}$, its hyper-extension is defined by:

$${}^*A \;=\; \{\sigma + \mathfrak{m} \mid \sigma : I \to A\} \;\subseteq\; {}^*\mathbb{R}.$$

So, *e.g.*, the *hyper-natural numbers* ${}^*\mathbb{N}$ are the cosets $\sigma + \mathfrak{m}$ of I-sequences $\sigma : I \to \mathbb{N}$ of natural numbers.

---

[8] The existence of non-principal ultrafilters that are countably complete is equivalent to the existence of the so-called *measurable cardinals*, a kind of inaccessible cardinals studied in the hierarchy of large cardinals, and whose existence cannot be proved by ZFC. In consequence, if one sticks to the usual principles of mathematics, it is safe to assume that every non-principal ultrafilter is countably incomplete.

[9] See [5] for details.

- For every function $f : A \to B$ where $A, B \subseteq \mathbb{R}$, its hyper-extension ${}^*f : {}^*A \to {}^*B$ is defined by putting for every $\sigma : I \to A$:

$$ {}^*f(\sigma + \mathfrak{m}) \ = \ (f \circ \sigma) + \mathfrak{m}. $$

It can be directly verified that ${}^*\mathbb{R}$ is an ordered field whose positive elements are ${}^*\mathbb{R}^+ = \mathrm{Fun}(\mathbb{N}, \mathbb{R}^+)/\mathfrak{m}$, where $\mathbb{R}^+$ is the set of positive reals. By identifying each $r \in \mathbb{R}$ with the coset $c_r + \mathfrak{m}$ of the corresponding constant sequence, one obtains that $\mathbb{R}$ is a proper subfield of ${}^*\mathbb{R}$.

Notice that the above definitions, exactly as done with the ultrapower model, are naturally extended to hyper-extensions of sets of real tuples and of functions between sets of real tuples.

Remark 2.42. The algebraic approach presented here is basically equivalent to the ultrapower model. Indeed, for every function $f : I \to \mathbb{R}$, let us denote by $Z(f) = \{i \in I \mid f(i) = 0\}$ its zero-set. If $\mathfrak{m}$ is a maximal ideal of the ring $\mathrm{Fun}(I, \mathbb{R})$, then it is easily shown that the family $\mathscr{U}_{\mathfrak{m}} = \{Z(f) \mid f \in \mathfrak{m}\}$ is an ultrafilter on $\mathbb{N}$. Conversely, if $\mathscr{U}$ is an ultrafilter on $\mathbb{N}$, then $\mathfrak{m}_{\mathscr{U}} = \{f \mid Z(f) \in \mathscr{U}\}$ is a maximal ideal of the ring $\mathrm{Fun}(I, \mathbb{R})$. The correspondance between $\mathscr{U}$-equivalence classes $[\sigma]$ and cosets $\sigma + \mathfrak{m}_{\mathscr{U}}$ yields an isomorphism between the ultrapower $\mathbb{R}^I/\mathscr{U}$ and the quotient $\mathrm{Fun}(I, \mathbb{R})/\mathfrak{m}_{\mathscr{U}}$.

2.5 Internal and external objects

We are now ready to introduce a fundamental class of objects in nonstandard analysis, namely the internal objects. In a way, they are similar to the open sets in topology, or to the measurable sets in measure theory, because they are those objects that behave "nicely" in our theory.[10]

Recall that the *star map* does not preserve the properties of *powersets* and *function sets*. For instance, we have noticed in the previous sections that that there are (nonempty) sets in ${}^*\mathscr{P}(\mathbb{N})$ with no least element and there are (nonempty) sets in ${}^*\mathscr{P}(\mathbb{R})$ that are bounded but have no least upper bound (see Example 2.28 and Remark 2.29). However, by the *transfer principle*, the family $\mathscr{P}(A)$ of all subsets of a set A and ${}^*\mathscr{P}(A)$ satisfy the same properties; and similarly, the family $\mathrm{Fun}(A, B)$ of all functions $f : A \to B$ and ${}^*\mathrm{Fun}(A, B)$ satisfy the same properties. Let us now elaborate on this, and start with two easy observations.

Proposition 2.43.

1. *Every element of the hyper-extension ${}^*\mathscr{P}(A)$ is a subset of ${}^*A$, that is, ${}^*\mathscr{P}(A) \subseteq \mathscr{P}({}^*A)$;*
2. *Every element of the hyper-extension ${}^*Fun(A, B)$ is a function $f : {}^*A \to {}^*B$, that is, ${}^*Fun(A, B) \subseteq Fun({}^*A, {}^*B)$.*

Proof. (1). Apply *transfer* to the elementary property: $\forall x \in \mathscr{P}(A)\ \forall y \in x\ y \in A$. (2) Apply *transfer* to the elementary property: $\forall x \in \mathrm{Fun}(A, B)$ "x is a function" and $\mathrm{dom}(x) = A$ and $\mathrm{range}(x) \subseteq B$.

Consequently, it is natural to consider the elements in ${}^*\mathscr{P}(A)$ as the "nice" subsets of ${}^*A$; and the elements in ${}^*\mathrm{Fun}(A, B)$ as the "nice" functions from ${}^*A$ to ${}^*B$.

Definition 2.44. Let A, B be sets. The elements of ${}^*\mathscr{P}(A)$ are called the *internal subsets* of ${}^*A$ and the elements of ${}^*\mathrm{Fun}(A, B)$ are called the *internal functions* from A to B. More generally, an *internal* object is any element $B \in {}^*Y$ that belongs to some hyper-extension.

First examples of internal objects are given by the hyperreal numbers $\xi \in {}^*\mathbb{R}$, and also by all tuples of hyperreal numbers $(\xi_1, \ldots, \xi_k) \in {}^*\mathbb{R}^k$. Notice that hyper-extensions ${}^*X$ themselves are internal objects, since trivially ${}^*X \in {}^*\{X\} = \{{}^*X\}$.

- **Rule of thumb.** Properties about *subsets* of a set A transfer to the *internal subsets* of ${}^*A$; and properties about functions $f : A \to B$ trasfer to the *internal functions* from ${}^*A$ to ${}^*B$.

[10] We will make this last statement precise in Proposition XXX, when we will see that properties of subsets or of functions transfer to the corresponding internal objects.

For instance, the *well-ordering* property of $\mathbb{N}$ is transferred to: "Every nonempty *internal* subset of ${}^*\mathbb{N}$ has a least element"; and the *completeness* property of $\mathbb{R}$ transfers to: "Every nonempty *internal* subset of ${}^*\mathbb{R}$ that is bounded above has a least upper bound".

The following is a useful closure property of the class of internal objects.

Theorem 2.45 (Internal Definition Principle). *Let $\varphi(x,y_1,\ldots,y_k)$ be an elementary formula. If A is an internal set and $B_1,\ldots,B_n$ are internal objects, then the set $\{x\in A\mid\varphi(x,B_1,\ldots,B_n)\}$ is also internal.*

Proof. By assumption, there exists a family of sets $\mathscr{F}$ and sets Y_i such that $A\in{}^*\mathscr{F}$ and $B_i\in{}^*Y_i$ for $i=1,\ldots,n$. Pick any family $\mathscr{G}\supseteq\mathscr{F}$ that is closed under subsets, that is, $C'\subseteq C\in\mathscr{G}\Rightarrow C'\in\mathscr{G}$. (For example, one can take $\mathscr{G}=\bigcup\{\mathscr{P}(C)\mid C\in\mathscr{F}\}$.) Then the following is a true elementary property of the objects $\mathscr{G},Y_1,\ldots,Y_n$:

$$P(\mathscr{G},Y_1,\ldots,Y_n):\quad \forall x\in\mathscr{G}\,\forall y_1\in Y_1\ldots\forall y_n\in Y_n\,\exists z\in\mathscr{G}\text{ such that ``}z=\{t\in x\mid\varphi(t,y_1,\ldots,y_n)\}.\text{''}$$ [11]

By *transfer*, the property $P({}^*\mathscr{G},{}^*Y_1,\ldots,{}^*Y_n)$ is also true, and since $A\in{}^*\mathscr{G},B_i\in{}^*Y_i$, we obtain the existence of an internal set $C\in{}^*\mathscr{G}$ such that $C=\{t\in A\mid\varphi(x,B_1,\ldots,B_n)\}$, as desired.

As direct applications of the above principle, one obtains the following properties for the class of internal objects.

Proposition 2.46.

1. *The class $\mathscr{I}$ of internal sets is closed under unions, intersections, set-differences, finite sets and tuples, Cartesian products, and under images and preimages of internal functions.*
2. *If $A\in\mathscr{I}$ is an internal set, then the set of its internal subsets $\mathscr{P}(A)\cap\mathscr{I}\in\mathscr{I}$ is itself internal.*
3. *If A,B are internal sets, then the set of internal functions between them $Fun(A,B)\cap\mathscr{I}\in\mathscr{I}$ is itself internal.*

Proof. ......

Definition 2.47. An object that is not internal is called *external*.

Although "bad" with respect to *transfer*, there are relevant examples of external sets that are useful in the practice of nonstandard methods.

Example 2.48.

1. The set of infinitesimal hyperreal numbers is *external*. Indeed, it is a bounded subset of ${}^*\mathbb{R}$ without least upper bound.
2. The set of infinite hypernatural numbers is *external*. Indeed, it is a nonempty subset of ${}^*\mathbb{N}$ without a least element.
3. The set $\mathbb{N}$ of finite hypernatural numbers is *external*, otherwise the set-difference ${}^*\mathbb{N}\setminus\mathbb{N}$ of infinite numbers would be internal.

The above examples shows that ${}^*\mathscr{P}(\mathbb{N})\neq\mathscr{P}({}^*\mathbb{N})$ and ${}^*\mathscr{P}(\mathbb{R})\neq\mathscr{P}({}^*\mathbb{R})$. More generally, we have

Proposition 2.49.

1. *For every infinite set A, the set ${}^\sigma A=\{{}^*a\mid a\in A\}$ is external.*
2. *Every infinite hyperextension ${}^*A$ has external subsets, that is, the inclusion ${}^*\mathscr{P}(A)\subset\mathscr{P}({}^*A)$ is proper.*
3. *If the set A is infinite and B contains at least two elements, then the inclusion ${}^*Fun(A,B)\subset Fun({}^*A,{}^*B)$ is proper.*

[11] The subformula "$z=\{t\in x\mid\varphi(t,y_1,\ldots,y_n)\}$" is elementary because it denotes the conjuction of the two formulas:

$$\text{``}\forall t\in z\;(t\in x\text{ and }\varphi(t,y_1,\ldots,y_n))\text{''}\quad\text{and}\quad\text{``}\forall t\in x\;(\varphi(t,y_1,\ldots,y_n)\Rightarrow t\in z)\text{''}.$$

Proof. (1). Pick a surjective map $\psi : A \to \mathbb{N}$; then also the hyper-extension ${}^*\psi : {}^*A \to {}^*\mathbb{N}$ is surjective. If by contradiction ${}^\sigma A$ was internal, also its image under ${}^*\psi$ would be, and this is not possible, since

$$ {}^*\psi({}^\sigma A) \;=\; \{{}^*\psi({}^*a) \mid a \in A\} \;=\; \{{}^*(\psi(a)) \mid a \in A\} \;=\; \{\psi(a) \mid a \in A\} \;=\; \mathbb{N}. $$

(2). Notice first that A is infinite, because if $A = \{a_1, \dots, a_n\}$ was finite then also ${}^*A = \{{}^*a_1, \dots, {}^*a_n\}$ would be finite. Recall that ${}^*\mathscr{P}(A)$ is the set of all internal subsets of ${}^*A$. Since ${}^\sigma A \subset {}^*A$ is external by (1), ${}^\sigma A \in \mathscr{P}({}^*A) \setminus {}^*\mathscr{P}(A)$.

(3). Recall that ${}^*\mathrm{Fun}(A,B)$ is the set of all internal functions $f : {}^*A \to {}^*B$. Pick an external subset $X \subset A$, pick $b_1 \neq b_2$ in B, and let $f : {}^*A \to {}^*B$ be the function where $f(\alpha) = {}^*b_1$ if $\alpha \in X$ and $f(\alpha) = {}^*b_2$ if $\alpha \notin X$. Then f is external, otherwise the preimage $f^{-1}({}^*\{b_1\}) = X$ would be internal.

We warn the reader that getting familiar with the distinction between internal and external objects is probably the hardest part of learning nonstandard analysis.

2.5.1 Internal objects in the ultrapower model

The ultrapower model ${}^*\mathbb{R} = \mathbb{R}^I/\mathscr{U}$ that we introduced in Section 2.4 can be naturally extended so as to include also hyper-extensions of families of sets of real tuples, and of families of functions.

Let us start by observing that every I-sequence $T = \langle T_i \mid i \in I \rangle$ of sets of real numbers $T_i \subseteq \mathbb{R}$ determines a set $\widehat{T} \subseteq {}^*\mathbb{R}$ of hyperreal numbers in a natural way, by letting

$$ \widehat{T} \;=\; \big\{ [\sigma] \in {}^*\mathbb{R} \;\big|\; \{i \in I \mid \sigma(i) \in T_i\} \in \mathscr{U} \big\}. $$

Definition 2.50. If $\mathscr{F} \subseteq \mathscr{P}(\mathbb{R})$, then its *hyper-extension* ${}^*\mathscr{F} \subseteq {}^*\mathscr{P}(\mathbb{R})$ is defined as

$$ {}^*\mathscr{F} \;=\; \Big\{ \widehat{T} \;\Big|\; T : I \to \mathscr{F} \Big\}. $$

We remark that the same definition above also applies to families $\mathscr{F} \subseteq \mathscr{P}(\mathbb{R}^k)$ of sets of k-tuples, where for I-sequences $T : I \to \mathscr{P}(\mathbb{R}^k)$ one lets $\widehat{T} = \big\{ ([\sigma_1], \dots, [\sigma_k]) \in {}^*\mathbb{R}^k \;\big|\; \{i \in I \mid (\sigma_1(i), \dots, \sigma_k(i)) \in T_i\} \in \mathscr{U} \big\}$.

According to Definition 2.44, $A \subseteq {}^*\mathbb{R}$ is internal if and only if $A \in {}^*\mathscr{P}(\mathbb{R})$. So, in the ultrapower model, $A \subseteq {}^*\mathbb{R}$ is internal if and only if $A = \widehat{T}$ for some I-sequence $T : I \to \mathscr{P}(\mathbb{R})$.

Analogously as above, every I-sequence $F = \langle F_i \mid i \in I \rangle$ of real functions $F_i : \mathbb{R} \to \mathbb{R}$ determines a function $\widehat{F} : {}^*\mathbb{R} \to {}^*\mathbb{R}$ on the hyperreal numbers by letting for every $\sigma : I \to \mathbb{R}$:

$$ \widehat{F}([\sigma]) \;=\; [\langle F_i(\sigma(i)) \mid i \in I \rangle]. $$

The internal functions from ${}^*\mathbb{R}$ to ${}^*\mathbb{R}$ in the ultrapower model are precisely those that are determined by some I-sequence $F : I \to \mathrm{Fun}(\mathbb{R}, \mathbb{R})$.

Definition 2.51. If $\mathscr{G} \subseteq \mathrm{Fun}(\mathbb{R}, \mathbb{R})$, then its *hyper-extension* ${}^*\mathscr{G} \subseteq {}^*\mathrm{Fun}(\mathbb{R}, \mathbb{R})$ is defined as

$$ {}^*\mathscr{G} \;=\; \Big\{ \widehat{F} \;\Big|\; F : I \to \mathrm{Fun}(\mathbb{R}, \mathbb{R}) \Big\}. $$

If $F = \langle F_i \mid i \in I \rangle$ is an I-sequence of functions $F_i : \mathbb{R}^k \to \mathbb{R}$ of several variables, one extends the above definition by letting $\widehat{F} : {}^*\mathbb{R}^k \to {}^*\mathbb{R}$ be the function where for every $\sigma_1, \dots, \sigma_k : I \to \mathbb{R}$:

$$ \widehat{F}([\sigma_1], \dots, [\sigma_k]) \;=\; [\langle F_i(\sigma(i)) \mid i \in I \rangle]. $$

Indeed, also in this case, if $\mathscr{G} \subseteq \mathrm{Fun}(\mathbb{R}^k, \mathbb{R})$ then one puts ${}^*\mathscr{G} = \Big\{ \widehat{F} \;\Big|\; F : I \to \mathrm{Fun}(\mathbb{R}^k, \mathbb{R}) \Big\}$.

2.6 Hyperfinite sets

In this section we introduce a fundamental tool in nonstandard analysis, namely the class of *hyperfinite sets*. Although they may contain infinitely many elements, hyperfinite sets retain all nice "elementary properties" of finite sets; for this reason they are instrumental in applications as a convenient bridge between finite combinatorics and the *continuum*.

Definition 2.52. A *hyperfinite set* A is an element of the hyper-extension ${}^*\mathscr{F}$ of a family $\mathscr{F}$ of finite sets.

In particular, hyperfinite sets are internal objects.

Remark 2.53. In the ultrapower model, the hyperfinite subsets of ${}^*\mathbb{R}$ are defined according to Definition 2.50. Precisely, $A \subseteq {}^*\mathbb{R}$ is hyperfinite if and only if there exists a sequence $\langle T_i \mid i \in I\rangle$ of finite sets $T_i \subset \mathbb{R}$ such that $A = \widehat{T}$, that is, for every $\sigma : I \to \mathbb{R}$, $[\sigma] \in A \Leftrightarrow \{i \in I \mid \sigma(i) \in T_i\} \in \mathscr{U}$.

Let us start with the simplest properties of hyperfinite sets.

Proposition 2.54.

1. *A subset $A \subseteq {}^*X$ is hyperfinite if and only if $A \in {}^*Fin(X)$, where $Fin(X) = \{A \subseteq X \mid A \text{ is finite}\}$.*
2. *Every finite set of internal objects is hyperfinite.*
3. *A hyper-image ${}^*X$ is hyperfinite if and only if X is finite.*
4. *If $f : A \to B$ is an internal function, and $\Omega \subseteq A$ is a hyperfinite set, then its image $f(\Omega) = \{{}^*f(\xi) \mid \xi \in \Omega\}$ is hyperfinite as well. In particular, internal subsets of hyperfinite sets are hyperfinite.*

Proof. (1). If A is a hyperfinite subset of ${}^*X$, then A is internal, and hence $A \in {}^*\mathscr{P}(X)$. So, if $\mathscr{F}$ is a family of finite sets with $A \in {}^*\mathscr{F}$, then $A \in {}^*\mathscr{P}(X) \cap {}^*\mathscr{F} = {}^*(\mathscr{P}(X) \cap \mathscr{F}) \subseteq {}^*\text{Fin}(X)$. The converse implication is trivial.

(2). Let $A = \{a_1, \ldots, a_k\}$, and pick X_i such that $a_i \in {}^*X_i$. If $X = \bigcup_{i=1}^n X_i$, then $A \in {}^*\text{Fin}(X)$, as it is easily shown by applying *transfer* to the elementary property: "$\forall x_1, \ldots, x_k \in X \ \{x_1, \ldots, x_k\} \in \text{Fin}(X)$".

(3). Apply *transfer* to the following elementary property: "X is finite if and only if $X \in \mathscr{F}$ for some family $\mathscr{F}$ of finite sets."

(4). Pick X and Y with $A \in {}^*\mathscr{P}(X)$ and $B \in {}^*\mathscr{P}(Y)$. Then apply *transfer* to the property: "For every $C \in \mathscr{P}(X)$, for every $D \in \mathscr{P}(Y)$, for every $f \in \text{Fun}(C,D)$ and for every $F \in \text{Fin}(X)$ with $F \subseteq C$, the image $f(F) \in \text{Fin}(Y)$".

Example 2.55. For every pair $N < M$ of (possibly infinite) hypernatural numbers , the interval

$$[N,M] \;=\; \{\alpha \in {}^*\mathbb{N} \mid N \leq \alpha \leq M\}$$

is hyperfinite. Indeed, applying *transfer* to the property: "For every $x, y \in \mathbb{N}$ with $x < y$, the interval $[x,y] \in \text{Fin}(\mathbb{N})$", one obtains that $[N,M] \in {}^*\text{Fin}(\mathbb{N})$.[12] More generally, it is easily seen that also every bounded interval of hyperintegers is hyperfinite.

Definition 2.56. A *hyperfinite sequence* is an internal function whose domain is a hyperfinite set A.

Typical examples of hyperfinite sequences are defined on initial segments $[1,N] \subset {}^*\mathbb{N}$ of the hypernatural numbers. In this case we use notation $\langle \xi_\nu \mid \nu = 1, \ldots, N\rangle$

By *transfer* from the property: "For every nonempty finite set A there exists a unique $n \in \mathbb{N}$ such that A is in bijection with the segment $[1,n]$," one obtains that there is a well-posed definition of cardinality for hyperfinite sets.

Definition 2.57. The *internal cardinality* $|A|_h$ of a nonempty hyperfinite set A is the unique hypernatural number α such that there exists an internal bijection $f : [1,\alpha] \to A$.

Proposition 2.58.

1. *If the hyperfinite set A is finite, then $|A|_h = |A|$.*

[12] More formally, one transfers the formula: "$\forall x, y \in \mathbb{N}\ [(x < y \Rightarrow (\exists A \in \text{Fin}(\mathbb{N})\ \forall z\,(z \in A \leftrightarrow x \leq z \leq y))]$".

2. For any $\nu \in {}^*\mathbb{N}$*, we have* $|[1,\nu]|_h = \nu$*. More generally, we have* $|[\alpha,\beta]|_h = \beta - \alpha + 1$.

Proof. ......

When confusion is unlikely, we will drop the subscript and directly write $|A|$ to also denote the internal cardinality of a hyperfinite set A.

The following is a typical example of a property that hyperfinite sets inherit from finite sets. It is obtained by a straightforward application of *transfer*.

Proposition 2.59. *Every nonempty hyperfinite subset of* ${}^*\mathbb{R}$ *has a least and a greatest element.*

A relevant example of a hyperfinite set which is useful in applications is the following.

Definition 2.60. Fix an infinite $N \in {}^*\mathbb{N}$. The corresponding *hyperfinite grid* $\mathbb{H}_N \subset {}^*\mathbb{Q}$ is the hyperfinite set that determines a partition of the interval $[1,N] \subset {}^*\mathbb{R}$ of hyperreals into N-many intervals of equal infinitesimal length $1/N$. Precisely:

$$\mathbb{H}_N \;=\; \left\{ \pm \frac{\alpha}{N} \;\middle|\; \alpha = 0,1,\ldots,N \right\}.$$

We close this section with a couple of result about the (infinite) cardinalities of hyperfinite sets.

Proposition 2.61. *If* $\alpha \in {}^*\mathbb{N}$ *is infinite, then the corresponding interval* $[1,\alpha] \subset {}^*\mathbb{N}$ *has cardinality at least the cardinality of the* continuum*.*

Proof. For every real number $r \in (0,1)$, let

$$\psi(r) \;= \min\{\beta \in [1,\alpha] \mid r < \beta/\alpha\}.$$

Notice that the above definition is well-posed, because $\{\beta \in {}^*\mathbb{N} \mid r < \beta/\alpha\}$ is an internal bounded set of hypernatural numbers, and hence a hyperfinite set. The map $\psi : (0,1)_{\rightarrow}[1,\alpha]$ is 1-1; indeed, $\psi(r) = \psi(s) \Rightarrow |r-s| < 1/\alpha \Rightarrow r \sim s \Rightarrow r = s$ (recall that two real numbers that are infinitely close are necessarily equal). Thus, we obtain the desired inequality $\mathfrak{c} = |(0,1)| \leq |[1,\alpha]|$.

Corollary 2.62. *If A is internal, then either A is finite or A has at least the cardinality of the* continuum*. In consequence, every countably infinite set is external.*

Proof. It is easily seen by *transfer* that an internal set A is either hyperfinite, and hence it is in bijection with an interval $[1,\alpha] \subset {}^*\mathbb{N}$, or there exists an internal 1-1 function $f : {}^*\mathbb{N} \to A$. In the first case, if $\alpha \in \mathbb{N}$ is finite, then trivially A is finite; otherwise $|A| = [1,\alpha] \geq \mathfrak{c}$ by the previous proposition. In the second case, if α is any infinite hypernatural number , then $|A| \geq |{}^*\mathbb{N}| \geq |[1,\alpha]| \geq \mathfrak{c}$.

2.6.1 Hyperfinite sums

Similarly to finite sums of real numbers, one can consider *hyperfinite* sums of hyperfinite sets of hyperreal numbers.

Definition 2.63. If $f : A \to \mathbb{R}$ then for every nonempty hyperfinite subset $\Omega \subset {}^*A$, one defines the corresponding *hyperfinite sum* by setting:

$$\sum_{\xi \in \Omega} {}^*f(\xi) \;:=\; {}^*S_f(\Omega),$$

where $S_f : \mathrm{Fin}(A) \setminus \{\emptyset\} \to \mathbb{R}$ is the function $\{r_1 < \ldots < r_k\} \mapsto f(r_1) + \ldots + f(r_k)$.

As a particular case, if $a = \langle a_n \mid n \in \mathbb{N} \rangle$ is a sequence of real numbers and $\alpha \in {}^*\mathbb{N}$ is a hypernatural number , then the corresponding *hyperfinitely long sum* is defined as

$$\sum_{i=1}^{\alpha} a_i \;=\; {}^*S_a(\alpha)$$

where $S_a : \mathbb{N} \to \mathbb{R}$ is the function $n \mapsto a_1 + \ldots + a_n$.

Remark 2.64. More generally, the above definition can be extended to hyperfinite sums $\sum_{\xi\in\Omega} F(\xi)$ where $F : {}^*A \to {}^*\mathbb{R}$ is an internal function, and $\Omega \subseteq {}^*A$ is a nonempty hyperfinite subset. Precisely, in this case one sets $\sum_{\xi\in\Omega} F(\xi) = {}^*\mathscr{S}(F,\Omega)$, where $\mathscr{S} : \mathrm{Fun}(A,\mathbb{R}) \times (\mathrm{Fin}(A)\setminus\{\emptyset\}) \to \mathbb{R}$ is the function $(f,G) \mapsto \sum_{x\in G} f(x)$.

Let us mention in passing that hyperfinite sums can be used to directly define integrals. Indeed, if $N \in {}^*\mathbb{N}$ is any infinite hypernatural number and $\mathbb{H}$ is the corresponding hyperfinite grid (see Definition 2.60), then for every $f : \mathbb{R} \to \mathbb{R}$ and for every $A \subseteq \mathbb{R}$, one defines the *grid integral* by putting:

$$\int_A f(x)d_{\mathbb{H}}(x) \;=\; \mathrm{st}\left(\sum_{\xi\in\mathbb{H}\cap{}^*A} {}^*f(\xi)\right).$$

Notice that the above definition applies to *every* real function f and to *every* subset A. Moreover, it can be shown that if $f : [a,b] \to \mathbb{R}$ is a Riemann integrable function defined on an interval, then the grid integral coincides with the usual Riemann integral.

2.7 Overflow and underflow principles

Proposition 2.65 (Overflow principles).

1. *$A \subseteq \mathbb{N}$ is infinite if and only if its hyper-extension ${}^*A$ contains an infinite number.*
2. *If $B \subseteq {}^*\mathbb{N}$ is internal and $B\cap\mathbb{N}$ is infinite then B contains an infinite number.*
3. *If $B \subseteq {}^*\mathbb{N}$ is internal and $\mathbb{N} \subseteq B$ then $[1,\alpha] \subseteq B$ for some infinite $\alpha \in {}^*\mathbb{N}$.*

Proof. Item 1 follows from Propositions 2.22 and 2.27. For item 2, suppose that B does not contain an infinite number. Then B is bounded above in ${}^*\mathbb{N}$. By *transfer*, B has a maximum, which is necessarily an element of $\mathbb{N}$, contradicting that $B\cap\mathbb{N}$ is infinite. For item 3, let $C := \{\alpha \in {}^*\mathbb{N} \;:\; [1,\alpha] \subseteq B\}$. Then C is internal and $\mathbb{N} \subseteq C$ by assumption. By item 2 applied to C, there is $\alpha \in C$ that is infinite; this α is as desired.

Proposition 2.66 (Underflow principles).

1. *If $B \subseteq {}^*\mathbb{N}$ is internal and B contains arbitrarily small infinite numbers, then B contains a finite number.*
2. *If $B \subseteq {}^*\mathbb{N}$ is internal and $[\alpha,+\infty) \subseteq B$ for every infinite $\alpha \in {}^*\mathbb{N}$ then then $[n,+\infty) \subseteq B$ for some finite $n \in \mathbb{N}$.*

Proof. For item 1, suppose that B does not contain a finite number. Then the minimum of B is necessarily infinite, contradicting the assumption that B contains arbitrarily small infinite numbers. Item 2 follows by applying item 1 to the internal set $C := \{\alpha \in {}^*\mathbb{N} \;:\; [\alpha,+\infty) \subseteq B\}$.

In practice, one often says they are using *overflow* when they are using any of the items in Proposition 2.65 and likewise for *underflow*. Below we will show a use of *overflow* in graph theory.

2.7.1 An application to graph theory

Recall that a *graph* is a set V (the set of *vertices*) endowed with an anti-reflexive and symmetric binary relation E (the set of *edges*). Notice that if $G = (V,E)$ is a graph then also its hyper-extension ${}^*G = ({}^*V, {}^*E)$ is a graph. By assuming as usual that ${}^*v = v$ for all $v \in V$, one has that G is a sub-graph of ${}^*G$. A graph $G = (V,E)$ is *locally finite* if for every vertex $v \in V$, its *set of neighbors* $N_G(v) = \{u \in V \mid \{u,v\} \in E\}$ is finite. One has the following simple nonstandard characterization.

Proposition 2.67. *A graph $G = (V,E)$ is locally finite if and only if ${}^*(N_G(v)) \subseteq V$ for every $v \in V$.*

Proof. If G is locally finite then for every $v \in V$ the set of its neighbors $N_G(v) = \{u_1, \dots, u_n\}$ is finite, and so ${}^*N_G(v) = \{{}^*u_1, \dots, {}^*u_n\} = \{u_1, \dots, u_n\} \subseteq V$. Conversely, if G is not locally finite, then there exists a vertex $v \in V$ such that $N_G(v)$ is infinite, and we can pick an element $\tau \in {}^*(N_G(v)) \setminus N_G(v)$. Now, $\tau \notin V$, as otherwise $\tau \in {}^*(N_G(v)) \cap V = N_G(v)$, a contradiction.

Recall that a *finite path* in a graph $G = (V, E)$ is a finite sequence $\langle v_i \mid i = 1, \dots, n \rangle$ of pairwise distinct vertexes such that $\{v_i, v_{i+1}\} \in E$ for every $i < n$. A graph is *connected* if for every pair of distinct vertices u, u' there exists a finite path $\langle v_i \mid i = 1, \dots, n \rangle$ where $v_1 = u$ and $v_n = u'$. An *infinite path* is a sequence $\langle v_i \mid i \in \mathbb{N} \rangle$ of pairwise distinct vertexes such that $\{v_i, v_{i+1}\} \in E$ for every $i \in \mathbb{N}$.

Theorem 2.68 (König's Lemma - I). *Every infinite connected graph that is locally finite contains an infinite path.*

Proof. Given a locally finite connected graph $G = (V, E)$ where V is infinite, pick $u \in V$ and $\tau \in {}^*V \setminus V$. Since G is connected, by *transfer* there exists a hyperfinite sequence $\langle v_i \mid i = 1, \dots, \mu \rangle$ for some $\mu \in {}^*\mathbb{N}$ where $v_1 = u$ and $\{v_i, v_{i+1}\} \in {}^*E$ for every $i < \mu$. By local finiteness, ${}^*(N_G(v_1)) \subseteq V$ and so $v_2 \in V$ and $\{v_1, v_2\} \in E$. Then, by induction, it is easily verified that the restriction $\langle v_i \mid i \in \mathbb{N} \rangle$ of the above sequence to the finite indexes is an infinite path in G.

A simple but relevant application of *overflow* proves the following equivalent formulation in terms of trees.

Theorem 2.69 (König's Lemma - II). *Every infinite, finitely branching tree has an infinite path.*

Proof. Let T_n denote the nodes of the tree of height n. Since T is finitely branching, each T_n is finite. Since T is infinite, each $T_n \neq \emptyset$. By *overflow*, there is $N > \mathbb{N}$ such that $T_N \neq \emptyset$. Fix $x \in T_N$. Then $\{y \in T \mid y \leq x\}$ is an infinite branch in T.

2.8 The saturation principle

The *transfer principle* is all that one needs to to develop the machinery of nonstandard analysis, but for advanced applications another property is also necessary, namely:

Definition 2.70. *Countable Saturation Principle:* Suppose $\{B_n\}_{n \in \mathbb{N}} \subseteq {}^*A$ is a countable family of internal sets with the finite intersection property. Then $\bigcap_{n \in \mathbb{N}} B_n \neq \emptyset$.

Countable saturation will be instrumental in the definition of *Loeb measures*. In several contexts, stronger saturation principles are assumed where also families of larger size are allowed. Precisely, if κ is a given uncountable cardinal, then one considers the following.

Definition 2.71. κ-*saturation property*: If $\mathscr{B} \subseteq {}^*A$ is a family of internal subsets of cardinality where $|\mathscr{B}| < \kappa$, and if $\mathscr{B}$ has the finite intersection property, then $\bigcap_{B \in \mathscr{B}} B \neq \emptyset$.

Notice that, in this terminology, countable saturation is $\aleph_1$-saturation.

In addition to countable saturation, in the applications presented in this book, we will only use the following weakened version of κ-saturation, where only families of hyper-extensions are considered.

Definition 2.72. κ-*enlarging property*: Suppose $\mathscr{F} \subseteq \mathscr{P}(A)$ has cardinality $|\mathscr{F}| < \kappa$. If $\mathscr{F}$ has the finite intersection property, then $\bigcap_{F \in \mathscr{F}} {}^*F \neq \emptyset$.[13]

As a first important application of the enlarging property, one obtains that sets are included in a hyperfinite subset of their hyper-extension.

Proposition 2.73. *If the κ-enlarging property holds, then for every set X of cardinality $|X| < \kappa$ there exists a hyperfinite subset $H \subseteq {}^*X$ such that $X \subseteq H$.*

[13] We remark that the enlarging property is strictly weaker than saturation, in the sense that for every infinite κ there are models of nonstandard analysis where the κ-enlarging property holds but κ-saturation fails.

Proof. For each $a \in X$, let $X_a := \{Y \subseteq X \ : \ Y \text{ is finite and } a \in X\}$. One then applies the κ-enlarging property to the family $\mathscr{F} := \{X_a \ : \ a \in X\}$ to obtain $H \in \bigcap_{a\in X} {}^*X_a$; such H is as desired.

Although it will not play a role in this book, we would be remiss if we did not mention the following example:

Example 2.74. Let (X,τ) be a topological space with *character* $< \kappa$, that is, such that each point $x \in X$ has a base of neighborhoods $\mathscr{N}_x$ of cardinality less than κ. If we assume the κ-enlarging property, the intersection $\mu(x) = \bigcap_{U\in\mathscr{N}_x} {}^*U \neq \emptyset$. In the literature, $\mu(x)$ is called the *monad* of x. Monads are the basic ingredient in applying nonstandard analysis to topology, starting with the following characterizations (see, *e.g.*, [37] Ch.III):

- X is Hausdorff if and only if $\mu(x) \cap \mu(y) = \emptyset$ whenever $x \neq y$;
- $A \subseteq X$ is *open* if and only if for every $x \in A$, $\mu(a) \subseteq {}^*A$;
- $C \subseteq X$ is *closed* if and only if for every $x \notin C$, $\mu(x) \cap {}^*C = \emptyset$;
- $K \subseteq X$ is *compact* if and only if ${}^*K \subseteq \bigcup_{x\in K} \mu(x)$.

2.8.1 Saturation in the ultrapower model

Let us show here that the ultrapower model ${}^*\mathbb{R} = \mathbb{R}^I/\mathscr{U}$ introduced in Section 2.4 also accommodates saturation. Let us start with a direct combinatorial proof in the case of ultrapowers modulo ultrafilters on $\mathbb{N}$.

Theorem 2.75. *For every non-principal ultrafilter $\mathscr{U}$ on $\mathbb{N}$, the corresponding ultrapower model satisfies countable saturation.*

Proof. Let $\{B_n\}$ be a countable family of internal subsets of ${}^*\mathbb{R}$ with the finite intersection property. For every n, pick a function $T_n : \mathbb{N} \to \mathscr{P}(\mathbb{R})$ such that

$$B_n \;=\; \widehat{T_n} \;=\; \{[\sigma] \in {}^*\mathbb{R} \mid \{i \in \mathbb{N} \mid \sigma(i) \in T_n\} \in \mathscr{U}\}\,.$$

For any fixed n, pick an element $\tau(n) \in T_1(n) \cap \cdots \cap T_n(n)$ if that intersection is nonempty; otherwise, pick an element $\tau(n) \in T_1(n) \cap \cdots \cap T_{n-1}(n)$ if that intersection is nonempty; and so forth until $\tau(n)$ is defined. We agree that $\tau(n) = 0$ in case $T_1(n) = \emptyset$. By the definition of τ, one has the following property:

- If $T_1(n) \cap \cdots \cap T_k(n) \neq \emptyset$ and $n \geq k$ then $\tau(n) \in T_1(n) \cap \ldots \cap T_k(n)$.

Now let k be fixed. By the finite intersection property, $\widehat{T_1} \cap \ldots \cap \widehat{T_k} \neq \emptyset$, so there exists $\sigma : \mathbb{N} \to \mathbb{R}$ such that $\Lambda_j = \{i \in \mathbb{N} \mid \sigma(i) \in T_j(i)\} \in \mathscr{U}$ for every $j = 1, \ldots, k$. In particular, the set of indexes $\Gamma(k) = \{i \in \mathbb{N} \mid T_1(i) \cap \ldots \cap T_k(i) \neq \emptyset\} \in \mathscr{U}$ because it is a superset of $\Lambda_1 \cap \ldots \cap \Lambda_k \in \mathscr{U}$. But then the set $\{i \in \mathbb{N} \mid \tau(i) \in T_1(i) \cap \ldots \cap T_k(i)\} \in \mathscr{U}$ because it is a superset of $\{i \in \Gamma(k) \mid i \geq k\} \in \mathscr{U}$. We conclude that $[\tau] \in \widehat{T_1} \cap \ldots \cap \widehat{T_k}$. As this holds for every k, the proof is completed.

The above result can be extended to all ultrapower models where the ultrafilter $\mathscr{U}$ on I is *countably incomplete* (recall that every non-principal ultrafilter on $\mathbb{N}$ is countably incomplete).

Theorem 2.76. *For every infinite cardinal κ there exist ultrafilters $\mathscr{U}$ on the set $I = Fin(\kappa)$ of finite parts of κ such that the corresponding ultrapower model satisfies the κ^+-enlarging property.*

Proof. For every $x \in \kappa$, let $\widehat{x} = \{a \in I \mid x \in a\}$. Then trivially the family $\mathscr{X} = \{\widehat{x} \mid x \in \kappa\}$ has the finite intersection property. We claim that every ultrafilter $\mathscr{U}$ that extends $\mathscr{X}$ has the desired property.

Suppose that the family $\mathscr{F} = \{B_x \mid x \in \kappa\} \subseteq \mathscr{P}(A)$ satisfies the finite intersection property. Then we can pick a sequence $\sigma : I \to A$ such that $\sigma(a) \in \bigcap_{x\in a} A_x$ for every $a \in I$. The proof is completed by noticing that $[\sigma] \in {}^*A_x$ for every $x \in \kappa$, since $\{a \in I \mid \sigma(a) \in A_x\} \supseteq \widehat{x} \in \mathscr{U}$.

A stronger result holds, but we will not prove it here because it takes a rather technical proof, and we do not need that result in the applications presented in this book.

Theorem 2.77. *For every infinite cardinal κ there exist ultrafilters $\mathscr{U}$ on κ (named κ^+-*good ultrafilters*) such that the corresponding ultrapower models satisfy the κ^+-saturation property.*

Proof. See [8, §6.1].

2.9 Hyperfinite approximation

As established in Proposition 2.73, in sufficiently saturated structures, hyperfinite sets can be conveniently used as "approximations" of infinite structure. The fact that they behave as finite sets makes them particularly useful objects in applications of nonstandard analysis. In this section we will see a few examples to illustrate this. We assume that the nonstandard extension satisfies the κ-enlarging property, where κ is larger than the cardinality of the objects under consideration.

Theorem 2.78. *Every infinite set can be linearly ordered.*

Proof. Let X be an infinite set and take hyperfinite $H \subseteq {}^*X$ such that $\{{}^*x \mid x \in X\} \subseteq H$. By *transfer* applied to the corresponding property of finite sets, H can be linearly ordered, whence so can $\{{}^*x \mid x \in X\}$, and hence X.

The next theorem is a generalization of the previous one:

Theorem 2.79. *Every partial order on a set can be extended to a linear order.*

Proof. We leave it as an easy exercise by induction to show that every partial order on a finite set can be extended to a linear order. Thus, we may precede as in the previous theorem. This time, H is endowed with the partial order it inherits from ${}^*X$, whence, by *transfer*, this partial order can be extended to a linear order. This linear order restricted to X extends the original partial order on X.

Theorem 2.80. *A graph is k-colorable if and only if every finite subgraph is k-colorable.*

Proof. Suppose that G is a graph such that every finite subgraph is k-colorable. Embed G into a hyperfinite subgraph H of ${}^*G$. By *transfer*, H can be k-colored. The restriction of this k-coloring to G is a k-coloring of G.

The next instance actually plays an important role. Say that $f : \mathbb{N} \to \mathbb{N}$ is *fixed-point free* if $f(n) \neq n$ for all $n \in \mathbb{N}$.

Theorem 2.81. *Suppose that $f : \mathbb{N} \to \mathbb{N}$ is fixed-point free. Then there is a function $c : \mathbb{N} \to \{1,2,3\}$ (that is, a* 3-coloring of *$\mathbb{N}$) such that $c(g(n)) \neq c(n)$ for all $n \in \mathbb{N}$.*

Proof. In order to use hyperfinite approximation, we first need a finitary version of the theorem:

Claim: For every finite subset $F \subseteq \mathbb{N}$, there is a 3-coloring c_F of F such that $c(f(n)) \neq c(n)$ whenever $n, f(n) \in F$.
Proof of Claim: We prove the claim by induction on the cardinality of F, the case $|F| = 1$ being trivial since F never contains both n and $f(n)$. Now suppose that $|F| > 1$. Fix $m \in F$ such that $|f^{-1}(m) \cap F| \leq 1$; such an m clearly exists by the Pigeonhole principle. Let $G := F \setminus \{m\}$. By the induction assumption, there is a 3-coloring c_G of G such that $c(f(n)) \neq c(n)$ whenever $n, f(n) \in G$. One extends c_G to a 3-coloring c_F of F by choosing $c_F(m)$ different from $c_G(f(m))$ (if $f(m) \in G$) and different from $c_G(k)$ if $k \in G$ is such that $f(k) = m$ (if there is such k); since we have three colors to choose from, this is clearly possible. The coloring c_F is as desired.

Now that the claim has been proven, let $H \subseteq {}^*\mathbb{N}$ be hyperfinite such that $\mathbb{N} \subseteq H$. By *transfer*, there is an internal 3-coloring c_H of H such that $c(f(n)) \neq c(n)$ whenever $n, f(n) \in H$. Since $n \in \mathbb{N}$ implies $n, f(n) \in H$, we see that $c_H|\mathbb{N}$ is a 3-coloring of H as desired.

2.10 Further reading

We finish this chapter with a few suggestions for further readings. A rigorous formulation and a detailed proof of the *transfer principle* can be found in Ch.4 of the textbook [19], where the *ultrapower* model is considered. See also §4.4 of [8] for the foundations of nonstandard analysis in full generality. A nice introduction of nonstandard methods for number theorists, including a number of examples, is given in [30] (see also [27]). Finally, a full development of nonstandard analysis can be found in several monographies of the existing literature; see *e.g.* the classical H.J. Keisler's book [32], or the comprehensive collections of surveys in [2].

Chapter 3
Hyperfinite generators of ultrafilters

Throughout this chapter, we fix an infinite set S and we assume that ${}^*s = s$ for every $s \in S$, so that $S \subseteq {}^*S$.

3.1 Hyperfinite generators

An important observation is that elements of ${}^*S$ generate ultrafilters on S:

Exercise 3.1. Suppose that $\alpha \in {}^*S$. Set $\mathscr{U}_\alpha := \{A \subseteq S \ : \ \alpha \in {}^*A\}$.

1. $\mathscr{U}_\alpha$ is an ultrafilter on S.
2. $\mathscr{U}_\alpha$ is principal if and only if $\alpha \in S$.

We call $\mathscr{U}_\alpha$ the *ultrafilter on S generated by* α. Note that in the case that $\alpha \in S$, there is no conflict between the notation $\mathscr{U}_\alpha$ in this chapter and the notation $\mathscr{U}_\alpha$ from Chapter 1.

Exercise 3.2. For $k \in \mathbb{N}$ and $\alpha \in {}^*\mathbb{N}$, show that $k\mathscr{U}_\alpha = \mathscr{U}_{k\alpha}$.

Recall from Exercise 1.11 that, for for every function $f : S \to T$ and for every ultrafilter $\mathscr{U}$ on S, the *image ultrafilter* $f(\mathscr{U})$ is the ultrafilter on T defined by setting

$$f(\mathscr{U}) \ = \ \{B \subseteq T \mid f^{-1}(B) \in \mathscr{U}\}.$$

Exercise 3.3. Show that $f(\mathfrak{U}_\alpha) = \mathfrak{U}_{{}^*f(\alpha)}$.

Since there are at most $2^{2^{|S|}}$ ultrafilters on S, if the nonstandard extension is κ-saturated for $\kappa > 2^{2^{|S|}}$, then $|{}^*S| > 2^{2^{|S|}}$ and we see that there must exist distinct $\alpha, \beta \in {}^*S \setminus S$ such that $\mathscr{U}_\alpha = \mathscr{U}_\beta$. This leads to the following notion, which is of central importance in Part II of this book.

Definition 3.4. Given $\alpha, \beta \in {}^*S$, we say that α and β are *u-equivalent*, written $\alpha \sim \beta$, if $\mathscr{U}_\alpha = \mathscr{U}_\beta$.

Here are some useful properties of this relation on ${}^*S$:

Proposition 3.5.

1. *If $\alpha, \beta \in S$, then $\alpha \sim \beta$ if and only if $\alpha = \beta$.*
2. *Suppose that $f : \mathbb{N} \to \mathbb{N}$ and $\alpha \sim \beta$. Then $f(\alpha) \sim f(\beta)$.*
3. *Suppose that $f : \mathbb{N} \to \mathbb{N}$ and α is such that $f(\alpha) \sim \alpha$. Then $f(\alpha) = \alpha$.*

Proof. Items (1) and (2) are easy and left to the reader. We now prove (3). Suppose that $f(\alpha) \neq \alpha$. Let $A := \{n \in \mathbb{N} \ : \ f(n) \neq n\}$. Let $g : \mathbb{N} \to \mathbb{N}$ be fixed-point free such that $f|A = g|A$. By Theorem 2.81, there is a 3-coloring c of $\mathbb{N}$ such that $c(g(n)) \neq c(n)$ for all $n \in \mathbb{N}$. In particular, $c(g(\alpha)) \neq c(\alpha)$. Since $\alpha \in {}^*A$, we have $f(\alpha) = g(\alpha)$, so $c(f(\alpha)) \neq c(\alpha)$. Setting, $i := c(\alpha)$ and $X := \{n \in \mathbb{N} \ : \ c(n) = i\}$, we have that $\alpha \in {}^*X$ but $f(\alpha) \notin {}^*X$, whence $f(\alpha) \not\sim \alpha$.

We have seen that elements of ${}^*S$ generate ultrafilters on S. Under sufficient saturation, the converse holds:

Proposition 3.6. *Assume that the nonstandard universe has the $(2^{|S|})^+$-enlarging property. Then for every $\mathscr{U} \in \beta S$, there is $\alpha \in {}^*S$ such that $\mathscr{U} = \mathscr{U}_\alpha$.*

Proof. Fix $\mathscr{U} \in \beta S$. It is clear that $\mathscr{U}$ is a family of subsets of S of cardinality $|\mathscr{U}| \leq 2^{|S|}$ with the finite intersection property, whence, by the $(2^{|S|})^+$-enlarging property, there is $\alpha \in \bigcap_{A \in \mathscr{U}} {}^*A$. Observe now that $\mathscr{U} = \mathscr{U}_\alpha$.

Exercise 3.7. Assume the $(2^{|S|})^+$-enlarging property. Show that for every non-principal $\mathscr{U} \in \beta S \setminus S$ there exist $|{}^*\mathbb{N}|$-many α such that $\mathscr{U} = \mathscr{U}_\alpha$.

By the previous proposition, the map $\alpha \mapsto \mathscr{U}_\alpha : {}^*S \to \beta S$ is surjective. This suggests that we define a topology on ${}^*S$, called the *u-topology on* ${}^*S$, by declaring the sets ${}^*A$, for $A \subseteq S$, to be the basic open sets.[1] This topology, while (quasi)compact by saturation, is not Hausdorff. In fact, $\alpha, \beta \in {}^*S$ are *not* separated in the *u*-topology precisely when $\alpha \sim \beta$. Passing to the separation, we get a compact Hausdorff space ${}^*S/\sim$ and the surjection ${}^*S \to \beta S$ defined above descends to a homeomorphism between the quotient space ${}^*S/\sim$ and βS. So, while βS is the "largest" Hausdorff compactification of the discrete space S, a (sufficiently saturated) hyper-extension of S is an even larger space, which is still compact (but non-Hausdorff) and have βS as a quotient.

3.2 The case of a semigroup again

Let us now suppose, once again, that S is the underlying set of a semigroup $(S, \cdot)$. One might guess that, for $\alpha, \beta \in {}^*S$, we have that the equation $\mathscr{U}_{\alpha \cdot \beta} = \mathscr{U}_\alpha \odot \mathscr{U}_\beta$ holds. Unfortunately, this is not the case:

Example 3.8. Fix $\alpha \in {}^*\mathbb{N} \setminus \mathbb{N}$. We show that there is $\beta \in {}^*\mathbb{N}$ such that $\mathscr{U}_\alpha \oplus \mathscr{U}_\beta \neq \mathscr{U}_\beta \oplus \mathscr{U}_\alpha$. For this β, we must have that either $\mathscr{U}_\alpha \oplus \mathscr{U}_\beta \neq \mathscr{U}_{\alpha+\beta}$ or $\mathscr{U}_\beta \oplus \mathscr{U}_\alpha \neq \mathscr{U}_{\beta+\alpha}$.

Without loss of generality, we may assume that α is even. (The argument when α is odd is exactly the same.) Take $\nu \in {}^*\mathbb{N}$ such that $\nu^2 \leq \alpha < (\nu+1)^2$. Let $A = \bigcup_{n \text{ even}} [n^2, (n+1)^2)$. First suppose that $(\nu+1)^2 - \alpha$ is finite. In this case, we let $\beta := \nu^2$. Note that $\{n \in \mathbb{N} \,:\, (A-n) \in \mathscr{U}_\alpha\} = \{n \in \mathbb{N} \,:\, n + \alpha \in {}^*A\}$ is finite by assumption, whence not in $\mathscr{U}_\beta$. Consequently, $A \notin \mathscr{U}_\beta \oplus \mathscr{U}_\alpha$. However, since $\alpha - \beta$ is necessarily infinite, we have $\{n \in \mathbb{N} \,:\, (A-n) \in \mathscr{U}_\beta\} = \{n \in \mathbb{N} \,:\, n + \beta \in {}^*A\} = \mathbb{N}$, whence a member of $\mathscr{U}_\alpha$ and thus $A \in \mathscr{U}_\alpha \oplus \mathscr{U}_\beta$.

If $(\nu+1)^2 - \alpha$ is infinite, then set $\beta := (\nu+1)^2$. An argument analogous to the argument in the previous paragraph shows that $A \in \mathscr{U}_\alpha \oplus \mathscr{U}_\beta$ but $A \notin \mathscr{U}_\beta \oplus \mathscr{U}_\alpha$.

Remark 3.9. The previous argument also gives a nonstandard proof of the fact that the center of $(\beta\mathbb{N}, \oplus)$ is precisely the set of principal ultrafilters.

The previous example notwithstanding, there is a connection between $(\beta S, \cdot)$ and the nonstandard extension of the semigroup $(S, \cdot)$. To see this, for notational cleanliness, let us switch over to writing the semigroup operation of S by $+$ (even though the semigroup need not be commutative). Fix $\alpha, \beta \in {}^*S$. For $A \subseteq S$, set

$$A_\beta := A - \mathscr{U}_\beta = \{s \in S \,:\, A - s \in \mathscr{U}_\beta\} = \{s \in S \,:\, s + \beta \in {}^*A\}.$$

By transfer, we have that ${}^*A_\beta = \{\gamma \in {}^*S \,:\, \gamma + {}^*\beta \in {}^{**}A\}$. It follows that, for $A \subseteq S$, we have

$$A \in \mathscr{U}_\alpha \oplus \mathscr{U}_\beta \Leftrightarrow A_\beta \in \mathscr{U}_\alpha \Leftrightarrow \alpha \in {}^*A_\beta \Leftrightarrow \alpha + {}^*\beta \in {}^{**}A \Leftrightarrow A \in \mathscr{U}_{\alpha + {}^*\beta},$$

that is, $\mathscr{U}_\alpha \oplus \mathscr{U}_\beta = \mathscr{U}_{\alpha + {}^*\beta}$.

Wait! What is ${}^{**}A$? And what is ${}^*\beta$? Well, our intentional carelessness was intended to motivate the need to be able to take nonstandard extensions of nonstandard extensions, that is, to be able to consider *iterated nonstandard extensions*. Once we

[1] This topology is usually named "*S*-topology" in the literature of nonstandard analsyis, where the "S" stands for "standard".

give this precise meaning in the next chapter, the above informal calculation will become completely rigorous and we have a precise connection between the operation $\oplus$ on βS and the operation $+$ on ${}^{**}S$.

We should also mention that it is possible for the equality $\mathscr{U}_\alpha \oplus \mathscr{U}_\beta = \mathscr{U}_{\alpha+\beta}$ to be valid. Indeed, this happens when α and β are *independent* in a certain sense; see [].

Chapter 4
Many stars: iterated nonstandard extensions

4.1 The foundational perspective

As we saw in the previous chapter, it is useful in applications to consider iterated hyper-extensions of the natural numbers, namely ${}^*\mathbb{N}$, ${}^{**}\mathbb{N}$, ${}^{***}\mathbb{N}$, and so forth. A convenient foundational framework where such iterations make sense can be obtained by considering models of nonstandard analysis where the *standard universe* and the *nonstandard universe* coincide.[1] In other words, one works with a *star map*

$$*:\mathbb{V}\to\mathbb{V}$$

from a universe into itself. Clearly, in this case every hyper-extension ${}^*X$ belongs to the universe $\mathbb{V}$, so one can apply the *star map* to it, and obtain the “second level” hyper-extension ${}^{**}X$; and so forth.

Let us stress that the *transfer principle* in this context must be handled with much care. The crucial point to keep in mind is that in the equivalence

$$P(A_1,\dots,A_n) \iff P({}^*A_1,\dots,{}^*A_n),$$

the considered objects $A_1,\dots,A_n$ could be themselves iterated hyper-extensions; in this case, one simply has to add one more “star”. Let us elaborate on this with a few examples.

Example 4.1. Recall that $\mathbb{N}$ is an initial segment of ${}^*\mathbb{N}$, that is,

$$\mathbb{N}\subset{}^*\mathbb{N} \text{ and } \forall x\in\mathbb{N}\;\forall y\in{}^*\mathbb{N}\setminus\mathbb{N}\;\; x<y.$$

Thus, by *transfer*, we obtain that:

$${}^*\mathbb{N}\subset{}^{**}\mathbb{N} \text{ and } \forall x\in{}^*\mathbb{N}\;\forall y\in{}^{**}\mathbb{N}\setminus{}^*\mathbb{N}\;\; x<y.$$

This means that ${}^*\mathbb{N}$ is a proper initial segment of the double hyper-image ${}^{**}\mathbb{N}$, that is, every element of ${}^{**}\mathbb{N}\setminus{}^*\mathbb{N}$ is larger than all element in ${}^*\mathbb{N}$.

Example 4.2. If $\eta\in{}^*\mathbb{N}\setminus\mathbb{N}$, then by *transfer* ${}^*\eta\in{}^{**}\mathbb{N}\setminus{}^*\mathbb{N}$, and hence $\eta<{}^*\eta$. Then, again by *transfer*, one obtains that the elements ${}^*\eta,{}^{**}\eta\in{}^{***}\mathbb{N}$ are such that ${}^*\eta<{}^{**}\eta$; and so forth.

The above example clarifies that the simplifying assumption ${}^*r=r$ that was adopted for every $r\in\mathbb{R}$ cannot be extended to hold for all hypernatural numbers . Indeed, we just proved that $\eta\neq{}^*\eta$ for every $\eta\in{}^*\mathbb{N}\setminus\mathbb{N}$.

Example 4.3. Since $\mathbb{R}\subset{}^*\mathbb{R}$, by *transfer* it follows that ${}^*\mathbb{R}\subset{}^{**}\mathbb{R}$. If $\varepsilon\in{}^*\mathbb{R}$ is a positive infinitesimal, that is, if $0<\varepsilon<r$ for every positive $r\in\mathbb{R}$, then by *transfer* we obtain that $0<{}^*\varepsilon<\xi$ for every positive $\xi\in{}^*\mathbb{R}$. In particular, ${}^*\varepsilon<\varepsilon$.

Recall that, by Proposition 2.19, for every elementary formula $\varphi(x,y_1,\dots,y_n)$ and for all objects $B,A_1,\dots,A_n$, one has that

$${}^*\{y\in B\mid P(y,A_1,\dots,A_n)\} \;=\; \{y\in{}^*B\mid P(y_1,{}^*A_1,\dots,{}^*A_n)\}. \quad (\dagger)$$

[1] A construction of such star maps is given in Section A.1.4 of the foundational appendix.

Of course one can apply the above property also when (some of) the parameters are hyper-extensions.

Remark 4.4. In nonstandard analysis, a hyper-extension ${}^*A$ is often called a "standard" set. This terminology comes from the fact that – in the usual approaches – one considers a star map $*:\mathbb{S}\to\mathbb{V}$ between the "standard universe" $\mathbb{S}$ and a "nonstandard universe" $\mathbb{V}$. Objects $A\in\mathbb{S}$ are named "standard" and, with some ambiguity, also their hyper-extensions ${}^*A$ are named "standard".[2] Let us stress that the name "standard" would be misleading in our framework, where there is just one single universe, namely the universe of *all* mathematical objects. Those objects of our universe that happen to be in the range of the star map, are called hyper-extensions.

4.2 Revisiting hyperfinite generators

In this subsection, we let $(S,+)$ denote an infinite semigroup. Now that we have the ability to take iterated nonstandard extensions, we can make our discussion from the end of Section 3.2 precise:

Proposition 4.5. *For $\alpha,\beta\in{}^*S$, we have $\mathscr{U}_\alpha\oplus\mathscr{U}_\beta=\mathscr{U}_{\alpha+{}^*\beta}$.*

Proof. By equation (†) from the previous section, we have that ${}^*A_\beta=\{\gamma\in{}^*S\ :\ \gamma+{}^*\beta\in{}^{**}A\}$. It follows that, for $A\subseteq S$, we have

$$A\in\mathscr{U}_\alpha\oplus\mathscr{U}_\beta\Leftrightarrow A_\beta\in\mathscr{U}_\alpha\Leftrightarrow\alpha\in{}^*A_\beta\Leftrightarrow\alpha+{}^*\beta\in{}^{**}A\Leftrightarrow A\in\mathscr{U}_{\alpha+{}^*\beta},$$

that is, $\mathscr{U}_\alpha\oplus\mathscr{U}_\beta=\mathscr{U}_{\alpha+{}^*\beta}$.

Exercise 4.6. The *tensor product* $\mathscr{U}\otimes\mathscr{V}$ of two ultrafilters on S is the ultrafilter on $S\times S$ defined by:

$$\mathscr{U}\otimes\mathscr{V}\ =\ \{C\subseteq S\times S\mid\{s\in S\mid C_s\in\mathscr{V}\}\in\mathscr{U}\},$$

where $C_s=\{t\in S\mid(s,t)\in C\}$ is vertical s-fiber of C. If $\alpha,\beta\in{}^*S$, prove that $\mathscr{U}_\alpha\otimes\mathscr{U}_\beta=\mathscr{U}_{(\alpha,{}^*\beta)}$.

We can extend this discussion to elements of higher nonstandard iterates of the universe. Indeed, given $\alpha\in{}^{k*}S$, we can define $\mathscr{U}_\alpha:=\{A\subseteq S\ :\ \alpha\in{}^{k*}A\}$.

Exercise 4.7. For $\alpha\in{}^{k*}S$, prove that $\mathscr{U}_\alpha=\mathscr{U}_{{}^*\alpha}$.

For $\alpha,\beta\in\bigcup_k{}^{k*}S$, we define $\alpha\sim\beta$ if and only if $\mathscr{U}_\alpha=\mathscr{U}_\beta$ remains true. Note that α and β may live in different levels of the iterated nonstandard extensions.

Exercise 4.8. Prove that, for $\alpha_0,\ldots,\alpha_k\in{}^*\mathbb{N}$ and $a_0,\ldots,a_k\in\mathbb{N}$, one has

$$a_0\mathscr{U}_{\alpha_0}\oplus\cdots\oplus a_k\mathscr{U}_{\alpha_k}=\mathscr{U}_{a_0\alpha_0+a_1{}^*\alpha_1+\cdots+a_k{}^{k*}\alpha_k}.$$

Exercise 4.9.

1. Suppose that $\alpha,\alpha',\beta,\beta'\in{}^*\mathbb{N}$ are such that $\alpha\sim\alpha'$ and $\beta\sim\beta'$. Prove that $\alpha+{}^*\beta\sim\alpha'+{}^*\beta'$.
2. Find $\alpha,\alpha',\beta,\beta'$ as above with $\alpha+\beta\not\sim\alpha'+\beta'$.

4.3 The iterated ultrapower perspective

The ultrapower model does not naturally accommodate iterations of hyper-extensions, and in fact, one can be easily puzzled when thinking of iterated hyper-extensions in terms of "iterated ultrapowers". Let us try to clarify this point.

[2] To avoid ambiguity, some authors call the hyper-extensions ${}^*A\in\mathbb{V}$ "internal-standard".

Let us fix an ultrafilter $\mathscr{U}$ on $\mathbb{N}$. Since one can take the ultrapower $\mathbb{N}^{\mathbb{N}}/\mathscr{U}$ of $\mathbb{N}$ to get a nonstandard extension of $\mathbb{N}$, it is natural to take an ultrapower $(\mathbb{N}^{\mathbb{N}}/\mathscr{U})^{\mathbb{N}}/\mathscr{U}$ of $\mathbb{N}^{\mathbb{N}}/\mathscr{U}$ to get a further nonstandard extension. The diagonal embedding $d : \mathbb{N}^{\mathbb{N}}/\mathscr{U} \to (\mathbb{N}^{\mathbb{N}}/\mathscr{U})^{\mathbb{N}}/\mathscr{U}$ is the map where $d(\alpha)$ is the equivalence class in $(\mathbb{N}^{\mathbb{N}}/\mathscr{U})^{\mathbb{N}}/\mathscr{U}$ of the sequence that is constantly α. We define ${}^*\alpha$ as $d(\alpha)$, but, unlike the first time when we took an ultrapower and identified $n \in \mathbb{N}$ with $d(n)$, let us refrain from identifying α with ${}^*\alpha$. Indeed, recall that, according to the theory developed in the first section of this chapter, ${}^*\alpha$ is supposed to be infinitely larger than α. How do we reconcile this fact with the current construction? Well, unlike the first time we took an ultrapower, a new phenomenon has occurred. Indeed, we now have a second embedding $d_0^{\mathscr{U}} : \mathbb{N}^{\mathbb{N}}/\mathscr{U} \to (\mathbb{N}^{\mathbb{N}}/\mathscr{U})^{\mathbb{N}}/\mathscr{U}$ given by taking the ultrapower of the diagonal embedding $d_0 : \mathbb{N} \to \mathbb{N}^{\mathbb{N}}/\mathscr{U}$.[3] Precisely, if $\alpha = [\sigma] \in \mathbb{N}^{\mathbb{N}}/\mathscr{U}$ where $\sigma : \mathbb{N} \to \mathbb{N}$, then $d_0^{\mathscr{U}}(\alpha) = [([c_{\sigma(1)}],[c_{\sigma(2)}],[c_{\sigma(3)}],\ldots)]$. It is thus through this embedding that we identify $\alpha \in \mathbb{N}^{\mathbb{N}}/\mathscr{U}$ with its image $d_0^{\mathscr{U}}(\alpha) \in (\mathbb{N}^{\mathbb{N}}/\mathscr{U})^{\mathbb{N}}/\mathscr{U}$.

It is now straightforward to see that $\alpha < d(\alpha)$ for all $\alpha \in \mathbb{N}^{\mathbb{N}}/\mathscr{U} \setminus \mathbb{N}$. For example, if $\alpha = [(1,2,3,\ldots)] \in \mathbb{N}^{\mathbb{N}}/\mathscr{U}$, then we identify α with $[([c_1],[c_2],[c_3],\ldots)] \in (\mathbb{N}^{\mathbb{N}}/\mathscr{U})^{\mathbb{N}}/\mathscr{U}$. Since $[c_n] < \alpha$ for all n, we have that $\alpha < [(\alpha,\alpha,\alpha,\ldots)] = d(\alpha) = {}^*\alpha$.

Also, it is also straightforward to see that defining ${}^{**}f$ as $(f^{\mathscr{U}})^{\mathscr{U}}$ extends ${}^*f = f^{\mathscr{U}}$ for any function $f : \mathbb{N} \to \mathbb{N}$. Indeed, if $\alpha = [\sigma] \in \mathbb{N}^{\mathbb{N}}/\mathscr{U}$, then we have that

$$(f^{\mathscr{U}})^{\mathscr{U}}(\alpha) = (f^{\mathscr{U}})^{\mathscr{U}}(d_0^{\mathscr{U}}(\alpha)) = [(f^{\mathscr{U}}([c_{\sigma(1)}]), f^{\mathscr{U}}([c_{\sigma(2)}]),\ldots)] = [([c_{f(\sigma(1))}],[c_{f(\sigma(2))}],\ldots)] = d_0^{\mathscr{U}}([f\circ\sigma]) = [f\circ\sigma] = f^{\mathscr{U}}(\alpha).$$

[3] Every map $f : A \to B$ yields a natural map $f^{\mathscr{U}} : A^{\mathbb{N}}/\mathscr{U} \to B^{\mathbb{N}}/\mathscr{U}$ between their ultrapowers, by setting $f^{\mathscr{U}}([\sigma]) = [f\circ\sigma]$ for every $\sigma : \mathbb{N} \to A$.

Chapter 5
Idempotents

5.1 The existence of idempotents in semitopological semigroups

Definition 5.1. Suppose that $(S,\cdot)$ is a semigroup. We say that $e \in S$ is *idempotent* if $e \cdot e = e$.

The following classical theorem of Ellis is the key to much of what we do.

Theorem 5.2. *Suppose that $(S,\cdot)$ is a compact semitopological semigroup. Then S has an idempotent element.*

Proof. Let $\mathscr{S}$ denote the set of nonempty closed subsemigroups of S. It is clear that the intersection of any descending chain of elements of $\mathscr{S}$ is also an element of $\mathscr{S}$, whence by Zorn's lemma, we may find $T \in \mathscr{S}$ that is minimal.

Fix $s \in T$; we show that s is idempotent. Set $T_1 := Ts$. Note that $T_1 \neq \emptyset$ as $T \neq \emptyset$. Since S is a semitopological semigroup and T is compact, we have that T_1 is also compact. Finally, note that T_1 is also a subsemigroup of S:

$$T_1 \cdot T_1 = (Ts)(Ts) \subseteq T \cdot T \cdot T \cdot s \subseteq T \cdot s = T_1.$$

We thus have that $T_1 \in \mathscr{S}$. Since $s \in T$, we have that $T_1 \subseteq T$, whence by minimality of T, we have that $T_1 = T$. In particular, the set $T_2 := \{t \in T \ : \ t \cdot s = s\}$ is not empty. Note that T_2 is also a closed subset of T, whence compact. Once again, we note that T_2 is a subsemigroup of S. Indeed, if $t,t' \in T_2$, then $tt' \in T$ and $(tt') \cdot s = t \cdot (t' \cdot s) = t \cdot s = s$. We thus have that $T_2 \in \mathscr{S}$. By minimality of T, we have that $T_2 = T$. It follows that $s \in T_2$, that is, $s \cdot s = s$.

The previous theorem and Theorem 1.19 immediately give the following:

Corollary 5.3. *Let $(S,\cdot)$ be a semigroup and let T be any nonempty closed subsemigroup of $(\beta S,\odot)$. Then T contains an idempotent element.*

We refer to idempotent elements of βS as *idempotent ultrafilters*. Thus, the previous corollary says that any nonempty closed subsemigroup of βS contains an idempotent ultrafilter.

Given the correspondence between ultrafilters on S and elements of ${}^*S$, it is natural to translate the notion of idempotent ultrafilter to the setting of ${}^*S$. Suppose that $\alpha \in {}^*S$ is such that $\mathscr{U}_\alpha$ is an idempotent ultrafilter on S. We thus have that $\mathscr{U}_\alpha = \mathscr{U}_\alpha \odot \mathscr{U}_\alpha = \mathscr{U}_{\alpha \cdot {}^*\alpha}$. This motivates the following:

Definition 5.4. $\alpha \in {}^*S$ is *u-idempotent* if $\alpha \cdot {}^*\alpha \sim \alpha$.

We thus see that $\alpha \in {}^*S$ is u-idempotent if and only if $\mathscr{U}_\alpha$ is an idempotent ultrafilter on S. The following exercise gives a nonstandard proof of [6, Theorem 2.10].

Exercise 5.5.

1. Suppose that $\alpha \in {}^*\mathbb{N}$ is idempotent. Prove that $2\alpha + {}^{**}\alpha$, $2\alpha + {}^*\alpha + {}^{**}\alpha$, and $2\alpha + 2{}^*\alpha + {}^{**}\alpha$ all generate the same ultrafilter, namely $2\mathscr{U}_\alpha \oplus \mathscr{U}_\alpha$.

2. Suppose that $\mathscr{U} \in \beta\mathbb{N}$ is idempotent and $A \in 2\mathscr{U} \oplus \mathscr{U}$. Prove that A contains a 3-termed arithmetic progression.

We now seek an analog of the above fact that nonempty closed subsemigroups of βS contain idempotents. Suppose that $T \subseteq \beta S$ is a subsemigroup and that $\alpha, \beta \in {}^*S$ are such that $\mathscr{U}_\alpha, \mathscr{U}_\beta \in T$. Since $\mathscr{U}_{\alpha \cdot {}^*\beta} = \mathscr{U}_\alpha \odot \mathscr{U}_\beta \in T$, we are led to the following definition:

Definition 5.6. $T \subseteq {}^*S$ is a *u-subsemigroup* if, for any $\alpha, \beta \in T$, there is $\gamma \in T$ such that $\alpha \cdot {}^*\beta \sim \gamma$.

We thus have the following:

Corollary 5.7. *Suppose that $T \subseteq {}^*S$ is a nonempty closed u-subsemigroup. Then T contains a u-idempotent element.*

5.2 Partial semigroups

We will encounter the need to apply the above ideas to the broader context of partial semigroups.

Definition 5.8. A *partial semigroup* is a set S endowed with a partially defined binary operation $(s,t) \mapsto s \cdot t$ that satisfies the following form of the associative law: given $s_1, s_2, s_3 \in S$, if either of the products $(s_1 \cdot s_2) \cdot s_3$ or $s_1 \cdot (s_2 \cdot s_3)$ are defined, then so is the other and the products are equal. The partial semigroup $(S, \cdot)$ is *directed* if, for any finite subset F of S, there exists $t \in S$ such that the product $s \cdot t$ is defined for every $s \in F$.

For the rest of this chapter, we assume that $(S, \cdot)$ is a directed partial semigroup.

Definition 5.9. We call $\mathscr{U} \in \beta S$ *cofinite* if, for all $s \in S$, we have $\{t \in S \,:\, s \cdot t \text{ is defined }\} \in \mathscr{U}$. We let γS denote the set of all cofinite elements of βS.

Exercise 5.10. γS is a *nonempty* closed subset of βS.

We can define an operation $\odot$ on γS by declaring, for $\mathscr{U}, \mathscr{V} \in \gamma S$ and $A \subseteq S$, that $A \subseteq \mathscr{U} \odot \mathscr{V}$ if and only if

$$\{s \in S \,:\, \{t \in S \,:\, s \cdot t \text{ is defined and } s \cdot t \in A\} \in \mathscr{V}\} \in \mathscr{U}.$$

Note that the operation $\odot$ is a totally defined operation on γS even though the original operation $\cdot$ was only a partially defined operation.

The next fact is very important but is somewhat routine given everything that has been proven thus far. We thus leave the proof as a (lengthy) exercise; one can also consult [45, pages 31 and 32].

Theorem 5.11. $(\gamma S, \odot)$ *is a compact semitopological semigroup. Consequently, every nonempty closed subsemigroup of γS contains an idempotent element.*

We once again give the nonstandard perspective on the preceding discussion. Note that ${}^*S$ is naturally a partial semigroup with the nonstandard extension of the partial seimgroup operation. We say that $\alpha \in {}^*S$ is *cofinite* if $s \cdot \alpha$ is defined for every $s \in S$. We leave it to the reader to check that α is cofinite if and only if $\mathscr{U}_\alpha$ is a cofinite element of βS. Consequently, Theorem 5.11 implies that any nonempty closed u-subsemigroup of the set of cofinite elements of ${}^*S$ contains an idempotent element.

Exercise 5.12. Without using Theorem 5.11, prove that, for any cofinite $\alpha, \beta \in {}^*S$, there is cofinite $\gamma \in {}^*S$ such that $\alpha \cdot {}^*\beta \sim \gamma$. Compare your proof to the proof that $\mathscr{U} \odot \mathscr{V} \in \gamma S$ whenever $\mathscr{U}, \mathscr{V} \in \gamma S$.

Chapter 6
Loeb measure

6.1 Premeasures and measures

Fix a set X. A nonempty set $\mathscr{A} \subseteq \mathscr{P}(X)$ is an *algebra* if it is closed under unions, intersections, and complements, that is, if $A, B \in \mathscr{A}$, then $A \cup B$, $A \cap B$, and $X \setminus A$ all belong to $\mathscr{A}$. If $\mathscr{A}$ is an algebra of subsets of X, then $\emptyset, X \in \mathscr{A}$. An algebra $\mathscr{A}$ on X is said to be a σ-*algebra* if it is also closed under countable unions, that is, if $A_1, A_2, \ldots$ all belong to $\mathscr{A}$, then so does $\bigcup_{n=1}^{\infty} A_n$. A σ-algebra is then automatically closed under countable intersections.

Exercise 6.1. Suppose that X is a set and $\mathscr{O} \subseteq \mathscr{P}(X)$ is an arbitrary collection of subsets of X. Prove that there is a smallest σ-algebra Ω containing $\mathscr{O}$. We call this σ-algebra the *σ-algebra generated by $\mathscr{O}$* and denote it by $\sigma(\mathscr{O})$.

Remark 6.2. When trying to prove that every element of $\sigma(\mathscr{O})$ has a certain property, one just needs to show that the set of elements having that property contains $\mathscr{O}$ and is a σ-algebra.

Suppose that $\mathscr{A}$ is an algebra on X. A *pre-measure* on $\mathscr{A}$ is a function $\mu : \mathscr{A} \to [0, +\infty]$ satisfying the following two axioms:

- $\mu(\emptyset) = 0$;
- (Countable Additivity) If $A_1, A_2, \ldots,$ all belong to $\mathscr{A}$, are pairwise disjoint, *and* $\bigcup_{n=1}^{\infty} A_n$ belongs to Ω, then $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.

If $\mathscr{A}$ is a σ-algebra, then a pre-measure is called a *measure*. If μ is a measure on X and $\mu(X) = 1$, then we call μ a *probability measure on X*.

Exercise 6.3. Fix $n \in \mathbb{N}$ and suppose that $X = \{1, 2, \ldots, n\}$. Let $\mathscr{A} := \mathscr{P}(X)$. Then $\mathscr{A}$ is an algebra of subsets of X that is actually a σ-algebra for trivial reasons. Define the function $\mu : \mathscr{A} \to [0, 1]$ by $\mu(A) = \frac{|A|}{n}$. Then μ is a probability measure on $\mathscr{A}$, called the *normalized counting measure*.

Exercise 6.4. Suppose that $\mu : \mathscr{A} \to [0, +\infty]$ is a pre-measure. Prove that $\mu(A) \leq \mu(B)$ for all $A, B \in \mathscr{A}$ with $A \subseteq B$.

For subsets A, B of X, we define the *symmetric difference of A and B* to be $A \triangle B := (A \setminus B) \cup (B \setminus A)$.

Exercise 6.5. Suppose that $\mathscr{A}$ is an algebra and $\mu : \sigma(\mathscr{A}) \to [0, \infty]$ is a measure. Prove that, for every $A \in \sigma(\mathscr{A})$ with $\mu(A) < \infty$ and every $\varepsilon \in \mathbb{R}^{>0}$, there is $B \in \mathscr{A}$ such that $\mu(A \triangle B) < \varepsilon$.

For our purposes, it will be of vital importance to know that a pre-measure μ on an algebra $\mathscr{A}$ can be extended to a measure on a σ-algebra $\sigma(\mathscr{A})'$ extending $\mathscr{A}$, a process which is known as *Carathéodory extension*. We briefly outline how this is done. The interested reader can consult any good book on measure theory for all the glorious details; see for instance [44, Section 1.7].

Fix an algebra $\mathscr{A}$ of subsets of X and a pre-measure μ on $\mathscr{A}$. For arbitrary $A \subseteq X$, we define the *outer measure of A* to be

$$\mu^+(A) := \inf\{\sum_{n\in\mathbb{N}} \mu(B_n) \mid A \subseteq \bigcup_{n\in\mathbb{N}} B_n, \text{ each } B_n \in \mathscr{A}\}.$$

Note that $\mu^+(A) = \mu(A)$ for all $A \in \mathscr{A}$. Now although μ^+ is defined on all of $\mathscr{P}(X)$ (which is certainly a σ-algebra), it need not be a measure. However, there is a canonical σ-sub-algebra $\mathscr{A}_m$ of $\mathscr{P}(X)$, the so-called *Carathéodory measurable* or *μ^+-measurable subsets of X*, on which μ^+ is a measure. These are the sets $A \subseteq X$ such that

$$\mu^+(E) = \mu^+(A \cap E) + \mu^+(E \setminus A)$$

for every other set $E \subset X$. Let us collect the relevant facts here:

Fact 6.6 *Let X be a set, $\mathscr{A}$ an algebra of subsets of X, and $\mu : \mathscr{A} \to [0,\infty]$ a pre-measure on $\mathscr{A}$ with associated outer measure μ^+ and σ-algebra of μ^+-measurable sets $\mathscr{A}_m$. Further suppose that μ is* σ-finite*, meaning that we can write $X = \bigcup_{n\in\mathbb{N}} X_n$ with each $X_n \in \mathscr{A}$ and $\mu(X_n) < \infty$.*

1. *$\sigma(\mathscr{A}) \subseteq \mathscr{A}_m$ and $\mu^+|\mathscr{A} = \mu$.*
2. *(Uniqueness) If $\mathscr{A}'$ is another σ-algebra on X extending $\mathscr{A}$ and $\mu' : \mathscr{A}' \to [0,\infty]$ is a measure on $\mathscr{A}'$ extending μ, then μ^+ and μ' agree on $\mathscr{A}_m \cap \mathscr{A}'$ (and, in particular, on $\sigma(\mathscr{A})$).*
3. *(Completeness) If $A \subseteq B \subseteq X$ are such that $B \in \mathscr{A}_m$ and $\mu^+(B) = 0$, then $A \in \mathscr{A}_m$ and $\mu^+(A) = 0$.*
4. *(Approximation Results)*

   a. *If $A \in \mathscr{A}_m$, then there is $B \in \sigma(\mathscr{A})$ containing A such that $\mu^+(B \setminus A) = 0$. (So $\mathscr{A}_m$ is the* completion *of $\sigma(\mathscr{A})$.)*
   b. *If $A \in \mathscr{A}_m$ is such that $\mu^+(A) < \infty$, then for every $\varepsilon \in \mathbb{R}^{>0}$, there is $B \in \mathscr{A}$ such that $\mu(A \triangle B) < \varepsilon$.*
   c. *Suppose that $A \subseteq X$ is such that, for every $\varepsilon \in \mathbb{R}^{>0}$, there is $B \in \mathscr{A}$ such that $\mu(A \triangle B) < \varepsilon$. Then $A \in \mathscr{A}_m$.*

Example 6.7 (Lebesgue measure). Suppose that $X = \mathbb{R}$ and $\mathscr{A}$ is the collection of *elementary sets*, namely the finite unions of intervals. Define $\mu : \mathscr{A} \to [0,\infty]$ by declaring $\mu(I) = \text{length}(I)$ and $\mu(I_1 \cup \cdots \cup I_n) = \sum_{i=1}^n \mu(I_j)$ whenever $I_1, \ldots, I_n$ are pairwise disjoint. The above outer-measure procedure yields the σ-algebra $\mathscr{A}_m$, which is known as the σ-algebra of *Lebesgue measurable subsets of* $\mathbb{R}$ and usually denoted by $\mathfrak{M}$. The measure μ^+ is often denoted by λ and is referred to as *Lebesgue measure*. The σ-algebra $\sigma(\mathscr{A})$ in this case is known as the σ-algebra of *Borel subsets of* $\mathbb{R}$, usually denoted by $\mathscr{B}$. It can also be seen to be the σ-algebra generated by the open intervals.

6.2 The definition of Loeb measure

How do we obtain pre-measures in the nonstandard context? Well, we obtain them by looking at normalized counting measures on hyperfinite sets. Suppose that X is a hyperfinite set. We set $\mathscr{A}$ to be the set of *internal* subsets of X. Then $\mathscr{A}$ is an algebra of subsets of X that is not (in general) a σ-algebra. For example, if $X = [1,N] \subseteq {}^*\mathbb{N}$ for some $N \in {}^*\mathbb{N} \setminus \mathbb{N}$, then for each $n \in \mathbb{N}$, $A_n := \{n\}$ belongs to $\mathscr{A}$, but $\bigcup_n A_n = \mathbb{N}$ does not belong to $\mathscr{A}$ as $\mathbb{N}$ is not internal.

If $A \in \mathscr{A}$, then A is also hyperfinite. We thus define a function $\mu : \mathscr{A} \to [0,1]$ by $\mu(A) := \text{st}\left(\frac{|A|}{|X|}\right)$. We claim that μ_X is a pre-measure. It is easily seen to be *finitely additive*, that is, $\mu(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^n \mu(A_i)$ whenever $A_1, \ldots, A_n \in \mathscr{A}$ are disjoint. But how do we verify countable additivity?

Exercise 6.8. If $A_1, A_2, \ldots$ all belong to $\mathscr{A}$ and $\bigcup_{n=1}^\infty A_n$ also belongs to $\mathscr{A}$, then there is $k \in \mathbb{N}$ such that $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^k A_n$.

Thus, by the exercise, countable additivity is a trivial consequence of finite additivity in this context. We may thus apply the Carathéodory extension theorem from the previous section to obtain a probability measure $\mu^+ : \mathscr{A} \to [0,1]$ extending μ. The measure μ^+ is called the *Loeb measure on X* and will be denoted μ_X. The elements of $\mathscr{A}$ are referred to as the *Loeb measurable subsets of X* and will be denoted by $\mathscr{L}_X$.

Lemma 6.9. *If $B \in \mathscr{L}_X$, then*

$$\mu_X(B) = \inf\{\mu_X(A) \mid A \textit{ is internal and } B \subseteq A\}.$$

Proof. The inequality $\leq$ is clear. Towards the other inequality, fix $\varepsilon \in \mathbb{R}^{>0}$; we need to find internal A such that $B \subseteq A$ and $\mu_X(A) \leq \mu_X(B) + \varepsilon$. Fix an increasing sequence of internal sets $(A_n \mid n \in \mathbb{N})$ such that $B \subseteq \bigcup_{n\in\mathbb{N}} A_n$ and $\mu_X(A_n) < \mu_X(B) + \varepsilon$ for every $n \in \mathbb{N}$. By countable saturation, we extend this sequence to an internal sequence $(A_n \mid n \in {}^*\mathbb{N})$. By transfer, for each $k \in \mathbb{N}$, we have

$$(\forall n \in {}^*\mathbb{N})(n \leq k \to (A_n \subseteq A_k \text{ and } \mu_X(A_n) < \mu_X(B) + \varepsilon)).$$

By , there is $K > \mathbb{N}$ such that $\mu_X(A_K) \leq \mu_X(B) + \varepsilon$. This concludes the proof.

Lemma 6.10. *If $B \in \mathscr{L}_X$, then, for every $\varepsilon \in \mathbb{R}^{>0}$, there are internal subsets C, A of X such that $C \subseteq B \subseteq A$ and $\mu_X(A \setminus C) < \varepsilon$.*

Proof. Fix $\varepsilon > 0$. By Lemma 6.9 applied to B, there is an internal set A containing B such that $\mu_X(A) < \mu_X(B) + \frac{\varepsilon}{2}$. By Lemma 6.9 applied to $A \setminus B$, there is an internal set R containing $A \setminus B$ such that $\mu_X(R) < \mu_X(A \setminus B) + \frac{\varepsilon}{2} < \varepsilon$. Set now $C := A \setminus R$ and observe that C is an internal set contained in B. Furthermore we have that $\mu_X(A \setminus C) \leq \mu_X(R) < \varepsilon$. This concludes the proof.

There are many interesting things to say about Loeb measure. It is crucial for applications of nonstandard analysis to many different areas of mathematics. More information on the Loeb measure can be found in [1, 2]. We will see later in this book that Loeb measure allows us to treat densities on the natural numbers as measures, allowing us to bring in tools from measure theory and ergodic theory into combinatorial number theory.

6.3 Lebesgue measure via Loeb measure

The purpose of this section is to see that Lebesgue measure can be constructed using a suitable Loeb measure. The connection between these measures serves as a useful motivation for the results of Chapter 13 on sumsets of sets of positive density.

Theorem 6.11. *Suppose that $N > \mathbb{N}$ and consider the hyperfinite set $X := \{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N}{N} = 1\}$ and the function* $\mathrm{st}: X \to [0,1]$. *Define a σ-algebra $\mathscr{A}$ on $[0,1]$ by $A \in \mathscr{A}$ if and only if $\mathrm{st}^{-1}(A) \in \mathscr{L}_X$. For $A \in \mathscr{A}$, define $\nu(A) := \mu_X(\mathrm{st}^{-1}(A))$. Then $\mathscr{A}$ is the algebra of Lebesgue measurable subsets of $[0,1]$ and ν is Lebesgue measure.*

We outline the proof of this theorem in a series of steps. We denote by $\mathscr{B}$ the σ-algebra of Borel subsets of $[0,1]$, by $\mathscr{M}$ the σ-algebra of measurable subsets of $[0,1]$, and by λ the Lebesgue measure on $\mathscr{M}$.

Exercise 6.12. Prove that $\mathscr{A}$ is a σ-algebra and ν is a measure on $\mathscr{A}$.

Exercise 6.13. Fix $a, b \in [0,1]$ with $a < b$.

1. Prove that $X \cap (a,b)^* \in \mathscr{L}_X$ and $\mu_X(X \cap (a,b)^*) = b - a$.
2. Prove that $\mathrm{st}^{-1}((a,b)) = \bigcup_{n\in\mathbb{N}}(X \cap (a + \frac{1}{n}, b - \frac{1}{n})^*)$.
3. Prove that $(a,b) \in \mathscr{A}$ and $\nu((a,b)) = b - a$.

We now use the fact that λ is the only probability measure on $\mathscr{B}$ satisfying $\lambda(a,b) = b - a$ and that is invariant under translations modulo 1 to conclude that $\mathscr{B} \subseteq \mathscr{A}$ and $\nu|_{\mathscr{B}} = \lambda|_{\mathscr{B}}$.

Exercise 6.14. Conclude that $\mathscr{M} \subseteq \mathscr{A}$ and $\nu|_{\mathfrak{M}} = \lambda|_{\mathfrak{M}}$. (Hint: Use Fact 6.6.)

Exercise 6.15. Show that $\mathscr{A} \subseteq \mathscr{M}$. (Hint: if $B \in \mathscr{A}$, then by Lemma 6.10, there are internal $C, D \subseteq X$ such that $C \subseteq \mathrm{st}^{-1}(B) \subseteq D$ and $\mu_X(D \setminus C) < \varepsilon$. Set $C' := \mathrm{st}(C)$ and $D' := [0,1] \setminus \mathrm{st}(X \setminus D)$. Notice that C' is closed and D' is open, whence $C', D' \in \mathscr{B} \subseteq \mathscr{A}$. Prove that $C \subseteq \mathrm{st}^{-1}(C')$ and $\mathrm{st}^{-1}(D') \subseteq D$. Conclude that $B \in \mathscr{M}$.)

6.4 Integration

There is a lot to say about the nonstandard theory of integration. We will focus on the Loeb measure μ_X obtained from a hyperfinite set X. In this section, X always denotes a hyperfinite set.

First, if $F : X \to {}^*\mathbb{R}$ is an internal function such that $F(x)$ is finite for μ_X-almost every $x \in X$, we define $\operatorname{st}(F) : X \to \mathbb{R}$ by $\operatorname{st}(F)(x) := \operatorname{st}(F(x))$ whenever $F(x)$ is finite. (Technically speaking, $\operatorname{st}(F)$ is only defined on a set of measure 1, but we will ignore this minor point.) If $f : X \to \mathbb{R}$ is a function and $F : X \to {}^*\mathbb{R}$ is an internal function such that $f(x) = \operatorname{st}(F)(x)$ for μ_X-almost every $x \in X$, we call F a *lift* of f. We first characterize which functions have lifts.

Proposition 6.16. *$f : X \to \mathbb{R}$ has a lift if and only if f is μ_X-measurable.*

Proof. If F is a lift of f, then for any $r \in \mathbb{R}$, we have

$$\mu_X\left(\{x \in X \ : \ f(x) < r\} \triangle \bigcup_{n\in\mathbb{N}} \left\{x \in X \ : \ F(x) < r - \frac{1}{n}\right\}\right) = 0.$$

Since the latter set is clearly measurable and μ_X is a complete measure, it follows that $\{x \in X \ : \ f(x) < r\}$ is measurable, whence f is μ_X-measurable.

For the converse, suppose that f is μ_X-measurable and fix a countable open basis $\{V_n\}$ for $\mathbb{R}$. For $n \in \mathbb{N}$, set $U_n := f^{-1}(V_n) \in \mathscr{L}_X$. By Lemma 6.10, one can find, for every $n \in \mathbb{N}$, an increasing sequence $(A_{n,m})$ of internal subsets of U_n such that $\mu_X(A_{n,m}) \geq \mu_X(U_n) - 2^{-m}$ for every $m \in \mathbb{N}$. It follows that the subset

$$X_0 := X \setminus \bigcup_{n\in\mathbb{N}} \left(U_n \setminus \bigcup_{m\in\mathbb{N}} A_{n,m}\right)$$

of X has μ_X-measure 1. Observe now that, for every $n, m \in \mathbb{N}$, there exists an internal function $F : X \to {}^*\mathbb{R}$ such that $F\left(A_{\ell,k}\right) \subset {}^*V_\ell$ for $k \leq m$ and $\ell \leq n$. Therefore, by saturation, there exists an internal function $F : X \to {}^*\mathbb{R}$ such that $F(A_{n,m}) \subset {}^*V_n$ for every $n, m \in \mathbb{N}$. It is clear that $f(x) = \operatorname{st}(F(x))$ for every $x \in X_0$, whence F is a lift of f.

The rest of this section is devoted towards understanding $\int f d\mu_X$ (in the case that f is μ_X-integrable) and the "internal integral" $\frac{1}{|X|}\sum_{x\in X} F(x)$ of a lift F of f. We first treat a special, but important, case.

Lemma 6.17. *Suppose that $F : X \to {}^*\mathbb{R}$ is an internal function such that $F(x)$ is finite for all $x \in X$. Then $\operatorname{st}(F)$ is μ_X-integrable and*

$$\int \operatorname{st}(F) d\mu_X = \operatorname{st}\left(\frac{1}{|X|}\sum_{x\in X} F(x)\right).$$

Proof. Note first that the assumptions imply that there is $m \in \mathbb{N}$ such that $|F(x)| \leq m$ for all $x \in X$. It follows that $\operatorname{st}(F)$ is μ_X-integrable. Towards establishing the displayed equality, note that, by considering positive and negative parts, that we may assume that F is nonnegative. Fix $n \in \mathbb{N}$. For $k \in \{0, 1, \ldots, mn-1\}$, set $A_k := \{x \in X \ : \ \frac{k}{n} \leq F(x) < \frac{k+1}{n}\}$, an internal set. Since $\sum_k \frac{k}{n}\chi_{A_k}$ is a simple function below $\operatorname{st}(F)$, we have that $\sum_k \frac{k}{n}\mu_X(A_k) \leq \int \operatorname{st}(F) d\mu_X$. However, we also have

$$\sum_k \frac{k}{n}\mu_X(A_k) = \operatorname{st}\left(\frac{1}{|X|}\sum_k \sum_{x\in A_k} \frac{k}{n}\right) \geq \operatorname{st}\left(\frac{1}{|X|}\sum_k \sum_{x\in A_k} (F(x) - \frac{1}{n})\right) = \operatorname{st}\left(\frac{1}{|X|}\sum_{x\in X} F(x)\right) - \frac{1}{n}.$$

It follows that $\operatorname{st}(\frac{1}{|X|}\sum_{x\in X} F(x)) \leq \int \operatorname{st}(F) d\mu_X + \frac{1}{n}$; since n was arbitrary, we have that $\operatorname{st}(\frac{1}{|X|}\sum_{x\in X} F(x)) \leq \int \operatorname{st}(F) d\mu_X$.

We leave the proof of the inequality $\int \operatorname{st}(F) d\mu_X \leq \operatorname{st}(\frac{1}{|X|}\sum_{x\in X} F(x))$ to the reader.

We now seek to extend the previous lemma to cover situations when F is not necessarily bounded by a standard number. Towards this end, we need to introduce the appropriate nonstandard integrability assumption. A μ_X-measurable internal function $F : X \to {}^*\mathbb{R}$ is called *S-integrable* if:

1. The quantity
$$\frac{1}{|X|}\sum_{x\in X}|F(x)|$$
is finite, and
2. for every internal subset A of X with $\mu_X(A)=0$, we have
$$\frac{1}{|X|}\sum_{x\in A}|F(x)|\approx 0.$$

Here is the main result of this section:

Theorem 6.18. *Suppose that $f : X \to \mathbb{R}$ is a μ_X-measurable function. Then f is μ_X-integrable if and only if f has an S-integrable lifting. In this case, for any S-integrable lift F of f and any internal subset B of X, we have*
$$\int_B f d\mu_X = \mathrm{st}\left(\frac{1}{|X|}\sum_{x\in B}F(x)\right).$$

Proof. We first note that, by taking positive and negative parts, we may assume that f is nonnegative. Moreover, by replacing f with $f\cdot\chi_B$, we may assume that $B = X$.

We first suppose that $F : X \to {}^*\mathbb{R}$ is a nonnegative S-integrable function such that $F(x)$ is finite for μ_X-almost every x. For $n \in {}^*\mathbb{N}$, set $B_n := \{x \in X\ :\ F(x) \geq n\}$.

Claim 1: For every infinite $N \in {}^*\mathbb{N}$, we have
$$\frac{1}{|X|}\sum_{x\in B_N}F(x)\approx 0.$$

Proof of Claim 1: Observe that
$$\frac{N|B_N|}{|X|}\leq\frac{1}{|X|}\sum_{x\in B_N}F(x)\leq\frac{1}{|X|}\sum_{x\in X}F(x)$$
Therefore
$$\frac{|B_N|}{|X|}\leq\frac{1}{N}\frac{1}{|X|}\sum_{x\in X}F(x)\approx 0$$
since, by assumption, $\frac{1}{|X|}\sum_{x\in X}F(x)$ is finite. It follows from the assumption that F is S-integrable that
$$\frac{1}{|X|}\sum_{x\in B_N}F(x)\approx 0.$$

In the rest of the proof, we will use the following notation: given a nonnegative internal function $F : X \to {}^*\mathbb{R}$ and $m \in {}^*\mathbb{N}$, we define the internal function $F_m : X \to {}^*\mathbb{R}$ by $F_m(x) = \min\{F(x), m\}$. Observe that $F_m(x) \leq F_{m+1}(x) \leq F(x)$ for every $m \in {}^*\mathbb{N}$ and every $x \in X$. It follows from the Monotone Convergence Theorem and the fact that, for μ_X-almost every $x \in X$, the sequence $(\mathrm{st}(F_m(x))\ :\ m \in \mathbb{N})$ converges to $\mathrm{st}(F(x))$, that $\int \mathrm{st}(F_m)d\mu_X \to \int \mathrm{st}(F)d\mu_X$.

Claim 2: We have
$$\mathrm{st}\left(\frac{1}{|X|}\sum_{x\in X}F(x)\right)=\lim_{m\to+\infty}\mathrm{st}\left(\frac{1}{|X|}\sum_{x\in X}F_m(x)\right).$$

Proof of Claim 2: It is clear that
$$\lim_{m\to\infty}\mathrm{st}\left(\frac{1}{|X|}\sum_{x\in X}F_m(x)\right)\leq\mathrm{st}\left(\frac{1}{|X|}\sum_{x\in X}F(x)\right).$$
For the other inequality, fix $M \in {}^*\mathbb{N}$ infinite and observe that

$$\begin{aligned}\frac{1}{|X|}\sum_{x\in X}F(x) &= \frac{1}{|X|}\sum_{x\in B_M}F(x)+\frac{1}{|X|}\sum_{x\in X\setminus B_M}F(x)\\ &\approx \frac{1}{|X|}\sum_{x\in X\setminus B_M}F(x)\\ &= \frac{1}{|X|}\sum_{x\in X\setminus B_M}F_M(x)\\ &\leq \frac{1}{|X|}\sum_{x\in X}F_M(x).\end{aligned}$$

Thus, given any $\varepsilon>0$, we have that $\frac{1}{|X|}\sum_{x\in X}F(x)\leq\frac{1}{|X|}\sum_{x\in X}F_M(x)+\varepsilon$ for all infinite M, whence, by underflow, we have that $\frac{1}{|X|}\sum_{x\in X}F(x)\leq\frac{1}{|X|}\sum_{x\in X}F_m(x)+\varepsilon$ for all but finitely many $m\in\mathbb{N}$. It follows that $\mathrm{st}\left(\frac{1}{|X|}\sum_{x\in X}F(x)\right)\leq\lim_{m\to+\infty}\mathrm{st}\left(\frac{1}{|X|}\sum_{x\in X}F_m(x)\right)$, as desired.

By Lemma 6.17, Claim 2, and the discussion preceding Claim 2, we have that $\mathrm{st}(F)$ is μ_X-integrable and $\int\mathrm{st}(F)d\mu=\mathrm{st}\left(\frac{1}{|X|}\sum_{x\in X}F(x)\right)$, as desired.

We now suppose that f is a nonnegative μ_X-integrable function. We must show that f has an S-integrable lifting. Let F be any nonnegative lifting of f. Note that, for every infinite $M\in{}^*\mathbb{N}$, that F_M is also a lifting of f. We will find an infinite $M\in{}^*\mathbb{N}$ such that F_M is also S-integrable.

By the Monotone Convergence Theorem, for every $\varepsilon>0$, we have that

$$\left|\int\mathrm{st}(F)d\mu_X-\int\mathrm{st}(F_m)d\mu_X\right|<\varepsilon$$

holds for all but finitely many $m\in\mathbb{N}$. Therefore, by Lemma 6.17, we have that

$$\left|\int\mathrm{st}(F)d\mu_X-\frac{1}{|X|}\sum_{x\in X}F_m(x)\right|<\varepsilon$$

holds for all but finitely many $m\in\mathbb{N}$. By transfer, there exists infinite $M\in{}^*\mathbb{N}$ such that

$$\int\mathrm{st}(F)d\mu_X=\mathrm{st}\left(\frac{1}{|X|}\sum_{x\in X}F_M(x)\right)$$

and

$$\int f d\mu_X=\mathrm{st}\left(\frac{1}{|X|}\sum_{x\in X}F_M(x)\right).$$

We show that the function F_M is S-integrable. Suppose that B is an internal subset of X such that $\mu_X(B)=0$. Set

$$r:=\mathrm{st}\left(\frac{1}{|X|}\sum_{x\in B}|F_M(x)|\right).$$

We wish to show that $r=0$. Towards this end, fix $m\in\mathbb{N}$. Then we have that

$$\begin{aligned}r+\int\mathrm{st}(F_m)d\mu_X &= r+\int_{X\setminus B}\mathrm{st}(F_m)d\mu_X\approx r+\frac{1}{|X|}\sum_{x\in X\setminus B}F_m(x)\\ &\leq r+\frac{1}{|X|}\sum_{x\in X\setminus B}F_M(x)\approx\frac{1}{|X|}\sum_{x\in X}F_M(x)\approx\int\mathrm{st}(F)d\mu_X.\end{aligned}$$

Letting $m\to+\infty$, we obtain that $r=0$, as desired.

Corollary 6.19. *Suppose $f \in L^1(X, \mathscr{L}_X, \mu_X)$ and $\varepsilon > 0$. Then there exists internal functions $F, G : X \to {}^*\mathbb{R}$ such that $F \leq f \leq G$ μ_X-almost everywhere and*

$$\max\left\{\left|\int_B f d\mu_X - \frac{1}{|X|}\sum_{x\in B} F(x)\right|, \left|\int_B f d\mu_X - \frac{1}{|X|}\sum_{x\in B} G(x)\right|\right\} \leq \varepsilon$$

for every internal subset B of X.

Proof. Let $H : X \to {}^*\mathbb{R}$ be a lifting of f. Set $F := H - \varepsilon/2$ and $G := H + \varepsilon/2$. Since $\operatorname{st}(H(x)) = f(x)$ for μ_X-almost every $x \in X$, we conclude that $F(x) \leq f(x) \leq G(x)$ for μ_X-almost every $x \in X$. Furthermore, if B is an internal subset of X, then by Lemma 6.18, we have that

$$\left|\int_B f d\mu_X - \frac{1}{|X|}\sum_{x\in B} F(x)\right| \leq \varepsilon/2 + \left|\int_B f d\mu_X - \frac{1}{|X|}\sum_{x\in B} H(x)\right| \leq \varepsilon$$

and

$$\left|\int_B f d\mu_X - \frac{1}{|X|}\sum_{x\in B} G(x)\right| \leq \varepsilon/2 + \left|\int_B f d\mu_X - \frac{1}{|X|}\sum_{x\in B} H(x)\right| \leq \varepsilon.$$

This concludes the proof.

6.5 Product measure

Suppose that $(X, \mathscr{A}_X, \nu_X)$ and $(Y, \mathscr{A}_Y, \nu_Y)$ are two probability measure spaces. We can then form their *product* as follows: first, set $\mathscr{A}$ to be the set of finite unions of rectangles of the form $A \times B$, where $A \in \mathscr{A}_X$ and $B \in \mathscr{A}_Y$; elements of $\mathscr{A}$ are called *elementary sets*. It is an exercise to show that $\mathscr{A}$ is an algebra of subsets of $X \times Y$ and that every element of $\mathscr{A}$ can be written as a finite union of *disjoint* such rectangles. We can then define a pre-measure ν on $\mathscr{A}$ by $\mu(\bigcup_{i=1}^n (A_i \times B_i)) := \sum_{i=1}^n (\nu_X(A_i) \cdot \nu_Y(B_i))$. Applying the outer measure procedure, we get a measure $\nu_X \otimes \nu_Y : \mathscr{A}_m \to [0,1]$ extending ν. We denote $\mathscr{A}_m$ by $\mathscr{A}_X \otimes \mathscr{A}_Y$.

The following situation will come up in Chapter 17: suppose that X and Y are hyperfinite sets and we construct the Loeb measure spaces $(X, \mathscr{L}_X, \mu_X)$ and $(Y, \mathscr{L}_Y, \mu_Y)$. We are thus entitled to consider the product measure space $(X \times Y, \mathscr{L}_X \otimes \mathscr{L}_Y, \mu_X \otimes \mu_Y)$. However, $X \times Y$ is itself a hyperfinite set, whence we can consider its Loeb measure space $(X \times Y, \mathscr{L}_{X\times Y,L}, \mu_{X\times Y})$. There is a connection:

Exercise 6.20. Show that $\mathscr{L}_X \otimes \mathscr{L}_Y$ is a sub-σ-algebra of $\mathscr{L}_{X\times Y}$ and that $\mu_{X\times Y}|_{(\mathscr{L}_X\otimes\mathscr{L}_Y)} = \mu_X \otimes \mu_Y$.

In the proof of the triangle removal lemma in Chapter 17, we will need to use the following Fubini-type theorem for Loeb measure on a hyperfinite set.

Theorem 6.21. *Suppose that X and Y are hyperfinite sets and $f : X \times Y \to \mathbb{R}$ is a bounded $\mathscr{L}_{X\times Y}$-measurable function. For $x \in X$, let $f_x : Y \to \mathbb{R}$ be defined by $f_x(y) := f(x,y)$. Similarly, for $y \in Y$, let $f^y : X \to \mathbb{R}$ be defined by $f^y(x) := f(x,y)$. Then:*

1. *f_x is $\mathscr{L}_Y$-measurable for μ_X-almost every $x \in X$;*
2. *f^y is $\mathscr{L}_X$-measurable for μ_Y-almost every $y \in Y$;*
3. *The double integral can be computed as an iterated integral:*

$$\int_{X\times Y} f(x,y) d\mu_{X\times Y}(x,y) = \int_X \left(\int_Y f_x(y) d\mu_Y(y)\right) d\mu_X(x) = \int_Y \left(\int_X f^y(x) d\mu_X(x)\right) d\mu_Y(y).$$

Proof. After taking positive and negative parts, it suffices to consider the case that f is positive. Furthermore, by the Monotone Convergence Theorem, it suffices to consider the case that f is a step function. Then, by linearity, one can restrict to the case that $f = \chi_E$ is the characteristic function of a Loeb measurable set $E \subseteq X \times Y$. Now Lemma 6.10 and a further application of the Monotone Convergence Theorem allows one to restrict to the case that E is internal. In this case, for $x \in X$ we have that $\int_Y \chi_E(x,y)d\mu_Y(y) = \operatorname{st}\left(\frac{|E_x|}{|Y|}\right)$, where $E_x := \{y \in Y \ : \ (x,y) \in E\}$. By Theorem 6.18, we thus have

$$\int_X \left(\int_Y \chi_E(x,y)d\mu_Y(y)\right) d\mu_X(x) \approx \frac{1}{|X|}\sum_{x\in X}\frac{|E_x|}{|Y|} = \frac{|E|}{|X||Y|} \approx \int_{X\times Y}\chi_E(x,y)d\mu_{X\times Y}(x,y).$$

The other equality is proved in the exact same way.

6.6 Ergodic theory of hypercycle systems

Definition 6.22. If $(X,\mathscr{B},\mu)$ is a probability space, we say that a bijection $T : X \to X$ is a *measure-preserving transformation* if, for all $A \in \mathscr{B}$, $T^{-1}(A) \in \mathscr{B}$ and $\mu(T^{-1}(A)) = \mu(A)$. The tuple $(X,\mathscr{B},\mu,T)$ is called a *measure-preserving dynamical system*. A measure-preserving dynamical system $(Y,\mathscr{C},\nu,S)$ is a *factor* of $(X,\mathscr{B},\mu,T)$ if there is a function $\pi : X \to Y$ such that, for $A \subseteq Y$, $A \in \mathscr{C}$ if and only if $\pi^{-1}(A) \in \mathscr{B}$, $\nu = \pi_*\mu$—which means $\nu(A) = \mu\left(\pi^{-1}(A)\right)$ for every $A \in \mathscr{C}$—and $(S\circ\pi)(x) = (\pi\circ T)(x)$ for μ-almost every $x \in X$.

Example 6.23. Suppose that $X = [0,N-1]$ is an infinite hyperfinite interval. Define $S : X \to X$ by $S(x) = x+1$ if $x < N$ and $S(N-1) = 0$. Then S is a measure-preserving transformation and the dynamical system $(X,\mathscr{L}_X,\mu_X,S)$ will be referred to as a *hypercycle system*.

The hypercycle system will play an important role later in the book. In particular, we will need to use the *pointwise ergodic theorem* for the hypercycle system. While the proof of the general ergodic theorem is fairly nontrivial, the proof for the hypercycle system, due to Kamae [31], is much simpler. In the rest of this section, we fix a hypercycle system (X,Ω_X,μ_X,S).

Theorem 6.24 (The ergodic theorem for the hypercycle system). *Suppose that $f \in L^1(X,\Omega,\mu)$. Define*

$$\hat{f}(x) := \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(S^i x)$$

whenever this limit exists. Then:

1. *$\hat{f}(x)$ exists for almost all $x \in X$;*
2. *$\hat{f} \in L^1(X,\Omega,\mu)$;*
3. *$\int_X f d\mu = \int_X \hat{f} d\mu$.*

Proof. Without loss of generality, we may assume that $X = [0,N-1]$ for some $N > \mathbb{N}$ and $f(x) \geq 0$ for μ_X-almost every $x \in X$. We set

$$\overline{f}(x) := \limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(S^i x)$$

and

$$\underline{f}(x) := \liminf_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(S^i x).$$

Note that $\overline{f},\underline{f}$ are μ_X-measurable and S-invariant. It suffices to show that $\overline{f},\underline{f} \in L^1(X,\Omega,\mu)$ and that

$$\int_X \overline{f} d\mu \leq \int_X f d\mu \leq \int_X \underline{f} d\mu.$$

Towards this end, fix $\varepsilon > 0$ and $m \in \mathbb{N}$. By Lemma 6.19, we may find internal functions $F,G : [0,N-1] \to {}^*\mathbb{R}$ such that:

- for all $x \in X$, we have $f(x) \leq F(x)$ and $G(x) \leq \min\{\bar{f}(x),m\}$;

- for every internal subset B of X

$$\max\left\{\left|\int_B f\,d\mu - \frac{1}{N}\sum_{x\in B} F(x)\right|, \left|\int_B \min\left\{\bar{f},m\right\}d\mu - \frac{1}{N}\sum_{x\in B} G(x)\right|\right\} < \varepsilon.$$

By definition of $\bar{f}$, for each $x \in X$, there is $n \in \mathbb{N}$ such that $\min\{\bar{f}(x), m\} \leq \frac{1}{n}\sum_{i=0}^{n-1} f(S^i x) + \varepsilon$. For such an n and $k = 0, 1, \ldots, n-1$, we then have that

$$G(S^k x) \leq \min\{\bar{f}(S^k x), m\} = \min\{\bar{f}(x), m\} \leq \frac{1}{n}\sum_{i=0}^{n-1} f(S^i x) + \varepsilon \leq \frac{1}{n}\sum_{i=0}^{n-1} F(S^i x) + \varepsilon,$$

whence it follows that

$$\sum_{i=0}^{n-1} G(S^i x) \leq \sum_{i=0}^{n-1} F(S^i x) + n\varepsilon. \tag{6.1}$$

Since the condition in (6.1) is internal, the function $\rho : X \to {}^*\mathbb{N}$ that sends x to the least n making (6.1) hold for x is internal. Note that $\rho(x) \in \mathbb{N}$ for all $x \in K$, whence $\sigma := \max_{x\in X} \rho(x) \in \mathbb{N}$.

Now one can start computing the sum $\sum_{x=0}^{N} G(x)$ by first computing

$$\sum_{x=0}^{\rho(0)-1} G(x) = \sum_{x=0}^{\rho(0)-1} G(S^x 0),$$

which is the kind of sum appearing in (6.1). Now in order to continue the computation using sums in which (6.1) applies, we next note that

$$\sum_{x=\rho(0)}^{\rho(0)+\rho(\rho(0))-1} G(x) = \sum_{x=0}^{\rho(\rho(0))-1} G(S^x \rho(0)).$$

This leads us to define, by internal recursion, the internal sequence (ℓ_j) by declaring $\ell_0 := 0$ and $\ell_{j+1} := \ell_j + \rho(\ell_j)$. It follows that, we have

$$\sum_{x=0}^{\ell_J-1} G(x) = \sum_{j=0}^{J-1}\sum_{i=0}^{\rho(\ell_j)-1} G(S^i \rho(\ell_j)) \leq \sum_{j=0}^{J-1}\sum_{j=0}^{\rho(\ell_j)-1} F(S^i x) + \rho(\ell_j)\varepsilon = \sum_{x=0}^{\ell_J-1} F(x) + \ell_J \varepsilon.$$

As a result, we have that, whenever $\ell_J < N$,

$$\frac{1}{N}\sum_{x=0}^{\ell_J-1} G(x) \leq \frac{1}{N}\sum_{x=0}^{\ell_J-1} F(x) + \varepsilon.$$

Now take J such that $N - \sigma \leq \ell_J < N$. Since $\sigma \in \mathbb{N}$ and $G(x) \leq m$ for every $x \in X$, we have that

$$\begin{aligned}\int_X \min\{\bar{f}, m\} d\mu &\leq \frac{1}{N}\sum_{x=0}^{N-1} G(x) + \varepsilon \approx \frac{1}{N}\sum_{x=0}^{\ell_J-1} G(x) + \varepsilon \\ &\leq \frac{1}{N}\sum_{x=0}^{\ell_J-1} F(x) + 2\varepsilon \approx \frac{1}{N}\sum_{x=0}^{N-1} F(x) + 2\varepsilon \leq \int_X f\,d\mu + 3\varepsilon.\end{aligned}$$

Letting $m \to \infty$ and then $\varepsilon \to 0$, we get that $\overline{f} \in L^1(X, \Omega, \mu)$ and $\int_X \overline{f} d\mu \leq \int_X f d\mu$. The inequality $\int_X f d\mu \leq \int_X \underline{f} d\mu$ is proven similarly.

In [31], Kamae uses the previous theorem to prove the ergodic theorem for an arbitrary measure-preserving dynamical system. In order to accomplish this, he proves the following result, which is interesting in its own right.

Theorem 6.25 (Universality of the hypercycle system). *Suppose that $(Y,\mathscr{B},\nu)$ is a standard probability space*[1] *and $T:Y\to Y$ is an measure-preserving transformation. Then $(Y,\mathscr{B},\nu,T)$ is a factor of the hypercycle system (X,Ω_X,μ_X,S).*

Proof. As before, we may assume that $X=[0,N-1]$ for some $N>\mathbb{N}$. Without loss of generality, we can assume that $(Y,\mathscr{B},\nu)$ is atomless, and hence isomorphic to $[0,1]$ endowed with the Borel σ-algebra and the Lebesgue measure. Consider the Borel map $r:[0,1]\to[0,1]^{\mathbb{N}}$ given by $r(y)(n)=h(T^n y)$ and the measure $r_*\nu$ on the Borel σ-algebra of $[0,1]^{\mathbb{N}}$. Then r defines an isomorphism between $(Y,\mathscr{B},\nu,T)$ and a factor of the unilateral Bernoulli shift on $[0,1]^{\mathbb{N}}$. Therefore, it is enough to consider the case when $(Y,\mathscr{B},\nu,T)$ is the unilateral Bernoulli shift on $[0,1]^{\mathbb{N}}$ endowed with the Borel σ-algebra $\mathscr{B}$ and some shift-invariant Borel probability measure ν.

We now define the factor map $\pi:X\to[0,1]^{\mathbb{N}}$. In order to do this, we fix $\alpha\in[0,1]^{\mathbb{N}}$ such that $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\alpha)=\int_{[0,1]^{\mathbb{N}}}f(y)d\nu$ for all $f\in C([0,1]^{\mathbb{N}})$; such an α is called *typical* in [31] and is well-known to exist.[2]

By transfer, one can identify ${}^*([0,1]^{\mathbb{N}})$ with the set of internal functions from ${}^*\mathbb{N}$ to ${}^*[0,1]$. By compactness of $[0,1]^{\mathbb{N}}$, one can deduce that, given $\xi\in{}^*([0,1]^{\mathbb{N}})$, there exists a unique element $\operatorname{st}(\xi)\in[0,1]^{\mathbb{N}}$ such that $\xi\approx\operatorname{st}(\xi)$, in the sense that, for every open subset U of $[0,1]^{\mathbb{N}}$, one has that $\xi\in{}^*U$ if and only if $\operatorname{st}(\xi)\in{}^*U$. Concretely, one can identify $\operatorname{st}(\xi)$ with the element of $[0,1]^{\mathbb{N}}$ such that $\operatorname{st}(\xi)(n)=\operatorname{st}(\xi(n))$ for $n\in\mathbb{N}$.

The function $\mathbb{N}\to[0,1]^{\mathbb{N}}$, $n\mapsto T^n\alpha$ has a nonstandard extension ${}^*\mathbb{N}\to{}^*([0,1]^{\mathbb{N}})$. Given $i\in[0,N-1]$, define $\pi(i):=\operatorname{st}(T^i\alpha)$. We must show that $\pi_*\mu_X=\nu$ and that $(T\circ\pi)(i)=(\pi\circ S)(i)$ for μ_X-almost every $i\in[0,N-1]$. For $f\in C([0,1]^{\mathbb{N}})$, we have that

$$\int_{[0,1]^{\mathbb{N}}}f(y)d\nu=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\alpha)\approx\frac{1}{N}\sum_{i=0}^{N-1}f\left(T^i\alpha\right)\approx\int_X(f\circ\pi)\,d\mu_X.$$

Note that the first step uses the fact that α is typical and the last step uses the fact that f is continuous and Theorem 6.18. This shows that

$$\int_{[0,1]^{\mathbb{N}}}fd\nu=\int_X(f\circ\pi)\,d\mu_X=\int_{[0,1]^{\mathbb{N}}}fd\pi_*\mu_X$$

and hence $\nu=\pi_*\mu_X$.

To finish, we show that $(T\circ\pi)(i)=(\pi\circ S)(i)$ for μ_X-almost every $i\in X$. Fix $i\in[0,N-2]$. Then we have

$$T(\pi(i))=T(\operatorname{st}(T^i\alpha))=\operatorname{st}(T^{i+1}\alpha))=\pi(S(i)),$$

where the second equality uses the fact that T is continuous.

From Theorems 6.24 and 6.25, we now have a proof of the ergodic theorem for measure-preserving systems based on standard probability spaces. It only requires one more step to obtain the ergodic theorem in general.

Corollary 6.26 (The ergodic theorem). *Suppose that $(Y,\mathscr{B},\nu,T)$ is a measure-preserving dynamical system and $f\in L^1(X,\Omega,\mu)$. Define $\hat{f}(x):=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^ix)$ whenever this limit exists. Then:*

1. *$\hat{f}(x)$ exists for almost all $x\in Y$;*
2. *$\hat{f}\in L^1(Y,\mathscr{B},\nu)$;*
3. *$\int_Y fd\nu=\int_Y\hat{f}d\nu$.*

Proof. Let $\tau:Y\to\mathbb{R}^{\mathbb{N}}$ be given by $\tau(y)(n):=f(T^ny)$. Let $\mathscr{C}$ denote the Borel σ-algebra of $\mathbb{R}^{\mathbb{N}}$. Let σ be the shift operator on $\mathbb{R}^{\mathbb{N}}$. Let $g:\mathbb{R}^{\mathbb{N}}\to\mathbb{R}$ be given by $g(\alpha)=\alpha(0)$. It is then readily verified that the conclusion of the ergodic theorem for $(Y,\mathscr{B},\nu,T,f)$ is equivalent to the truth of the ergodic theorem for $(\mathbb{R}^{\mathbb{N}},\mathscr{C},\tau_*\nu,\sigma,g)$, which, as we mentioned above, follows from Theorems 6.24 and 6.25.

[1] Unfortunately, *standard* is used in a different sense than in the rest of this book. Indeed, here, a standard probability space is simply a probability space which is isomorphic to a quotient of $[0,1]$ endowed with the Borel σ-algebra and Lebesgue measure.

[2] Of course, one can use the ergodic theorem to prove the existence of typical elements. However, we need a proof that typical elements exist that does not use the ergodic theorem. One can see, for example, [31, Lemma 2] for such a proof.

Part II
Ramsey theory

Chapter 7
Ramsey's Theorem

7.1 Infinite Ramsey's Theorem

Recall that a *graph* is a pair (V,E) where V is the set of *vertices*, and the set of *edges* $E \subseteq V \times V$ is an anti-reflexive and symmetric binary relation on V. If $X \subseteq V$ is such that $(x,x') \in E$ (resp. $(x,x') \notin E$) for all distinct $x,x' \in X$, we say that X is a *clique* (resp. *anticlique*) in (V,E).

Theorem 7.1 (Ramsey's theorem for pairs). *If (V,E) is an infinite graph, then (V,E) either contains an infinite clique or an infinite anticlique.*

Proof. Let ξ be an element of ${}^*V$ that does not belong to V. Consider the element $(\xi, {}^*\xi) \in {}^{**}V$. There are now two possibilities: either $(\xi, {}^*\xi) \in {}^{**}E$ or $(\xi, {}^*\xi) \notin {}^{**}E$. We only treat the first case, the second case being entirely similar. We recursively define a one-to-one sequence (x_n) in V such that the set $\{x_n : n \in \mathbb{N}\}$ forms a clique in (V,E). Towards this end, suppose that $d \in \mathbb{N}$ and $x_0, \ldots, x_{d-1}$ are distinct elements of V such that, for all $1 \leq i < j < d$, we have

- $(x_i, x_j) \in E$, and
- $(x_i, \xi) \in {}^*E$.

Consider now the statement "there exists $y \in {}^*V$ such that, for $i < d$, y is different from x_i, and $(x_i, y) \in {}^*E$, and $(y, {}^*\xi) \in {}^{**}E$", whose truth is witnessed by ξ. It follows by transfer that there exists $x_d \in V$ different from x_i for $i < d$, such that $(x_i, x_d) \in E$ for $i < d$, and $(x_d, \xi) \in {}^*E$. This concludes the recursive construction.

In order to prove the full Ramsey theorem, we need the notion of a hypergraph. Given $m \in \mathbb{N}$, an *m-regular hypergraph* is a set V of vertices together with a subset E of V^m that is permutation-invariant and has the property that $(x_1, \ldots, x_m) \in E$ implies that $x_1, \ldots, x_m$ are pairwise distinct. A *clique* (resp. *anticlique*) for (V,E) is a subset Y of V with the property that $(y_1, \ldots, y_m) \in E$ (resp. $(y_1, \ldots, y_m) \notin E$) for any choice of pairwise distinct elements $y_1, \ldots, y_m$ of Y.

Theorem 7.2 (Ramsey's theorem). *If (V,E) is an infinite m-regular hypergraph, then (V,E) contains an infinite clique or an infinite anticlique.*

Proof. For simplicity, we consider the case when $m = 3$. Let ξ be an element of ${}^*V$ that does not belong to V. As before, there are now two cases, depending on whether $(\xi, {}^*\xi, {}^{**}\xi)$ belongs to ${}^{***}E$ or not. Once again, we only treat the first case.

We recursively define a one-to-one sequence (x_n) of elements of V such that $\{x_n : n \in \mathbb{N}\}$ forms a clique for V. Towards this end, suppose that $d \in \mathbb{N}$ and $x_0, \ldots, x_{d-1}$ are distinct elements of V such that, for all $1 \leq i < j < k < d$, we have:

- $(x_i, x_j, x_k) \in E$,
- $(x_i, x_j, \xi) \in {}^*E$, and
- $(x_i, \xi, {}^*\xi) \in {}^{**}E$.

Consider now the statement "there exists $y \in {}^*V$ such that y is different from x_i for $1 \leq i < d$, $(x_i, x_j, y) \in {}^*E$ for every $1 \leq i < j < d$, $(x_i, y, {}^*\xi) \in {}^{**}E$ for every $1 \leq i < d$, and $(y, {}^*\xi, {}^{**}\xi) \in {}^{***}E$." Note that ξ witnesses the truth of the statement in the nonstandard extension. Therefore, by transfer, there is an element x_d of V distinct from x_i for $1 \leq i < d$ for which the above items remain true for all $1 \leq i < j < k \leq d$. This completes the recursive construction.

Ramsey's theorem is often stated in the language of colorings. Given a set X and $m \in \mathbb{N}$, we let $X^{[m]}$ denote the set of m-element subsets of X. We often identify $X^{[m]}$ with the set of pairs $\{(x_1, \ldots, x_m) \in X^m \ : \ x_1 < \cdots < x_m\}$. Given $k \in \mathbb{N}$, a k-coloring of $X^{[m]}$ is a function $c : X^{[m]} \to \{1, \ldots, k\}$. In this vein, we often refer to the elements of $\{1, \ldots, k\}$ as colors. Finally, a subset $Y \subseteq X$ is *monochromatic for the coloring c* if the restriction of c to $Y^{[m]}$ is constant. Here is the statement of Ramsey's theorem for colorings.

Corollary 7.3. *For any $k, m \in \mathbb{N}$, any infinite set V, and any k-coloring c of $V^{[m]}$, there is an infinite subset of V that is monochromatic for the coloring c.*

Proof. By induction, it suffices to consider the case $k = 2$. We identify a coloring $c : V^{[m]} \to \{1, 2\}$ with the m-regular hypergraph (V, E) satisfying $(x_1, \ldots, x_m) \in E$ if and only if $c(\{x_1, \ldots, x_m\}) = 1$ for distinct $x_1, \ldots, x_m \in V$. An infinite clique (resp. anticlique) in (V, E) corresponds to an infinite set with color 1 (resp. 2), whence the corollary is merely a restatement of our earlier version of Ramsey's theorem.

Remark 7.4. Ramsey's Theorem cannot be extended to finite colorings of the infinite parts $V^{[\infty]} = \{A \subseteq V \mid A \text{ is infinite}\}$. Indeed, pick a copy of the natural numbers $\mathbb{N} \subseteq V$, pick an infinite $\xi \in {}^*\mathbb{N} \setminus \mathbb{N}$, and for $A \in V^{[\infty]}$ set $c(A) = 1$ if the internal cardinality $|{}^*A \cap [1, \alpha]|$ is odd, and $c(A) = 2$ otherwise. Then $c : V^{[\infty]} \to \{1, 2\}$ is a 2-coloring with the property that $X^{[\infty]}$ is *not* monochromatic for any infinite $X \subseteq V$ since, *e.g.*, $c(X) \neq c(X \setminus \{x\})$ for every $x \in X$.

7.2 Finite Ramsey Theorem

Corollary 7.3 is often referred to as the infinite Ramsey theorem. We now deduce from it the finite Ramsey theorem. We first need a bit of notation.

Definition 7.5. Given $k, l, m, n \in \mathbb{N}$, we write $l \to (n)^m_k$ if every coloring of $[l]^{[m]}$ with k colors has a homogeneous set of size n.

Corollary 7.6 (Finite Ramsey Theorem). *For every $k, m, n \in \mathbb{N}$, there is $l \in \mathbb{N}$ such that $l \to (k)^n_m$.*

Proof. Suppose the theorem is false for a particular choice of k, m, n. Then for every $l \in \mathbb{N}$, there is a "bad" coloring $c : [l]^{[m]} \to \{1, \ldots, k\}$ with no monochromatic subset of size n. We can form a finitely branching tree of bad colorings with the partial order being inclusion. Since there is a bad coloring for every such l, we have that the tree is infinite. By König's Lemma, there is an infinite branch. This branch corresponds to a coloring of $[\mathbb{N}]^{[m]} \to \{1, \ldots, k\}$ with no monochromatic subset of size n, contradicting the Infinite Ramsey Theorem.

7.3 Rado's Path Decomposition Theorem

In this section, by a *path in* $\mathbb{N}$ we mean a (finite or infinite) injective sequence of natural numbers. For a finite path $(a_0, \ldots, a_n)$ from $\mathbb{N}$, we refer to a_n as the *end of the path*.

Suppose that $c : [\mathbb{N}]^2 \to \{1, \ldots, r\}$ is an r-coloring of $[\mathbb{N}]^2$. For $i \in \{1, \ldots, r\}$, we say that a path $P = (a_n)$ has color i if $c(\{a_n, a_{n+1}\}) = i$ for all n.

Theorem 7.7 (Radó's Path Decomposition Theorem). *Suppose that $c : [\mathbb{N}]^2 \to \{1, \ldots, r\}$ is an r-coloring of $[\mathbb{N}]^2$. Then there is a partition of $\mathbb{N}$ into paths $P_1, \ldots, P_r$ such that each P_i has color i.*

Proof. First, fix $\alpha \in {}^*\mathbb{N}$. For $m \in \mathbb{N}$ and $i \in \{1,\ldots,r\}$, we say that m has color i if $c(\{m,\alpha\}) = i$. We now recursively define disjoint finite paths $P_{1,k},\ldots,P_{r,k}$ such that, whenever $P_{i,k} \neq \emptyset$, then the end of $P_{i,k}$ has color i (in the sense of the previous sentence).

To start, we define $P_{i,0} = \emptyset$ for each $i = 1,\ldots,r$. Now assume that $P_{i,k-1}$ has been constructed for $i = 1,\ldots,r$. If k belongs to some $P_{i,k-1}$, then set $P_{i,k} := P_{i,k-1}$ for all $i = 1,\ldots,r$. Otherwise, let i be the color of k and let e be the end of $P_{i,k-1}$. Since $c(\{k,\alpha\}) = c(\{e,\alpha\}) = i$, by transfer, we can find $f \in \mathbb{N}$ larger than all numbers appearing in $\bigcup_{i=1}^{r} P_{i,k-1}$ such that $c(\{k,f\}) = c(\{e,f\}) = i$. We then set $P_{j,k} := P_{j,k-1}$ for $j \neq i$ and $P_{i,k} := P_{i,k-1}{}^\frown(f,k)$. Note that the recursive assumptions remain true.

For $i = 1,\ldots,r$, we now set $P_i := \lim_k P_{i,k}$ (in the sense of the product topology on $\mathbb{N}^{\mathbb{N}}$). It is clear that $P_1,\ldots,P_r$ are as desired.

For more on the metamathematics of Rado's Decomposition Theorem, see [9], whose ultrafilter proof of the theorem is essentially the proof given here.

Chapter 8
van der Waerden's and Hales-Jewett Theorems

8.1 van der Waerden's theorem

The van der Waerden theorem is one of the earliest achievements of what is now call Ramsey theory. Indeed, it was established by van der Waerden in 1928 [47], and thus predating Ramsey's theorem itself. The theorem is concerned with the notion of *arithmetic progressions* in the set $\mathbb{N}$ of natural numbers. More precisely, for $k \in \mathbb{N}$, a k-term arithmetic progression in $\mathbb{N}$ is a set of the form $a + d[0,k) := \{a, a+d, a+2d, \ldots, a+(k-1)d\}$ for some $a, d \in \mathbb{N}$. A k-term arithmetic progression is also called an arithmetic progression of length k. An arithmetic progression in ${}^*\mathbb{N}$ is defined in a similar fashion, where one can actually consider k-term arithmetic progressions for $k \in {}^*\mathbb{N}$.

A collection $\mathscr{C}$ of subsets of $\mathbb{N}$ is *partition regular* if it is closed under supersets and, whenever an element of $\mathscr{C}$ is partition into finitely many pieces, at least one of the pieces of the partition must belong to $\mathscr{C}$.

Theorem 8.1. *The following are equivalent:*

1. *If $\mathbb{N}$ is partitioned into finitely many colors, then some color contains arbitrarily long arithmetic progressions.*
2. *For every $r, k \in \mathbb{N}$, there is $l \in \mathbb{N}$ such that if $[1,l]$ is partitioned into r colors, then some color contains a k-term arithmetic progression.*
3. *The property of containing arbitrarily long arithmetic progressions is partition regular.*

Proof. (1)⇒(2) Suppose that (2) fails for some k, r. By , there is $L > \mathbb{N}$ and an internal r-coloring of $[1,L]$ with no monochromatic k-term arithmetic progression. By considering the restriction of c to $\mathbb{N}$, we get an r-coloring of $\mathbb{N}$ with no monochromatic k-term arithmetic progression, whence (1) fails.

(2)⇒(3) Suppose that (2) holds. Towards establishing (3), fix a set A containing arbitrarily long arithmetic progressions and a partition of A into two pieces $A = B_1 \sqcup B_2$. Fix $k \in \mathbb{N}$. Let l witness the truth of (2) with 2 colors and k-term arithmetic progressions. Fix an arithmetic progression $x + [0,l)d \subseteq A$. For $i = 1, 2$, let $C_i := \{n \in [0,l) \ : \ x + nd \in B_i\}$. Then there is $i \in \{1,2\}$ such that C_i contains an arithmetic progression $y + [0,k)e$. It follows that $(x+yd) + [0,k)de$ is a k-term arithmetic progression contained in B_i. Since some i must work for infinitely many k's, we see that some B_i contains arbitrarily long arithmetic progressions.

(3)⇒(1) This is obvious.

The following is a nonstandard presentation of the proof of van der Waerden's theorem from [22]; see also [40, Section 2.3]. First, some terminology. For $k, m \in \mathbb{N}$ and $g, h \in [0,k]^m$, we say that g and h are equivalent, written $g \equiv h$, if g and h agree up to the last occurrence of k.

Definition 8.2. For $k, m \in \mathbb{N}$, let $S(m,k,r,n)$ be the statement: for any r-coloring of $[1,n]$, there exist $a, d_0, \ldots, d_{m-1} \in [1,n]$ such that $a + k\sum_{j<m} d_j \in [1,n]$ and, for any $g, h \in [0,k]^m$ such that $g \equiv h$, the elements $a + \sum_{j<m} g_j d_j$ and $a + \sum_{j<m} h_j d_j$ have the same color. We then let $S(m,k)$ be the statement: for all $r \in \mathbb{N}$, there is $n \in \mathbb{N}$ such that $S(m,k,r,n)$ holds.

We first observe that even though the statement $S(m,k,r,n)$ considers colorings of $[1,n]$, it is readily verified that its truth implies the corresponding statement for colorings of any interval of length n.

We next observe that the finitary van der Waerden theorem is the statement that $S(k,1)$ holds for all $k \in \mathbb{N}$. Indeed, suppose that $S(k,1)$ holds and fix $r \in \mathbb{N}$. Fix $n \in \mathbb{N}$ such that $S(k,1,r,n)$ holds. Let $c : [1,n] \to [1,r]$ be an r-coloring of $[1,n]$. Then there is $a,d \in [1,n]$ such that $a+kd \in [1,n]$ and, since all elements of $[0,k]^1$ are equivalent, we get that $c(a+gd) = c(a+hd)$ for all $g,h \in [0,k]$, whence we get a monochromatic arithmetic progression of length $k+1$.

If $\nu \in {}^*\mathbb{N}$, then we also consider the internal statement $S(m,k,r,\nu)$ which is defined exactly as its standard counterpart except that it only considers internal r-colorings of $[1,\nu]$.

Lemma 8.3. *$S(k,m)$ is equivalent to the statement: for all $r \in \mathbb{N}$ and all $\nu \in {}^*\mathbb{N} \setminus \mathbb{N}$, we that have that $S(m,k,r,\nu)$ holds.*

Proof. First suppose that $S(k,m)$ holds. Given $r \in \mathbb{N}$, take $n \in \mathbb{N}$ such that $S(k,m,r,n)$ holds. Fix $\nu \in {}^*\mathbb{N} \setminus \mathbb{N}$ and consider an internal r-coloring c of $[1,\nu]$. Then $c|_{[1,n]}$ is an r-coloring of $[1,n]$, whence the validity of $S(k,m,r,n)$ yields the desired conclusion. Conversely, if $S(k,m,r,\nu)$ holds for all $\nu \in {}^*\mathbb{N} \setminus \mathbb{N}$, then by underflow there is $n \in \mathbb{N}$ such that $S(k,m,r,n)$ holds.

Theorem 8.4. *$S(k,m)$ holds for all $k,m \in \mathbb{N}$.*

Proof. Suppose, towards a contradiction, that $S(k,m)$ fails for the pair (k,m) and that (k,m) is lexicographically least with this property.

Claim: $m = 1$.
Proof of Claim: Suppose the claim is false. We obtain a contradiction by showing that $S(k,m,r,\nu)$ holds for all $r \in \mathbb{N}$ and all $\nu \in {}^*\mathbb{N} \setminus \mathbb{N}$. Towards this end, fix $r \in \mathbb{N}$, $\nu \in {}^*\mathbb{N} \setminus \mathbb{N}$, and an internal coloring $c : [1,\nu] \to [1,r]$. Since $S(k,m-1)$ is true, there is $M \in \mathbb{N}$ such that $S(k,m-1,r,M)$ is true. Write $\nu = NM + s$ with $0 \le s < M$. Note that $N \in {}^*\mathbb{N} \setminus \mathbb{N}$. Consider the internal coloring $c_N : [1,N] \to [1,r^M]$ given by

$$c_N(i) := (c((i-1)M+1), \ldots, c((i-1)M+M)).$$

Since $S(k,1,r,N)$ holds, there is an arithmetic progression $b+d, b+2d, \ldots, b+kd$ contained in $[1,N]$ that is monochromatic for the coloring c_N. Next, since $S(k,m-1,r,M)$ holds, by considering $c|_{[(b-1)M,bM]}$, we see that there are $a, d_0, \ldots, d_{m-2} \in [(b-1)M, bM]$ such that $a + k\sum_{j<m-1} d_j \in [(b-1)M, bM]$ and, for any $g,h \in [0,k]^{m-1}$ such that $g \equiv h$, the elements $a + \sum_{j<m} g_j d_j$ and $a + \sum_{j<m} h_j d_j$ have the same color with respect to c.

Set $d_{m-1} := dM$. We claim that $a, d_0, \ldots, d_{m-1}$ are as desired. First note that $a + k\sum_{j<m} d_j \le bM + kdM \le NM \le \nu$. Next suppose that $g,h \in [0,k]^m$ are such that $g \equiv h$. We wish to show that $a + \sum_{j<m} g_j d_j$ and $a + \sum_{j<m} h_j d_j$ have the same color. If the last occurrence of k is $m-1$, then this is obvious. Otherwise, we see that $g \mid m-1 = h \mid m-1$, whence by assumption $a + \sum_{j<m-1} g_j d_j$ and $a + \sum_{j<m-1} h_j d_j$ have the same color. Write $a + \sum_{j<m-1} g_j d_j = (b-1)M + p$ with $p \in [1,M]$. Then $a + \sum_{j<m} g_j d_j = (b-1)M + p + g_{M-1} dM = (b + g_{m-1} d - 1)M + p$, which has the same color as $(b-1)M + p$ by assumption. Likewise, $a + \sum_{j<m-1} h_j d_j = (b-1)M + q$ with $q \in [1,M]$, whence $a + \sum_{j<m} h_j d_j = (b-1)M + q + h_{M-1} dM = (b + h_{m-1} d - 1)M + q$, which has the same color as $(b-1)M + q$ by assumption. Thus, $a + \sum_{j<m-1} g_j d_j$ and $a + \sum_{j<m-1} h_j d_j$ have the same color, proving the claim.

Since $S(k,1)$ fails, necessarily we have $k > 1$. We will arrive at a contradiction by showing that $S(k,1)$ in fact holds. Fix $r \in \mathbb{N}$, $\nu \in {}^*\mathbb{N}$ infinite, and an internal r-coloring c of $[1,\nu]$. By minimality of $(k,1)$, we have that there exist $a, d_0, \ldots d_{r-1} \in [1,\nu]$ such that $a + r\sum_{j<r} d_j \in [1,\nu]$ and, for any $g,h \in [1,k-1]^r$ with $g \equiv h$, we have $a + \sum_{j<r} g_j d_j$ and $a + \sum_{j<r} h_j d_j$ have the same color. Observe that there are $r+1$ r-tuples that are obtained by concatenating a (possibly empty) r-tuple of $(k-1)$'s and a (possibly empty) r-tuple of 0's. Hence, by the pigeonhole principle, there exist $1 \le s < t \le r$ such that $a + (k-1)\sum_{i<s} d_i$ and $a + (k-1)\sum_{i<t} d_i$ have the same color. We also have that $a + (k-1)\sum_{i<s} d_i$ and $a + (k-1)\sum_{i<s} d_i + j\sum_{s \le i<t} d_i$ have the same color for every $j < k-1$. Therefore, setting $a' := a + (k-1)\sum_{i<s} d_i$ and $d' := \sum_{s \le i<t} d_i$, we have that $a' + jd'$, for $j < k$, all have the same color. Since $\nu \in {}^*\mathbb{N} \setminus \mathbb{N}$ and c were arbitrary, this witnesses that $S(k,1)$ holds, yielding the desired contradiction.

We will see in the next section that the Hales-Jewett theorem allows us to immediately conclude a generalization of the van der Waerden theorem.

8.2 The Hales-Jewett theorem

Let L be a *finite* set (alphabet). We use the symbol x to denote a *variable* not in L. We let W_L denote the set of finite strings of elements of L (called *words* in L), and W_{Lx} denote the set of finite strings of elements of $L \cup \{x\}$ with the property that x appears at least once (called *variable words*). We denote (variable) words by v, w, z and letters by a, b, c. If w is a variable word and a is a letter, then we denote by $w[a]$ the word obtained from w by replacing every occurrence of x with a. For convenience, we also set $w[x] := w$. The concatenation of two (variable) words v, w is denoted by $v^\frown w$.

Definition 8.5. Fix a sequence (w_n) of variable words

1. The *partial subsemigroup of W_L generated by* (w_n), denoted $[(w_n)]_{W_L}$, is the set of all words $w_{n_0}[a_0]^\frown \cdots {}^\frown w_{n_{k-1}}[a_{k-1}]$, where $k \in \mathbb{N}$, $n_0 < \cdots < n_{k-1}$, and $a_0, \ldots, a_{k-1} \in L$.
2. The *partial subsemigroup of W_{Lx} generated by* (w_n), denoted $[(w_n)]_{W_{Lx}}$, is the set of all words $w_{n_0}[\lambda_0]^\frown \cdots {}^\frown w_{n_{k-1}}[\lambda_{k-1}]$, where $k \in \mathbb{N}$, $n_0 < \cdots < n_{k-1}$, $\lambda_0, \ldots, \lambda_{k-1} \in L \cup \{x\}$, and some $\lambda_i = x$.

Theorem 8.6 (Infinite Hales-Jewett). *For every finite coloring of $W_L \cup W_{Lx}$ there exists an infinite sequence (w_n) of variable words such that $[(w_n)]_{W_L}$ and $[(w_n)]_{W_{Lx}}$ are both monochromatic.*

There is also a finitary version of the Hales-Jewett theorem. Suppose that $x_1, \ldots, x_m$ are variables. A variable word w in the variables $x_1, \ldots, x_m$ in the alphabet L is a string of symbols in $L \cup \{x_1, \ldots, x_m\}$ such that, for every $1 \leq i \leq m$, x_i occurs in w, and for every $1 \leq i < j \leq m$, every occurrence of x_i precedes every occurrence of x_j. The word $w[a_1, \ldots, a_m]$ obtained from w by substituting the variable x_i with the letter a_i for $i = 1, 2, \ldots, m$ is defined in the obvious way.

Corollary 8.7 (Finite Hales-Jewett). *For any finite alphabet L and any $r, m \in \mathbb{N}$ there exists $n \in \mathbb{N}$ such that for any r-coloring of the set $W_L(n)$ of L-words of length n there exist a variable word w of length n in the alphabet L and variables $x_1, \ldots, x_m$ such that the "combinatorial m-subspace" $\{w[a_1, \ldots, a_m] : a_1, \ldots, a_n \in L\}$ is monochromatic.*

A combinatorial m-subspace for $m = 1$ is usually called a *combinatorial line*.

Proof. We just establish the case that $m = 1$. We let $W_{Lx}(n)$ denote the elements of W_{Lx} of length n and $W_L(n)$ denote the elements of W_L of length n. Suppose, towards a contradiction, that there is $r \in \mathbb{N}$ such that, for each n, there is a "bad" r-coloring of $W_L(n)$ that admits no monochromatic combinatorial line. By a compactness argument there is an r-coloring c of W_L such that the restriction of c to $W_L(n)$ is a bad r-coloring for every $n \in \mathbb{N}$. By the Infinite Hales-Jewett Theorem, there is a sequence (w_i) for which $[(w_i)]_{W_L}$ is monochromatic. For $i = 1, 2, \ldots, m$, rename the variable x of w_i by x_i, and consider the variable word $w := w_1{}^\frown w_2{}^\frown \cdots {}^\frown w_m$ in the variables $\{x_1, \ldots, x_m\}$. If n is the length of w, then by the choice of $w_1, \ldots, w_m$ the combinatorial subspace $\{w[a_1, \ldots, a_m] : a_1, \ldots, a_n \in L\}$ is monochromatic. This contradicts the fact that the restriction of c to $W_L(n)$ is a bad r-coloring.

From the Hales-Jewett theorem one can deduce a multidimensional generalization of van der Waerden's theorem, known as *Gallai's theorem*.

Theorem 8.8 (Gallai). *Fix $d \in \mathbb{N}$, a finite $F \subset \mathbb{N}^d$, and $r \in \mathbb{N}$. Then there exists $n \in \mathbb{N}$ such that, for any r-coloring of $[-n, n]^d$, there exist $a \in \mathbb{N}^d$ and $c \in \mathbb{N}$ such that the* affine image $a + cF := \{a + cx : x \in F\}$ *of F is monochromatic.*

Proof. Consider the finite alphabet $L = F$. For $n \in \mathbb{N}$, consider the map $\Psi_n : W_L(n) \to \mathbb{N}^d$ defined by $\Psi_n((a_1, \ldots, a_n)) = a_1 + \cdots + a_n$. Observe that Ψ_n maps a combinatorial line to an affine image of F. Thus the conclusion follows from the finitary Hales-Jewett theorem.

In the rest of the section we present the proof of Theorem 8.6. Consider W_L and $W_L \cup W_{Lx}$ as semigroups with respect to concatenation. Thus their nonstandard extensions ${}^*W_L$ and ${}^*W_L \cup {}^*W_{Lx}$ have canonical semigroup operations with respect to the nonstandard extension of the concatenation operation, which we still denote by "$^\frown$". The elements of ${}^*W_L$ can be regarded as hyperfinite strings of elements of ${}^*L$, and similarly for ${}^*W_{Lx}$. For every $a \in L \cup \{x\}$ we also denote by $\varpi \mapsto \varpi[a]$ the nonstandard extension of the substitution operation $W_{Lx} \to W_L$, $w \mapsto w[a]$.

Lemma 8.9. *There exists a u-idempotent ϖ in ${}^*W_{Lx}$ and a u-idempotent $\upsilon \in {}^*W_L$ such that $\varpi^\frown {}^*\upsilon \sim \upsilon^\frown {}^*\varpi \sim \varpi$ and $\varpi[a] \sim \upsilon$ for every $a \in L$.*

Proof. Fix an enumeration $\{a_1,\dots,a_m\}$ of L. We define, by recursion on $k = 1,\dots,m$, u-idempotent elements $\varpi_1,\dots,\varpi_m$ of ${}^*W_{Lx}$ and $\upsilon_1,\dots,\upsilon_m$ of ${}^*W_L$ such that, for $1 \le i \le j \le m$,

1. $\varpi_j[a_i] \sim \upsilon_j$, and
2. $\varpi_j \sim \varpi_j{}^\frown {}^*\upsilon_i \sim \upsilon_i{}^\frown {}^*\varpi_j$.

Supposing this has been done, the conclusion of the lemma holds by taking $\varpi := \varpi_m$ and $\upsilon := \upsilon_m$.

To begin, we let ϖ_0 be any nontrival u-idempotent element of ${}^*W_{Lx}$ and set $\upsilon_1 := \varpi_0[a_1]$, which we note is an idempotent element of ${}^*W_L$. Let ρ_1 be an element of ${}^*W_{Lx}$ such that $\rho_1 \sim \varpi_0{}^\frown {}^*\upsilon_1$. Observe that $\rho_1[a_1] \sim \upsilon_1$ and $\rho_1{}^\frown {}^*\upsilon_1 \sim \rho_1$. Thus, the compact u-semigroup

$$\{z \in {}^*W_{Lx} : z[a_1] \sim \upsilon_1 \text{ and } z^\frown {}^*\upsilon_1 \sim z\}$$

is nonempty, whence it contains a u-idempotent β_1. We now fix $\varpi_1 \in {}^*W_{Lx}$ such that $\varpi_1 \sim \upsilon_1{}^\frown {}^*\beta_1$. It follows now that ϖ_1 is u-idempotent and ϖ_1 and υ_1 satisfy (1) and (2) above.

Suppose that ϖ_i, υ_i have been defined for $1 \le i \le k < m$ satisfying (1) and (2) above. Set $\upsilon_{k+1} := \varpi_k[a_{k+1}]$. Observe that $\upsilon_{k+1} \sim \upsilon_{k+1}{}^\frown {}^*\upsilon_i \sim \upsilon_i{}^\frown {}^*\upsilon_{k+1}$ for $1 \le i \le k+1$. Let ρ_{k+1} be an element of ${}^*W_{Lx}$ such that $\rho_{k+1} \sim \varpi_k{}^\frown {}^*\upsilon_{k+1}$. Observe that $\upsilon_i{}^\frown {}^*\rho_{k+1} \sim \rho_{k+1}{}^\frown {}^*\upsilon_i \sim \rho_{k+1}$ and $\rho_{k+1}[a_i] \sim \upsilon_{k+1}$ for $1 \le i \le k+1$. Thus, the compact u-semigroup

$$\{z \in {}^*W_{Lx} : z[a_i] \sim \upsilon_{k+1} \text{ and } z^\frown {}^*\upsilon_i \sim z \text{ for } 1 \le i \le k+1\}$$

is nonempty, whence it contains a u-idempotent element β_{k+1}. Finally, fix ϖ_{k+1} in ${}^*W_{Lx}$ such that $\varpi_{k+1} \sim \upsilon_{k+1}{}^\frown {}^*\beta_{k+1}$. It follows that ϖ_{k+1} is u-idempotent and (1) and (2) continue to hold for ϖ_i and υ_i for $1 \le i \le k+1$. This completes the recursive construction and the proof of the lemma.

In the statement of the following proposition, we assume that ϖ and υ are as in the conclusion of Lemma 8.9.

Proposition 8.10. *Suppose that $A \subset W_L$ and $B \subset W_{Lx}$ are such that $\upsilon \in {}^*A$ and $\varpi \in {}^*B$. Then there exists an infinite sequence (w_n) in W_{Lx} such that $[(w_n)]_{W_L}$ is contained in A and $[(w_n)]_{W_{Lx}}$ is contained in B.*

Proof. Set $C := A \cup B$. Observe that ϖ satisfies, for every $a, b \in L \cup \{x\}$,

$$\varpi[a] \in {}^*C$$
$$\varpi[a]^\frown {}^*\varpi[b] \in {}^{**}C.$$

Therefore, by transfer, there exists $w_0 \in W_{Lx}$ that satisfies, for every $a_0, a_1 \in L \cup \{x\}$,

$$w_0[a_0] \in C$$
$$w_0[a_0]^\frown \varpi[a_1] \in {}^*C.$$

From this we also have, for every $a_0, a_1, b \in L \cup \{x\}$, that,

$$w_0[a_0]^\frown \varpi[a_1]^\frown {}^*\varpi[b] \in {}^{**}C.$$

Therefore, by transfer, there exists $w_1 \in W_{Lx}$ that satisfies, for every $a_0, a_1, a_2 \in L \cup \{x\}$:

$$w_0[a_0] \in C$$
$$w_1[a_1] \in C$$
$$w_0[a_0]^\frown w_1[a_1] \in C$$
$$w_0[a_0]^\frown \varpi[a_2] \in {}^*C$$
$$w_1[a_1]^\frown \varpi[a_2] \in {}^*C$$
$$w_0[a_0]^\frown w_1[a_1]^\frown \varpi[a_2] \in {}^*C.$$

Proceeding recursively, one can assume that at the n-th step elements $w_0, \dots, w_{n-1}$ of W_{Lx} have been defined such that, for every $n_1 < \cdots < n_k < n$ and $a_0, \dots, a_{n-1}, a \in L \cup \{x\}$, one has that

$$w_{n_1}[a_{n_1}]^\frown \cdots ^\frown w_{n_k}[a_{n_k}] \in C$$
$$w_{n_1}[a_{n_1}]^\frown \cdots ^\frown w_{n_k}[a_{n_k}]^\frown \varpi[a] \in {}^*C.$$

From this one deduces also that for every $a, b \in L \cup \{x\}$ one has that

$$w_{n_1}[a_{n_1}]^\frown \cdots ^\frown w_{n_k}[a_{n_k}]^\frown \varpi[a]^{\frown *}\varpi[b] \in {}^{**}C.$$

Hence, by transfer one obtains $w_n \in W_{Lx}$ such that for every $n_1 < \cdots < n_k \leq n$ and $a_0, \dots, a_n, a \in L \cup \{x\}$, one has that

$$w_{n_1}[a_{n_1}]^\frown \cdots ^\frown w_{n_k}[a_{n_k}] \in C$$
$$w_{n_1}[a_{n_1}]^\frown \cdots ^\frown w_{n_k}[a_{n_k}]^\frown \varpi[a] \in {}^*C.$$

This concludes the recursive construction.

Theorem 8.6 now follows immediately from Proposition 8.10. Indeed, if $\{A_1, \dots, A_r\}$ is a finite coloring of $W_L \cup W_{Lx}$, then there exist $1 \leq i, j \leq r$ such that $\upsilon \in {}^*A_i$ and $\varpi \in {}^*A_j$.

Chapter 9
From Hindman to Gowers

9.1 Hindman's theorem

Hindman's theorem is another fundamental pigeonhole principle, which considers the combinatorial configurations provided by sets of finite sums of infinite sequences.

Definition 9.1. 1. Given $F \subseteq \mathbb{N}$ finite and (c_n) a sequence of distinct elements from A, define $c_F := \sum_{n \in F} c_n$, with the convention that $c_\emptyset = 0$.
2. Given a (finite or infinite) sequence (c_n) of distinct elements from $\mathbb{N}$, set $\mathrm{FS}((c_n)) := \{c_F \ : \ F \subseteq \mathbb{N}$ finite, nonempty$\}$.
3. We say that $A \subseteq \mathbb{N}$ is an *FS-set* if there is an infinite sequence (c_n) of distinct elements from $\mathbb{N}$ such that $\mathrm{FS}((c_n)) \subseteq A$.

We first note, using the notation from Section 8.1, that $S(m,2)$ implies the following theorem:

Theorem 9.2 (Folkman's theorem). *For any $m, r \in \mathbb{N}$, there is $n \in \mathbb{N}$ such that, for any r-coloring of $[1,n]$, there are $d_0, \ldots, d_{m-1} \in [1,n]$ such that $\mathrm{FS}(d_n)$ is monochromatic.*

In particular, for any finite coloring of $\mathbb{N}$, there are arbitrarily large finite finite sequences $(c_1, \ldots, c_n)$ in $\mathbb{N}$ such that $\mathrm{FS}(c_1, \ldots, c_n)$ is monochromatic. The main result of this chapter, due to Hindman, allows us to find an *infinite* sequence (c_n) in $\mathbb{N}$ such that $\mathrm{FS}((c_n))$ is monochromatic. Just as the infinite Ramsey theorem cannot just be deduced from its finite form, Hindman's theorem cannot simply be deduced from Folkman's theorem.

Theorem 9.3. *Suppose that $\alpha \in {}^*\mathbb{N}$ is u-idempotent. Then for every $A \subseteq \mathbb{N}$, if $\alpha \in {}^*A$, then A is an* FS*-set.*

Proof. We define by recursion $x_0 < x_1 < \cdots < x_n$ such that $x_F \in A$ and $x_F + \alpha \in {}^*A$ for any $F \subseteq \{0, 1, \ldots, n\}$. Note that, since α is idempotent, we also have that $x_F + \alpha + {}^*\alpha \in {}^{**}A$. Suppose that these have been defined up to n. The statement "there exists $w \in {}^*\mathbb{N}$ such that $w > x_n$ and, for every subset F of $\{0, 1, 2, \ldots, n\}$, $x_F + w \in {}^*A$ and $x_F + w + {}^*\alpha \in {}^{**}A$" holds, so by transfer, there exists $x_{n+1} \in \mathbb{N}$ larger than x_n such that $x_F + x_{n+1} \in A$ and $x_F + x_{n+1} + \alpha \in {}^*A$ for any $F \subseteq \{0, 1, \ldots, n\}$. This concludes the recursive construction.

Corollary 9.4 (Hindman). *For any finite coloring of $\mathbb{N}$, there is a color that is an* FS*-set.*

Proof. Let $\mathbb{N} := C_1 \sqcup \cdots \sqcup C_r$ be a finite coloring of $\mathbb{N}$. Let α be a u-idempotent element of ${}^*\mathbb{N}$ and let i be such that $\alpha \in {}^*C_i$. The result now follows from the previous theorem.

Lemma 9.5. *Suppose that (c_n) is a sequence of distinct elements from $\mathbb{N}$. Then there is an idempotent $\alpha \in {}^*\mathbb{N}$ such that $\alpha \in {}^*\mathrm{FS}((c_n))$.*

Proof. For each m, let U_m be the closed subset ${}^*\mathrm{FS}((c_n)_{n \geq m})$ of ${}^*\mathbb{N}$. By compactness, we have that $S := \bigcap_m U_m$ is a nonempty closed subset of ${}^*\mathbb{N}$. We claim that S is a u-subsemigroup of ${}^*\mathbb{N}$. Indeed, suppose that $\alpha, \beta \in S$ and let $\gamma \in {}^*\mathbb{N}$ such that $\alpha + {}^*\beta \sim \gamma$. We claim that $\gamma \in S$. Fix $m \in \mathbb{N}$. We must show that $\gamma \in {}^*\mathrm{FS}((c_n)_{n \geq m})$ or, equivalently, $\alpha + {}^*\beta \in {}^{**}\mathrm{FS}((c_n)_{n \geq m})$.

Write $\alpha = c_F$ for some hyperfinite $F \subseteq \{n \in {}^*\mathbb{N} \,:\, n \geq m\}$. By transferring the fact that $\beta \in \bigcap_m S_m$, there is hyperfinite $G \subseteq \{n \in {}^{**}\mathbb{N} \,:\, n > \max(F)\}$ such that ${}^*\beta = c_G$. It follows that $\alpha + {}^*\beta = c_F + c_G \in {}^{**}\mathrm{FS}((c_n)_{n \geq m})$.

It follows that S is a nonemtpy closed u-subsemigroup of ${}^*\mathbb{N}$, whence, by Corollary 5.7, there is an idempotent $\alpha \in S$, which, in particular, implies that $\alpha \in {}^*\mathrm{FS}((c_n))$.

Corollary 9.6 (Strong Hindman's Theorem). *Suppose that C is an FS-set and C is partitioned into finitely many pieces $C_1, \ldots, C_n$. Then some C_i is an FS-set.*

Proof. Take (c_n) such that $\mathrm{FS}((c_n)) \subseteq C$. Take $\alpha \in {}^*\mathbb{N}$ u-idempotent such that $\alpha \in {}^*\mathrm{FS}((c_n))$. Then $\alpha \in {}^*C$ as well, whence $\alpha \in {}^*C_i$ for a unique $i = 1, \ldots, n$, and this C_i is itself thus an FS-set.

Exercise 9.7. Let $\mathrm{Idem} := \{\alpha \in {}^*\mathbb{N} \,:\, \alpha \text{ is } u\text{-idempotent}\}$. Prove that $\alpha \in \overline{\mathrm{Idem}}$ if and only if: for every $A \subseteq \mathbb{N}$, if $\alpha \in {}^*A$, then A is an FS-set. Here, $\overline{\mathrm{Idem}}$ denotes the closure of Idem in the u-topology.

9.2 The Milliken-Taylor theorem

We denote by $\mathbb{N}^{[m]}$ the set of subsets of $\mathbb{N}$ of size m. We identify $\mathbb{N}^{[m]}$ with the set of ordered m-tuples of elements of $\mathbb{N}$ increasingly ordered. If F, G are finite subsets of $\mathbb{N}$, we write $F < G$ if either one of them is empty, or they are both nonempty and the maximum of F is smaller than the minimum of G. For $F \subseteq \mathbb{N}$ finite, we will also use the notation x_F for $\sum_{i \in F} x_i$, where we declare $x_F = 0$ when F is empty.

The goal of this section is to prove the following:

Theorem 9.8 (Milliken-Taylor). *For any $m \in \mathbb{N}$ and finite coloring of $\mathbb{N}^{[m]}$, there exists an increasing sequence (x_n) in $\mathbb{N}$ such that the set of elements of the form $\{x_{F_1}, \ldots, x_{F_m}\}$ for finite nonempty subsets $F_1 < \cdots < F_m$ of $\mathbb{N}$ is monochromatic.*

We note that the Milliken-Taylor theorem is a simultaneous generalization of Ramsey's theorem (by taking the finite sets $F_1, \ldots, F_m$ to have cardinality one) and Hindman's theorem (by taking $m = 1$).

The heart of the nonstandard approach is the following:

Proposition 9.9. *Suppose that $m \in \mathbb{N}$ and $\alpha \in {}^*\mathbb{N}$ is u-idempotent. If $A \subset \mathbb{N}^{[m]}$ is such that $\left\{\alpha, {}^*\alpha, \ldots, {}^{*m-1}\alpha\right\} \in {}^{*m}A$, then there exists an increasing sequence (x_n) in $\mathbb{N}$ such that $\{x_{F_1}, x_{F_2}, \ldots, x_{F_m}\} \in A$ for any finite nonempty subsets $F_1 < \cdots < F_m$ of $\mathbb{N}$.*

Proof. We define by recursion an increasing sequence (x_n) such that

$$\left\{x_{F_1}, x_{F_2}, \ldots, x_{F_j}, \alpha, {}^*\alpha, \ldots, {}^{*(m-j-1)}\alpha\right\} \in {}^{*(m-j)}A$$

and

$$\left\{x_{F_1}, x_{F_2}, \ldots, x_{F_{j-1}}, x_{F_j} + \alpha, {}^*\alpha, {}^{**}\alpha, \ldots, {}^{*(m-j)}\alpha\right\} \in {}^{*(m-j+1)}A$$

for every $1 \leq j \leq m$ and finite $F_1 < \cdots < F_j$ such that $F_1, \ldots, F_{j-1}$ are nonempty. It is clear that the sequence (x_n) satisfies the conclusion of the proposition.

Suppose that we have constructed $x_1 < \cdots < x_{n-1}$ satisfying the recursive construction (where of course now $F_1, \ldots, F_j$ are subsets of $\{1, \ldots, n-1\}$). Since α is u-idempotent, we also have, for any $1 \leq j \leq m$ and $F_1, \ldots, F_j$ as above, that

$$\left\{x_{F_1}, x_{F_2}, \ldots, x_{F_{j-1}}, x_{F_j} + \alpha + {}^*\alpha, {}^{**}\alpha, \ldots, {}^{*(m-j+1)}\alpha\right\} \in {}^{*(m-j+2)}A.$$

Therefore, by transfer there exists $x_n > x_{n-1}$ such that

$$\left\{x_{F_1}, x_{F_2}, \ldots, x_{F_{j-1}}, x_{F_j} + x_n, \alpha, {}^*\alpha, \ldots, {}^{*(m-j-1)}\alpha\right\} \in {}^{*(m-j)}A$$

and

$$\left\{x_{F_1}, x_{F_2}, \ldots, x_{F_{j-1}}, x_{F_j} + x_n + \alpha, {}^*\alpha, \ldots, {}^{*(m-j)}\alpha\right\} \in {}^{*(m-j+1)}A$$

for any $1 \leq j \leq m$ and $F_1 < \cdots < F_j$ contained in $\{1, 2, \ldots, n-1\}$ such that $F_1, \ldots, F_{j-1}$ are nonempty. This concludes the recursive construction and the proof of the proposition.

Theorem 9.8 follows immediately from Proposition 9.9. Indeed, suppose $\mathbb{N}^{[m]} = A_1 \sqcup \cdots \sqcup A_r$ is a partition of $\mathbb{N}^{[m]}$. Fix $\alpha \in {}^*\mathbb{N}$ a u-idempotent. Let $i \in \{1, \ldots, r\}$ be such that $\left\{\alpha, {}^*\alpha, \ldots, {}^{*m-1}\alpha\right\} \in {}^{*m}A_i$. Then A_i is the desired color.

Observe now that if $c \in \mathbb{N}$ and $\alpha \sim \alpha + {}^*\alpha$, then $c\alpha \sim c\alpha + c^*\alpha$. Hence the same proofs as above shows the following slight strengthening of the Milliken-Taylor theorem.

Theorem 9.10. *Suppose that $m \in \mathbb{N}$, $c_1, \ldots, c_m \in \mathbb{N}$, and $\alpha \in {}^*\mathbb{N}$ is u-idempotent. If $A \subset \mathbb{N}^{[m]}$ is such that $\left\{\alpha, {}^*\alpha, \ldots, {}^{*(m-1)}\alpha\right\} \in {}^{*m}A$, then there exists an increasing sequence (x_n) in $\mathbb{N}$ such that $\{c_1 x_{F_1}, \ldots, c_m x_{F_m}\} \in A$ for any finite nonempty subsets $F_1 < \cdots < F_m$ of $\mathbb{N}$.*

Corollary 9.11. *For any $m \in \mathbb{N}$, $c_1, \ldots, c_m \in \mathbb{N}$, and finite coloring of $\mathbb{N}^{[m]}$, there exists an increasing sequence (x_n) in $\mathbb{N}$ such that the set of elements of the form $\{c_1 x_{F_1}, \ldots, c_m x_{F_m}\} \in A$ for finite nonempty subsets $F_1 < \cdots < F_m$ of $\mathbb{N}$ is monochromatic.*

From the previous corollary, it is straightforward to deduce an "additive" version:

Corollary 9.12. *For any $m \in \mathbb{N}$, $c_1, \ldots, c_m \in \mathbb{N}$, and finite coloring of $\mathbb{N}$, there exists an increasing sequence (x_n) in $\mathbb{N}$ such that the set of elements of the form $c_1 x_{F_1} + \cdots + c_m x_{F_m} \in A$ for finite nonempty subsets $F_1 < \cdots < F_m$ of $\mathbb{N}$ is monochromatic.*

9.3 Gowers' theorem

Definition 9.13. For $k \in \mathbb{N}$, we let FIN_k denotes the set of functions $b : \mathbb{N} \to \{0, 1, \ldots, k\}$ with $\mathrm{Supp}(b)$ finite and such that k belongs to the range of b. Here, $\mathrm{Supp}(b) := \{n \in \mathbb{N} \ : \ b(n) \neq 0\}$, the *support of* b. We extend the definition of FIN_k to $k = 0$ by setting FIN_0 to consist of the function on $\mathbb{N}$ that is identically 0.

Note that, after identifying a subset of $\mathbb{N}$ with its characteristic function, FIN_1 is simply the set of nonempty finite subsets of $\mathbb{N}$.

We endow FIN_k with a partial semigroup operation $(b_0, b_1) \mapsto b_0 + b_1$ which is defined only when $\mathrm{Supp}(b_0) < \mathrm{Supp}(b_1)$.

By transfer, ${}^*\mathrm{FIN}_k$ is the set of internal functions $b : {}^*\mathbb{N} \to \{0, 1, \ldots, k\}$ with hyperfinite support that have k in their range. The partial semigroup operation on FIN_k extends also to ${}^*\mathrm{FIN}_k$. We say that $\alpha \in {}^*\mathrm{FIN}_k$ is *cofinite* if its support is disjoint from $\mathbb{N}$. Thus, if $\alpha, \beta \in {}^*\mathrm{FIN}_k$ are cofinite and $i < j$, then the sum ${}^{*i}\alpha + {}^{*j}\beta$ exists.

Gowers' original theorem considers the *tetris operation* $T : \mathrm{FIN}_k \to \mathrm{FIN}_{k-1}$ given by $T(b)(n) := \max\{b(n) - 1, 0\}$. In this section, we prove a more general version of Gowers' theorem by considering a wider variety of functions $\mathrm{FIN}_k \to \mathrm{FIN}_j$ for $j \leq k$. First, for $k \in \mathbb{N}$, by a *regressive map on* k or *generalized tetris operation*, we mean a nondecreasing surjection $f : [0, k] \to [0, f(k)]$. Given a regressive map f on k, one can define a corresponding operation $f : \mathrm{FIN}_k \to \mathrm{FIN}_{f(k)}$ by setting $f(b) := f \circ b$. Note also that if $l \leq k$, then $f|_{[0,l]}$ is a regressive map on l, whence we can also consider $f : \mathrm{FIN}_l \to \mathrm{FIN}_{f(l)}$.

Given $n \in \mathbb{N}$, we set $\mathrm{FIN}_{[0,n]} := \bigcup_{k=0}^{n} \mathrm{FIN}_k$. Note that $\mathrm{FIN}_{[0,n]}$ is also a partial semigroup given by pointwise addition and defined on pairs of functions with disjoint supports. If f is a regressive map on n, then as we already recalled, $f|_{[0,k]}$ is a regressive map on k for $1 \leq k \leq n$, whence f yields a function $f : \mathrm{FIN}_{[0,n]} \to \mathrm{FIN}_{[0,f(n)]}$.

Given a regressive map f on n, we get the nonstandard extension $f : {}^*\mathrm{FIN}_n \to {}^*\mathrm{FIN}_{f(n)}$ and $f : {}^*\mathrm{FIN}_{[0,n]} \to {}^*\mathrm{FIN}_{[0,f(n)]}$. In addition, if $\alpha, \beta \in {}^*\mathrm{FIN}_{[0,n]}$ are cofinite and $i < j$, then ${}^{*i}\alpha + {}^{*j}\beta$ exists and $f({}^{*i}\alpha + {}^{*j}\beta) = f({}^{*i}\alpha) + f({}^{*j}\beta)$.

If $\alpha_k \in {}^*\mathrm{FIN}_k$ for $k = 1, \ldots, n$, we say that $\alpha_1, \ldots, \alpha_n$ is *coherent* if $f(\alpha_k) \sim \alpha_{f(k)}$ for all $k = 1, \ldots, n$ and all regressive maps f on n. We let Z denote the compact u-semigroup consisting of cofinite coherent tuples. We note that Z is nonempty. Indeed, let $\alpha_1 \in {}^*\mathrm{FIN}_1$ be any cofinite element. For $k = 2, \ldots, n$, let $\alpha_k \in {}^*\mathrm{FIN}_k$ have the same support as α_1 and take only the values 0 and k. It is immediate that $(\alpha_1, \ldots, \alpha_n) \in Z$.

Finally, we introduce some convenient notation. Given $\alpha_0, \alpha_1, \ldots, \alpha_j \in {}^*\mathrm{FIN}_{[0,n]}$ and $j \in \mathbb{N}$, we set

$$\bigoplus_{i=0}^{j}\alpha_i := \alpha_0 +^* \alpha_1 + \cdots +^{*j} \alpha_j.$$

Thus, if each α_i is cofinite and f is a regressive map on n, we have the convenient equation

$$f(\bigoplus_{i=1}^{j}\alpha_i) = \bigoplus_{i=1}^{j} f(\alpha_i).$$

Lemma 9.14. *Fix $n \in \mathbb{N}$. Then, for $k = 1, \ldots, n$, there exist cofinite u-idempotents $\alpha_k \in {}^*\mathrm{FIN}_k$ such that:*

1. *$\alpha_1, \ldots, \alpha_n$ is a coherent sequence, and*
2. *$\alpha_j +^* \alpha_k \sim \alpha_k +^* \alpha_j \sim \alpha_k$ for every $1 \le j \le k \le n$.*

Proof. We define, by recursion on $k = 1, 2, \ldots, n$, a sequence of u-idempotents

$$\alpha^{(k)} = (\alpha_1^{(k)}, \ldots, \alpha_n^{(k)}) \in Z$$

such that, for $1 \le i \le j \le k \le n$, one has that

(a) $\alpha_i^{(k)} \sim \alpha_i^{(j)}$,
(b) $\alpha_j^{(k)} +^* \alpha_i^{(k)} \sim \alpha_j^{(k)}$.

To begin the construction, let $\alpha^{(1)}$ be any idempotent element of Z. Now suppose now that $k < n$ and $\alpha^{(1)}, \ldots, \alpha^{(k)}$ have been constructed satisfying (a) and (b). Consider the closed u-semigroup Z_k consisting of sequences $\beta = (\beta_1, \ldots, \beta_k) \in Z$ such that:

(i) $\beta_j \sim \alpha_j^{(k)}$ for $1 \le j \le k$, and
(ii) $\beta_j +^* \beta_i \sim \beta_j$ for $1 \le i < j \le n$ and $1 \le i \le k$.

We claim that Z_k is nonempty. Indeed, we claim it contains the sequence $\beta = (\beta_1, \ldots, \beta_k)$, where $\beta_j \in {}^*\mathrm{FIN}_j$ is such that

$$\beta_j \sim \bigoplus_{i=0}^{j-1}\alpha_{j-i}^{(k)}.$$

To see that β is coherent, fix a regressive map f on n. For a given $j \in [1,k]$, we have that

$$f(\beta_j) \sim \bigoplus_{i=0}^{j-1} f(\alpha_{j-i}^{(k)}) \sim \bigoplus_{i=0}^{j-1}\alpha_{f(j-i)}^{(k)} \sim \bigoplus_{i=0}^{f(j)-1}\alpha_{f(j)-i}^{(k)} \sim \beta_{f(j)}.$$

The second equivalence uses that $\alpha^{(k)}$ is coherent, while the third equivalence uses that f is a regressive map and that $\alpha^{(k)}$ is a u-idempotent. Next observe that, since $\alpha^{(k)}$ satisfies (b), we have that $\beta_j \sim \bigoplus_{i=0}^{j-k}\alpha_{j-i}^{(k)}$ for $j = 1, \ldots, n$, and, moreover, that $\beta_j \sim \alpha_j^{(k)}$ for $j = 1, 2, \ldots, k$. Thus, if $1 \le i < j \le n$ and $1 \le i \le k$, it follows that

$$\beta_j +^* \beta_i \sim \bigoplus_{i=0}^{j-k}\alpha_{j-i}^{(k)} +^{*k} \alpha_i \sim \bigoplus_{i=0}^{j-k}\alpha_{j-i}^{(k)},$$

where the last equivalence follows from (b). This concludes the proof that β belongs to Z_k.

Since Z_k is a nonempty closed u-semigroup, it contains an idempotent $\alpha^{(k+1)}$. It is clear that $\alpha^{(k+1)}$ satisfies (a) and (b). This concludes the recursive construction.

For $k = 1, \ldots, n$, we fix $\alpha_k \in {}^*\mathrm{FIN}_k$ such that

$$\alpha_k \sim \bigoplus_{i=1}^{k}\alpha_i^{(i)}.$$

We claim that $\alpha_1,\dots,\alpha_n$ are as in the conclusion of the lemma. Towards this end, first fix a regressive map f on n. We then have that

$$f(\alpha_j)\sim\bigoplus_{i=1}^{k}f(\alpha_i^{(i)})\sim\bigoplus_{i=1}^{k}\alpha_{f(i)}^{(i)}\sim\bigoplus_{i=1}^{f(k)}\alpha_i^{(i)}\sim\alpha_{f(j)},$$

where the second to last step uses the fact that f is a regressive map, that that the $\alpha_i^{(k)}$'s are u-idempotent, and that (a) holds. We thus have that $\alpha_1,\dots,\alpha_n$ are coherent. We now show that (2) holds. Fix $1\le j\le k\le n$. We then have

$$\alpha_k+{}^*\alpha_j\sim\bigoplus_{i=1}^{k}\alpha_i^{(i)}+\bigoplus_{i=1}^{j}\alpha_i^{(i)}\sim\bigoplus_{i=1}^{k}\alpha_i^{(i)}\sim\alpha_k,$$

where the second to last equivalence repeatedly uses the fact that $\alpha_k^{(k)}+{}^*\alpha_i^{(i)}\sim\alpha_k^{(k)}$ for $1\le i\le k$. A similar computation shows that $\alpha_j+{}^*\alpha_k\sim\alpha_k$, establishing (2) and finishing the proof of the lemma.

We say that a sequence (x_i) in FIN_n is a *block sequence* if $\mathrm{Supp}(x_i)<\mathrm{Supp}(x_j)$ for $i<j$.

Theorem 9.15. *Suppose that $\alpha_k\in{}^*\mathrm{FIN}_k$ for $k=1,2,\dots,n$ are as in the previous lemma. For $k=1,\dots,n$, suppose that $A_k\subset\mathrm{FIN}_k$ is such that $\alpha_k\in{}^*A_k$. Then there exists a block sequence (x_i) in FIN_n such that, for every finite sequence $f_1,\dots,f_\ell$ of regressive maps on n, we have $f_1(x_1)+\cdots+f_\ell(x_\ell)\in A_{\max(f_1(n),\dots,f_l(n))}$.*

Proof. By recursion on d, we define a block sequence (x_d) in FIN_n such that, for every sequence $f_1,\dots,f_{d+1}$ of regressive maps n, we have

$$f_1(x_1)+\cdots+f_d(x_d)\in A_{\max(f_1(n),\dots,f_d(n))}$$

and

$$f_1(x_1)+\cdots+f_d(x_d)+f_{d+1}(\alpha_n)\in{}^*A_{\max(f_1(n),\dots,f_{d+1}(n))}.$$

Suppose that $x_1,\dots,x_d$ has been constructed satisfying the displayed properties. Suppose that $f_1,\dots,f_{d+2}$ are regressive maps on n. Then since

$$f_{d+1}(\alpha_n)+f_{d+2}({}^*\alpha_n)\sim\alpha_{f_{d+1}(n)}+{}^*\alpha_{f_{d+2}(n)}\sim\alpha_{\max(f_{d+1}(n),f_{d+2}(n))}\sim f_{d+p}(\alpha_n),$$

where $p\in\{1,2\}$ is such that $\max(f_{d+1}(n),f_{d+2}(n))=f_{d+p}(n)$, the inductive hypothesis allows us to conclude that

$$f_1(x_1)+\cdots+f_{d-1}(x_d)+f_{d+1}(\alpha_n)+f_{d+2}({}^*\alpha_n)\in{}^{**}A_{\max(f_1(n),\dots,f_{d+2}(n))}.$$

Therefore, by transfer, we obtain $x_{d+1}\in\mathrm{FIN}_n$ such that $\mathrm{Supp}(x_{d+1})>\mathrm{Supp}(x_d)$, and, for any sequence $f_1,\dots,f_{d+2}$ of regressive maps on n, we have that

$$f_1(x_1)+\cdots+f_d(x_{d+1})\in A_{\max(f_1(n),\dots,f_{d+1}(n))}$$

and

$$f_1(x_1)+\cdots+f_{d+1}(x_{d+1})+f_{d+2}(\alpha_n)\in{}^*A_{\max(f_1(n),\dots,f_{d+2}(n))}.$$

This concludes the recursive construction.

Corollary 9.16 (Generalized Gowers). *For any finite coloring of FIN_n, there exists a block sequence (x_d) in FIN_n such that the set of elements of the form $f_1(x_1)+\cdots+f_l(x_l)$ for $l\in\mathbb{N}$ and regressive maps $f_1,\dots,f_\ell$ on n such that $n=\max(f_1(n),\dots,f_l(n))$ is monochromatic.*

Proof. If $\mathrm{FIN}_n=B_1\sqcup\cdots\sqcup B_r$ is a partition of FIN_n, apply the previous theorem with $A_n:=B_i$ where $\alpha_n\in{}^*B_i$.

Gowers' original theorem is a special case of the previous corollary by taking each f_i to be an iterate of the tetris operation. One can also obtain a common generalization of Gowers' theorem and the Milliken-Taylor theorem. We let $\mathrm{FIN}_k^{[m]}$ be the set of m-tuples $(x_1,\dots,x_m)$ in FIN_k such that $\mathrm{Supp}(x_i)<\mathrm{Supp}(x_j)$ for $1\le i<j\le m$. Suppose that (x_d) is a sequence in FIN_n. Suppose that $F=\{a_1,\dots,a_r\}$ is a finite nonempty subset of $\mathbb{N}$. We let $\mathscr{S}(F,k)$ be the set of tuples $f=(f_j)_{j\in F}$ such that $f_j:\{0,1,\dots,n\}\to\{0,1,\dots,k_j\}$ is a nondecreasnig surjection and $\max\{k_j:j\in F\}=k$. For such an element f we let x_f be

the $f_{a_1}(x_{a_1}) + \cdots + f_{a_r}(x_{a_r})$. When F is empty, by convention we let $\mathscr{S}(F,k)$ contain a single element $f = \varnothing$, and in such case $x_f = 0$.

Theorem 9.17. *Let $\alpha_1, \ldots, \alpha_n$ be as in Lemma 9.14. Suppose that $A_k \subset \mathrm{FIN}_k^{[m]}$ for $k = 1, 2 \ldots, n$ is such that $\left(\alpha_k, {}^*\alpha_k, \ldots, {}^{*(m-1)}\alpha_k\right) \in {}^{*m}A_k$. Then there exists a block sequence (x_d) in FIN_n such that, given $k \in \{1, \ldots, n\}$, nonempty finite subsets $F_1 < \ldots < F_m$ of $\mathbb{N}$, and $f_i \in \mathscr{S}(F_i, k)$ for $i = 1, \ldots, m$, we have that $\{x_{f_1}, \ldots, x_{f_m}\} \in A_k$.*

Proof. We define by recursion a block sequence (x_d) in FIN_n such that, for all $k \in \{1, \ldots, n\}$, all $1 \leq j \leq m$, all finite $F_1, \ldots, F_j \subseteq \mathbb{N}$ with $F_1 < \cdots < F_j$ and $F_1, \ldots, F_{j-1}$ nonempty, and all $f_i \in \mathscr{S}(F_i, k)$, we have

$$\left\{x_{f_1}, x_{f_2}, \ldots, x_{f_j}, \alpha_k, {}^*\alpha_k, \ldots, {}^{*(m-j-1)}\alpha_k\right\} \in {}^{*(m-j)}A_k$$

and

$$\left\{x_{f_1}, x_{f_2}, \ldots, x_{f_{j-1}}, x_{f_j} + \alpha_k, {}^*\alpha, {}^{**}\alpha_k, \ldots, {}^{*(m-j)}\alpha_k\right\} \in {}^{*(m-j+1)}A_k.$$

It is clear that the sequence (x_d) is as desired.

Suppose that $x_1, \ldots, x_d$ have been constructed satisfying the above assumption. From the properties of the sequence $\alpha_1, \ldots, \alpha_n$, we see that the second condition also implies, for all $1 \leq s \leq k$:

$$\{x_{f_1}, \ldots, x_{f_j} + \alpha_k + {}^*\alpha_s, {}^{**}\alpha_k, \ldots, {}^{*(m-j+1)}\alpha_k\} \in {}^{*(m-j+2)}A_k$$

and

$$\{x_{f_1}, \ldots, x_{f_j} + \alpha_s + {}^*\alpha_k, {}^{**}\alpha_k, \ldots, {}^{**(m-j+1)}\alpha_k\} \in {}^{*(m-j+2)}A_k.$$

It follows from transfer that we can find x_{d+1} with $\mathrm{Supp}(x_{d+1}) > \mathrm{Supp}(x_d)$ as desired.

Chapter 10
Partition regularity of equations

10.1 Rado's theorem

Definition 10.1. A polynomial $F(X_1,\ldots,X_n)$ is said to be *partition regular* if, for every finite partition $\mathbb{N}=C_1\sqcup\cdots\sqcup C_r$ of the natural numbers, there exists $i\in\{1,\ldots,r\}$ and there exist $x_1,\ldots,x_n\in C_i$ such that $F(x_1,\ldots,x_n)=0$. If, in addition, we can require $x_1,\ldots,x_n$ to be distinct, then we say that the equation is *injectively partition regular*.

Proposition 10.2. *Given an equation $F(X_1,\ldots,X_n)=0$, the following are equivalent:*

1. *$F(X_1,\ldots,X_n)=0$ is [injectively] partition regular;*
2. *there exists [distinct] $\alpha_1,\ldots,\alpha_n\in{}^*\mathbb{N}$ such that $\alpha_i\sim\alpha_j$ for all $i,j=1,\ldots,n$ and such that ${}^*F(\alpha_1,\ldots,\alpha_n)=0$;*
3. *there is an ultrafilter $\mathscr{U}$ on $\mathbb{N}$ such that, for every $A\in\mathscr{U}$, there exists [distinct] $x_1,\ldots,x_n\in A$ such that $F(x_1,\ldots,x_n)=0$.*

Proof. (1)⇒(2) Suppose that $F(X_1,\ldots,X_n)=0$ is [injectively] partition regular. Given $A\subseteq\mathbb{N}$, consider the set

$$X_A:=\{(\alpha_1,\ldots,\alpha_n)\in{}^*\mathbb{N}^n\ :\ [\bigwedge_{i\neq j}\alpha_i\neq\alpha_j\wedge]\bigwedge_{i,j}(\alpha_i\in{}^*A\leftrightarrow\alpha_j\in{}^*A)\wedge{}^*F(\alpha_1,\ldots,\alpha_n)=0\}.$$

Observe that the family $(X_A)_{A\subseteq\mathbb{N}}$ has the finite intersection property. Indeed, given $A_1,\ldots,A_m\subseteq\mathbb{N}$, let $C_1,\ldots,C_k$ be the atoms of the boolean algebra generated by $A_1,\ldots,A_m$. Since the equation $F(X_1,\ldots,X_n)=0$ is [injectively] partition regular, there is $i\in\{1,\ldots,k\}$ and [distinct] $x_1,\ldots,x_n\in C_i$ such that $F(x_1,\ldots,x_n)=0$; it follows that $(x_1,\ldots,x_n)\in\bigcap_{i=1}^m X_{A_i}$. Thus, by saturation, there is $(\alpha_1,\ldots,\alpha_n)\in\bigcap_{A\subseteq\mathbb{N}}X_A$. These $\alpha_1,\ldots,\alpha_n$ are as desired.

(2)⇒(3) Suppose that $\alpha_1,\ldots,\alpha_n$ are as in (2). Consider $\mathscr{U}:=\mathscr{U}_{\alpha_1}$. Observe that $\mathscr{U}:=\mathscr{U}_{\alpha_i}$ for $i=1,2,\ldots,n$. Suppose that $A\in\mathscr{U}$. Then the statement "there exist $i\in\{1,2,\ldots,r\}$ and [distinct] $x_1,\ldots,x_n\in{}^*A$ such that ${}^*F(x_1,\ldots,x_n)=0$" holds in the nonstandard universe, as witnessed by $\alpha_1,\ldots,\alpha_n$; the desired conclusion follows from transfer.

(3)⇒(1) This is trivial.

In the context of part (3) of the previous proposition, we say that $\mathscr{U}$ *witnesses the [injective] partition regularity* of $F(X_1,\ldots,X_n)=0$.

We use the above characterization of partition regularity to prove the following version of the classical theorem of Rado:

Theorem 10.3. *Suppose that $k>2$ and $c_1,\ldots,c_k\in\mathbb{Z}$ are such that $c_1+\cdots+c_k=0$. Then the equation $c_1X_1+\cdots+c_kX_k=0$ is injectively partition regular.*

Indeed, we will prove a strengthening of Rado's theorem below. First, given a polynomial $P(X):=\sum_{j=0}^n b_jX^j\in\mathbb{Z}[X]$ and $\xi\in{}^*\mathbb{Z}$, set $\tilde{P}(\xi):=\sum_{j=0}^n b_j{}^{j*}\xi\in{}^{(j+1)*}\mathbb{Z}$. We note the following corollary of Proposition 10.2.

Corollary 10.4. *Suppose that $c_1,\ldots,c_k\in\mathbb{Z}$ are such that there exist [distinct] polynomials $P_1(X),\ldots,P_k(X)\in\mathbb{Z}[X]$ and $\xi,\eta\in{}^*\mathbb{N}$ for which*

1. *$c_1P_1(X)+\cdots+c_kP_k(X)=0$, and*

2. $\tilde{P}_i(\xi) \sim \eta$ *for each* $i = 1, \ldots, k$.

Then $\mathscr{U}_\eta$ *witnesses that* $c_1X_1 + \cdots + c_kX_k = 0$ *is [injectively] partition regular.*

Proof. For each $i = 1, \ldots, k$, let $\alpha_i := \tilde{P}_i(\xi)$; by assumption, for each i we have $\mathscr{U}_{\alpha_i} = \mathscr{U}_\eta$. It is also clear that $c_1\alpha_1 + \cdots + c_k\alpha_k = 0$. By the previous proposition, we have that $\mathscr{U}_\eta$ witnesses the partition regularity of $c_1X_1 + \cdots + c_kX_k = 0$.

Suppose in addition that the P_i's are distinct; to conclude injective partition regularity, we must show that the α_i's are distinct. Suppose that $\alpha_i = \alpha_j$, that is, $\tilde{P}_i(\xi) = \tilde{P}_j(\xi)$. Write $P_i(X) := \sum_{l=0}^m r_l X^l$ and $P_j(X) = \sum_{l=0}^m s_l X^l$. We then have that $(r_m - s_m)^{m*}\xi = -\sum_{l=0}^{m-1}(r_l - s_l)^{l*}\xi$. The only way that this is possible is that $r_m = s_m = 0$; continuing inductively in this manner, we see that $P_i = P_j$, yielding the desired contradiction.

In light of the previous corollary, it will be useful to find a standard condition on a family of polynomials $P_1, \ldots, P_k \in \mathbb{Z}[X]$ such that, at least for idempotent $\xi \in {}^*\mathbb{N}$, we have that all $\tilde{P}_i(\xi)$'s are u-equivalent. The next definition captures such a condition.

Definition 10.5. We define the equivalence relation $\approx_u$ on finite strings of integers to be the smallest equivalence relation satisfying the following three properties:

- $\emptyset \approx_u \langle 0 \rangle$;
- If $a \in \mathbb{Z}$, then $\langle a \rangle \approx_u \langle a, a \rangle$;
- If $\sigma \approx_u \sigma'$ and $\tau \approx_u \tau'$, then concatenations $\sigma\tau \approx_u \sigma'\tau'$.

If $P, Q \in \mathbb{Z}[X]$ are polynomials, then we write $P \approx_u Q$ to mean that their strings of coefficients are u-equivalent.

Lemma 10.6. *Let* $P, Q \in \mathbb{Z}[X]$ *have positive leading coefficient. If* $P \approx_u Q$*, then for every idempotent* $\xi \in {}^*\mathbb{N}$*, we have* $\tilde{P}(\xi) \sim \tilde{Q}(\xi)$.

Proof. Fix an idempotent $\xi \in {}^*\mathbb{N}$. The lemma follows from the following facts:

- $\sum_{j=0}^m a_j{}^{j*}\xi \sim \sum_{j=0}^i a_j{}^{j*}\xi + a_i{}^{(i+1)*}\xi + \sum_{j=i+1}^m a_j{}^{(j+1)*}\xi$;
- If $\sum_{j=0}^m a_j{}^{j*}\xi \sim \sum_{j=0}^{m'} a_j'^{j*}\xi$ and $\sum_{j=0}^n b_j{}^{j*}\xi \sim \sum_{j=0}^{n'} b_j'^{j*}\xi$, then

$$\sum_{j=0}^m a_j{}^{j*}\xi + \sum_{j=0}^n b_j{}^{(j+m)*}\xi \sim \sum_{j=0}^{m'} a_j'^{j*}\xi + \sum_{j=0}^{n'} b_j'^{(j+m')*}\xi.$$

We should mention that the converse of the previous lemma is true in an even stronger form, namely that if $\tilde{P}(\xi) \sim \tilde{Q}(\xi)$ for *some* idempotent $\xi \in {}^*\mathbb{N}$, then $P \approx_u Q$. This follows from [38, Theorem T].

We can now give the nonstandard proof of the above mentioned version of Rado's theorem. In fact, we prove the more precise statement:

Theorem 10.7. *Suppose that* $k > 2$ *and* $c_1, \ldots, c_k \in \mathbb{Z}$ *are such that* $c_1 + \cdots + c_k = 0$*. Then there exists* $a_0 \ldots, a_{k-2} \in \mathbb{N}$ *such that, for every idempotent ultrafilter* $\mathscr{U}$*, we have that* $a_0\mathscr{U} \oplus \cdots \oplus a_{k-2}\mathscr{U}$ *witnesses the injective partition regularity of the equation* $c_1X_1 + \cdots + c_kX_k = 0$.

Proof. Without loss of generality, we will assume that $c_1 \geq c_2 \geq \cdots \geq c_k$. By Corollary 10.4 and Lemma 10.6, we need to find $a_0, \ldots, a_{k-2} \in \mathbb{N}$ and distinct $P_1(X), \ldots, P_k(X) \in \mathbb{N}_0[X]$ such that $c_1P_1(X) + \cdots + c_kP_k(X) = 0$ and such that $P_i(X) \approx_u \sum_{j=0}^{k-2} a_jX^j$ for each $i = 1, \ldots, k$. For appropriate $a_0, \ldots, a_{k-2}$, the following polynomials will be as needed:

- $P_1(X) := \sum_{j=0}^{k-2} a_jX^j + a_{k-2}X^{k-1}$;
- $P_i(X) := \sum_{j=0}^{k-i-1} a_jX^j + \sum_{j=k-i+1}^{k-1} a_{j-1}X^j$ for $2 \leq i \leq k-1$,
- $P_k(X) := a_0 + \sum_{j=1}^{k-1} a_{j-1}X^j$.

It is straightforward to check that $P_i(X) \approx_u \sum_{j=0}^{k-2} a_j X^j$ for each $i = 1,\dots,k$. Furthermore, since $a_0,\dots,a_{k-2}$ are nonzero, the polynomials $P_1(X),\dots,P_k(X)$ are mutually distinct. It remains to show that there are $a_0,\dots,a_{k-2} \in \mathbb{N}$ for which the $P_i(X)$'s are distinct and for which $c_1P_1(X)+\cdots+c_kP_k(X) = 0$. Since $c_1+\cdots+c_k = 0$, the constant and leading terms of $c_1P_1(X)+\cdots+c_kP_k(X)$ are zero. So the equation $c_1P_1(X)+\cdots+c_kP_k(X) = 0$ is equivalent to the system of equations $(c_1+\cdots+c_{k-i})\cdot a_{i-1}+(c_{k-i+2}+\cdots+c_k)\cdot a_{i-2}$ for $i = 1,2,\dots,k-1$. One can then easily define recursively elements $a_0,a_1,\dots,a_{k-2}$ satisfying all these equations.

10.2 Non-partition regularity of some equations

Nonstandard methods have also played a role in establishing the non-partition regularity of equations. We present here the simplest examples of this type of result.

Theorem 10.8 ([14]). *Let $P(x_1,\dots,x_h) := a_1x_1^{n_1}+\cdots+a_hx_h^{n_h}$, with $n_1 < \cdots < n_h$, where each $a_i \in \mathbb{Z}$ is odd and h is odd. Then $P(x_1,\dots,x_h) = 0$ is not partition regular.*

Proof. Suppose, towards a contradiction, that there are u-equivalent $\xi_1,\dots,\xi_h \in {}^*\mathbb{N}$ such that $P(\xi_1,\dots,\xi_h) = 0$. Let $f,g : \mathbb{N} \to \mathbb{N}$ be such that, for all $x \in \mathbb{N}$, we have $x = 2^{f(x)}g(x)$ with $g(x)$ odd. Then, for each $i,j = 1,\dots,h$, we have $f(\xi_i) \sim f(\xi_j)$. Set $\nu_i := f(\xi_i)$ and $\zeta_i := g(\xi_i)$.

We next claim that, for distinct $i,j \in \{1,\dots,h\}$, we have $n_i\nu_i \neq n_j\nu_j$. Indeed, if $n_i\nu_i = n_j\nu_j$, then $n_i\nu_i = n_j\nu_j \sim n_j\nu_i$, whence $n_i\nu_i = n_j\nu_i$ by Proposition 3.5 and hence $\nu_i = 0$. Since the ν_k's are all u-equivalent, it follows that $\nu_k = 0$ for each k, whence each ξ_i is odd. But then since h is odd, we have that $P(\xi_1,\dots,\xi_h)$ is odd, contradicting that $P(\xi_1,\dots,\xi_h) = 0$.

By the previous paragraph, we can let $i \in \{1,\dots,k\}$ be the unique index for which $n_i\nu_i < n_j\nu_j$ for all $j = 1,\dots,k$. By factoring out $2^{n_i\nu_i}$ from the equation $P(\xi_1,\dots,\xi_h) = 0$, we obtain the contradiction

$$0 = a_i\zeta_i^{n_i} + \sum_{j\neq i} a_j 2^{n_j\nu_j - n_i\nu_i}\zeta_j^{n_j} \equiv 1 \mod 2.$$

From the previous theorem, we see that many "Fermat-like" equations are not partition regular:

Corollary 10.9. *Suppose that k,m,n are distinct positive natural numbers. Then the equation $x^m+y^n = z^k$ is not partition regular.*

In [14], the previous corollary is extended to allow m and n to be equal, in which case the equations are shown to be not partition regular (as long as, in the case when $m = n = k-1$, one excludes the trivial solution $x = y = z = 2$). The methods are similar to the previous proof. To further illustrate the methods, we conclude by treating two simple cases.

Theorem 10.10. *The equation $x^2+y^2 = z$ is not partition regular.*

Proof. Notice first that the given equation does not have constant solutions. Then suppose, towards a contradiction, that α,β,γ are infinite hypernatural numbers such that $\alpha \sim \beta \sim \gamma$ and $\alpha^2+\beta^2 = \gamma$. Notice that α,β,γ are even numbers, since they cannot all be odd. Then we can write

$$\alpha = 2^a\alpha_1, \quad \beta = 2^b\beta_1, \quad \gamma = 2^c\gamma_1,$$

with positive $a \sim b \sim c$ and with $\alpha_1 \sim \beta_1 \sim \gamma_1$ odd.

Case 1: $a < b$. We then have that $2^{2a}(\alpha_1^2+2^{2b-2a}\beta_1^2) = 2^c\gamma_1$. Since $\alpha_1^2+2^{2b-2a}\beta_1^2$ and γ_1 are odd, it follows that $2a = c \sim a$, whence $2a = a$ by Proposition 3.5 and hence $a = 0$, a contradiction. If $b > a$ the proof is entirely similar.

Case 2: $a = b$. In this case we have the equality $2^{2a}(\alpha_1^2+\beta_1^2) = 2^c\gamma_1$. Since α_1,β_1 are odd, $\alpha_1^2+\beta_1^2 \equiv 2 \mod 4$, and so $2^c\gamma_1 = 2^{2a+1}\alpha_2$ for a suitable odd number α_2. But then $2a+1 = c \sim a$, whence $2a+1 = a$, and we again obtain a contradiction.

The following result was first proven by Csikivari, Gyarmati, and Sarkozy in [10].

Theorem 10.11. *If one excludes the trivial solution $x = y = z = 2$, then the equation $x+y = z^2$ is not partition regular.*

Proof. Suppose, towards a contradiction, that $\alpha,\beta,\gamma \in {}^*\mathbb{N}\setminus\mathbb{N}$ are such that $\alpha \sim \beta \sim \gamma$ and $\alpha+\beta = \gamma^2$. Since $\alpha \sim \beta \sim \gamma$, there is $i \in \{0,1,2,3,4\}$ such that $\alpha \equiv \beta \equiv \gamma \equiv i \mod 5$. Note that since $\alpha+\beta = \gamma^2$, we have that $i = 0$ or 2. Write

$$\alpha = 5^a\alpha_1 + i, \quad \beta = 5^b\beta_1 + i, \quad \gamma = 5^c\gamma_1 + i,$$

with $a,b,c \in {}^*\mathbb{N}_+$ and $\alpha_1,\beta_1,\gamma_1 \in {}^*\mathbb{N}$ not divisible by 5.

Next note that $\alpha_1 \sim \beta_1 \sim \gamma_1$ and $a \sim b \sim c$. Indeed, if $f : \mathbb{N} \to \mathbb{N}$ is the function such that $f(n)$ is the unique $k \not\equiv 0 \mod 5$ such that $n = 5^hk + j$ with $h > 0$ and $0 \le j \le 4$, then $\alpha_1 = f(\alpha)$, $\beta_1 = f(\beta)$, and $\gamma_1 = f(\gamma)$, whence $\alpha_1 \sim \beta_1 \sim \gamma_1$ by Proposition 3.5. The proof that $a \sim b \sim c$ is similar. As in the previous paragraph, we may take $j \in \{1,2,3,4\}$ such that $\alpha_1 \equiv \beta_1 \equiv \gamma_1 \equiv j \mod 5$.

Case 1: $i = 0$. Suppose first that $a < b$. We then have that $5^a(\alpha_1 + 5^{b-a}\beta_1) = 5^{2c}\gamma_1^2$. Since $\alpha_1 + 5^{b-a}\beta_1 \equiv j \not\equiv 0 \mod 5$ and $\gamma_1^2 \equiv j^2 \not\equiv 0 \mod 5$, it follows that $a = 2c$. However, $a \sim c$, so $2c \sim c$, whence $2c = c$ by Proposition 3.5, yielding a contradiction. The case that $b < a$ is treated similarly. Finally, suppose that $a = b$. We then have that $5^a(\alpha_1 + \beta_1) = 5^{2c}\gamma_1^2$; since $\alpha_1 + \beta_1 \equiv 2j \not\equiv 0 \mod 5$, as before we conclude that $a = 2c$ and arrive at a contradiction.

Case 2: $i = 2$. We first note that then $\gamma^2 - 4 = 5^{2c}\gamma_1^2 + 5^c \cdot 4\gamma_1 = 5^c(5^c\gamma_1^2 + 4\gamma_1)$ and that $5^c\gamma_1^2 + 4\gamma_1 \equiv 4j \not\equiv 0 \mod 5$. Now suppose that $a < b$. Then $\alpha + \beta - 4 = 5^a(\alpha_1 + 5^{b-a}\beta_1)$ and $\alpha_1 + 5^{b-a}\beta_1 \equiv j \not\equiv 0 \mod 5$. As a result, we have that $\alpha_1 + 5^{b-a}\beta_1 = 5^c\gamma_1^2 + 4\gamma_1$, whence $j \equiv 4j \mod 5$, which is a contradiction. The case that $a > b$ is treated similarly. Finally, suppose that $a = b$. Then $\alpha + \beta - 4 = 5^a(\alpha_1 + \beta_1)$. Since $\alpha_1 + \beta_1 \equiv 2j \mod 5$, we arrive at the contradiction $2j \equiv 4j \mod 5$.

Part III
Combinatorial Number Theory

Chapter 11
Densities and structural properties

11.1 Densities

In this section, A and B denote subsets of $\mathbb{N}$.

Definition 11.1.

1. The *upper density* of A is defined to be
$$\overline{d}(A) := \limsup_{n\to\infty} \delta(A,n).$$
2. The *lower density* of A is defined to be
$$\underline{d}(A) := \liminf_{n\to\infty} \delta(A,n).$$
3. If $\overline{d}(A) = \underline{d}(A)$, then we call this common value the *density of* A and denote it by $d(A)$.

The following exercise concerns the nonstandard characterizations of the aforementioned densities.

Exercise 11.2. Prove that

$$\overline{d}\,(A) = \max\{\mathrm{st}(\delta(A,N)) \ : \ N \in {}^*\mathbb{N}\setminus\mathbb{N}\} = \max\{\mu_N({}^*A) \ : \ N \in {}^*\mathbb{N}\setminus\mathbb{N}\}.$$

State and prove the corresponding statement for lower density.

The previous exercise illustrates why the nonstandard approach to densities is so powerful. Indeed, while densities often "feel" like measures, they lack some of the key properties that measures possess. However, the nonstandard approach allows us to treat densities as measures, thus making it possible to use techniques from measure theory and ergodic theory.

There is something artificial in the definitions of upper and lower density in that one is always required to take samples from initial segments of the natural numbers. We would like to consider a more uniform notion of density which allows one to consider sets that are somewhat dense even though they do not appear to be so when considering only initial segments. This leads us to the concept of (upper) Banach density. In order to defined Banach density, we first need to establish a basic lemma from real analysis, whose nonstandard proof is quite elegant.

Lemma 11.3 (Fekete). *Suppose that* (a_n) *is a* subadditive *sequence of positive real numbers, that is,* $a_{m+n} \leq a_m + a_n$ *for all* m,n. *Then the sequence* $\left(\frac{1}{n}a_n\right)$ *converges to* $\inf\left\{\frac{1}{n}a_n : n \in \mathbb{N}\right\}$.

Proof. After normalizing, we may suppose that $a_1 = 1$. This implies that $\frac{1}{n}a_n \leq 1$ for every $n \in \mathbb{N}$. Set $\ell := \inf\{\frac{1}{n}a_n \ : \ n \in \mathbb{N}\}$. By transfer, there exists $\nu_0 \in {}^*\mathbb{N}$ infinite such that $\frac{1}{\nu_0}a_{\nu_0} \approx \ell$. Furtermore $\mathrm{st}\left(\frac{1}{\nu}a_\nu\right) \geq \ell$ for every $\nu \in {}^*\mathbb{N}$. Fix an infinite $\mu \in {}^*\mathbb{N}$ and observe that for $\nu \geq \mu\nu_0$ one can write $\nu = r\nu_0 + s$ where $r \geq \mu$ and $s < \nu_0$. Therefore

$$\frac{1}{\nu}a_\nu \leq \frac{ra_{\nu_0} + a_s}{r\nu_0 + s} \leq \frac{a_{\nu_0}}{\nu_0} + \frac{a_s}{\mu s} \leq \frac{a_{\nu_0}}{\nu_0} + \frac{1}{\mu} \approx \frac{a_{\nu_0}}{\nu_0} \approx \ell.$$

It follows that $\frac{1}{\nu}a_\nu \approx \ell$ for every $\nu \geq \mu\nu_0$, whence by transfer we have that, for every $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that $\left|\frac{1}{n}a_n - \ell\right| < \varepsilon$ for every $n \geq n_0$. Therefore the sequence $\left(\frac{1}{n}a_n\right)$ converges to ℓ.

For each n, set

$$\Delta_n(A) := \max\{\delta(A,I) \ : \ I \subseteq \mathbb{N} \text{ is an interval of length } n\}.$$

It is straightforward to verify that $(\Delta_n(A))$ is subadditive, whence, by Feketes Lemma, we have that the sequence $((\Delta_n(A))$ converges to $\inf_n \Delta_n(A)$.

Definition 11.4. We define the *Banach density of A* to be

$$\mathrm{BD}(A) = \lim_{n\to\infty} \Delta_n(A) = \inf_n \Delta_n(A).$$

Remark 11.5. Unlike upper and lower densities, the notion of Banach density actually makes sense in any amenable (semi)group, although we will not take up this direction in this book.

If (I_n) is a sequence of intervals in $\mathbb{N}$ such that $\lim_{n\to\infty}|I_n| = \infty$ and $\mathrm{BD}(A) = \lim_{n\to\infty}\delta(A,I_n)$, then we say that (I_n) *witnesses the Banach density of A*.

Here is the nonstandard characterization of Banach density:

Exercise 11.6. For any $N \in {}^*\mathbb{N}\setminus\mathbb{N}$, we have

$$\mathrm{BD}(A) = \max\{\mathrm{st}(\delta({}^*A,I)) \ : \ I \subseteq {}^*\mathbb{N} \text{ is an interval of length } \mathbb{N}\}.$$

As above, if I is an infinite hyperfinite interval such that $\mathrm{BD}(A) = \mathrm{st}(\delta(A,I))$, we also say that *I witnesses the Banach density of A*.

Exercise 11.7. Give an example of a set $A \subseteq \mathbb{N}$ such that $\overline{d}(A) = 0$ but $\mathrm{BD}(A) > 0$.

Exercise 11.8. Prove that Banach density is translation-invariant: $\mathrm{BD}(A+n) = \mathrm{BD}(A)$.

Banach density is also subadditive:

Proposition 11.9. *For any $A,B \subseteq \mathbb{N}$, we have* $\mathrm{BD}(A\cup B) \leq \mathrm{BD}(A) + \mathrm{BD}(B)$.

Proof. Let I be an infinite hyperfinite interval witnessing the Banach density of $A\cup B$. Then

$$\mathrm{BD}(A\cup B) \leq \mathrm{st}(\delta(A,I)) + \mathrm{st}(\delta(B,I)) \leq \mathrm{BD}(A) + \mathrm{BD}(B).$$

The following "fattening" result is often useful.

Proposition 11.10. *If* $\mathrm{BD}(A) > 0$*, then* $\lim_{k\to\infty}\mathrm{BD}(A+[-k,k]) = 1$.

Proof. Set $r := \mathrm{BD}(A)$. For each k, set $a_k := \max_{x\in\mathbb{N}}|A\cap[x+1,x+k]|$, so $r = \lim_{k\to\infty} a_k/k$. By the Squeeze Theorem, it suffices to show that $\mathrm{BD}(A+[-k,k]) \geq \frac{r\cdot k}{a_k}$ for all k. Towards this end, fix $k \in \mathbb{N}$ and $N \in {}^*\mathbb{N}\setminus\mathbb{N}$ and take $x \in {}^*\mathbb{N}$ such that $s := |{}^*A\cap[x+1,x+N\cdot k]|/N\cdot k \approx r$. For $i = 0,1,\ldots,N-1$, set $J_i := [x+ik+1, x+(i+1)k]$. Set $\Lambda := \{i \mid {}^*A\cap J_i \neq \emptyset\}$; observe that Λ is internal. We then have

$$s = \frac{|{}^*A\cap[x+1,x+N\cdot k]|}{N\cdot k} = \frac{\sum_{i\in\Lambda}|{}^*A\cap J_i|}{N\cdot k} \leq \frac{|\Lambda|\cdot a_k}{N\cdot k},$$

whence we can conclude that $|\Lambda| \geq s\cdot N\cdot k/a_k$. Now note that if $i \in \Lambda$, then $J_i \subseteq {}^*A + [-k,k]$, so

$$\frac{|({}^*A+[-k,k])\cap[x+1,x+N\cdot k]|}{N\cdot k} \geq \frac{|\Lambda|\cdot k}{N\cdot k} \geq s\cdot k/a_k.$$

It follows that $\mathrm{BD}(A+[-k,k]) \geq r\cdot k/a_k$.

11.2 Structural properties

We now move on to consider structural notions of largeness. In this section, A continues to denote a subset of $\mathbb{N}$.

Definition 11.11. A is *thick* if and only if A contains arbitrarily long intervals.

Proposition 11.12. *A is thick if and only if there is an infinite hyperfinite interval I contained in ${}^*A$.*

Proof. The backwards direction follows directly from transfer. The forwards direction follows from applied to the set $\{\alpha \in {}^*\mathbb{N} \ : \ {}^*A$ contains an interval of length $\alpha\}$.

Corollary 11.13. *A is thick if and only if* $\mathrm{BD}(A) = 1$.

Proof. The forwards direction is obvious. For the backwards direction, let $N \in {}^*\mathbb{N}$ be divisible by all elements of $\mathbb{N}$ and let I be a hyperfinite interval of length N witnessing the Banach density of A. If A is not thick, then there is m such that $m \mid N$ and A does not contain any intervals of length m. Divide I into N/m many intervals of length m. By transfer, each such interval contains an element of ${}^*\mathbb{N} \setminus {}^*A$. Thus

$$\mathrm{BD}(A) = \mathrm{st}(\delta(A, I)) \leq \mathrm{st}\left(\frac{N - N/m}{N}\right) = 1 - 1/m.$$

Definition 11.14. A is *syndetic* if $\mathbb{N} \setminus A$ is not thick.

Equivalently, A is syndetic if there is m such that all gaps of A are of size at most m.

Proposition 11.15. *A is syndetic if and only if all gaps of ${}^*A$ are finite.*

Proof. The forward direction is immediate by transfer. For the backwards direction, consider the set

$$X := \{\alpha \in {}^*\mathbb{N} \ : \ \text{all gaps of } {}^*A \text{ are of size at most } \alpha\}.$$

By assumption, X contains all elements of ${}^*\mathbb{N} \setminus \mathbb{N}$, so by underflow, there is $m \in X \cap \mathbb{N}$. In particular, all gaps of A are of size at most m.

Definition 11.16. A is *piecewise syndetic* if there is a finite set $F \subseteq \mathbb{N}$ such that $A + F$ is thick.

Proposition 11.17. *If A is piecewise syndetic, then* $\mathrm{BD}(A) > 0$. *More precisely, if F is a finite set such that $\mathbb{N} = A + F$, then* $\mathrm{BD}(A) \geq 1/|F|$.

Proof. Take finite $F \subseteq \mathbb{N}$ such that $A + F$ is thick. Since Banach density is translation invariant, by Proposition 11.9, we have

$$1 = \mathrm{BD}(\mathbb{N}) = \mathrm{BD}(\bigcup_{x \in F} (A + x)) \leq |F| \cdot \mathrm{BD}(A).$$

The notion of being piecewise syndetic is very robust in that it has many interesting reformulations:

Proposition 11.18. *For $A \subseteq \mathbb{N}$, the following are equivalent:*

1. *A is piecewise syndetic;*
2. *there is $m \in \mathbb{N}$ such that $A + [0, m]$ is thick;*
3. *there is $k \in \mathbb{N}$ such that for every $N > \mathbb{N}$, there is a hyperfinite interval I of length N such that ${}^*A$ has gaps of size at most k on I;*
4. *for every $N > \mathbb{N}$, there is a hyperfinite interval I of length N such that all gaps of ${}^*A$ on I are finite;*
5. *there is $k \in \mathbb{N}$ and there is an infinite hyperfinite interval I such that ${}^*A$ has gaps of size at most k on I;*
6. *there is an infinite hyperfinite interval I such that all gaps of ${}^*A$ on I are finite;*

7. *there is $k \in \mathbb{N}$ such that, for every $n \in \mathbb{N}$, there is an interval $I \subseteq \mathbb{N}$ of length n such that the gaps of A on I are of size at most k;*
8. *there is a thick set B and a syndetic set C such that $A = B \cap C$.*

Proof. Clearly (1) and (2) are equivalent and (3) implies (4). Now assume that (3) fails. In particular, if X is the set of $k \in {}^*\mathbb{N}$ for which there is a hyperfinite interval I of length greater than k on which ${}^*A$ has gaps of size greater than k, then X contains all standard natural numbers. By , there is an infinite element of X, whence (4) fails. Thus, (3) and (4) are equivalent. (5) clearly implies (6) and (6) implies (5) follows from a familiar underflow argument. (5) and (7) are also equivalent by transfer-overflow.

We now show (2) impies (3). Fix $N > \mathbb{N}$. By (2) and transfer, there is an interval $[x, x+N) \subseteq {}^*A + [0,m]$. Thus, on $[x, x+N)$, ${}^*A$ has gaps of size at most m.

Clearly (3) $\Rightarrow$ (5). Now suppose that (5) holds. Choose $k \in \mathbb{N}$ and $M, N \in {}^*\mathbb{N}$ such that $M < N$ and $N - M > \mathbb{N}$ such that ${}^*A$ has gaps of size at most k on $[M,N]$. Then $[M+k, N] \subseteq {}^*A + [0,k]$. It follows by transfer that $A + [0,k]$ is thick, whence (2) holds.

Thus far, we have proven that (1)-(7) are equivalent. Now assume that (7) holds and take $k \in \mathbb{N}$ and intervals $I_n \subseteq \mathbb{N}$ of length n such that A has gaps of size at most k on each I_n. Without loss of generality, the I_n's are of distance at least $k+1$ from each other. Let $B := A \cup \bigcup_n I_n$ and let $C := A \cup (\mathbb{N} \setminus B)$. Clearly B is thick. To see that C is syndetic, suppose that J is an interval of size $k+1$ disjoint from C. Then J is disjoint from A and $J \subseteq B$, whence $J \subseteq \bigcup_n I_n$. Since the I_n's are of distance at least $k+1$ from each other, $J \subseteq I_n$ for some n. Thus, J represents a gap of A on I_n of size $k+1$, yielding a contradiction. It is clear that $A = B \cap C$.

Finally, we prove that (8) implies (7). Indeed, suppose that $A = B \cap C$ with B thick and C syndetic. Suppose that $k \in \mathbb{N}$ is such that all gaps of C are of size at most k. Fix $n \in \mathbb{N}$ and let I be an interval of length n contained in B. If J is an interval contained in I of size $k+1$, then $J \cap C \neq \emptyset$, whence $J \cap A \neq \emptyset$ and (7) holds.

Item (7) in the previous proposition explains the name piecewise syndetic. The following is not obvious from the definition:

Corollary 11.19. *The notion of being piecewise is partition regular, meaning that if A is piecewise syndetic and $A = A_1 \sqcup A_2$, then A_i is piecewise syndetic for some $i = 1,2$.*

Proof. Suppose that I is an infinite hyperfinite interval such that all gaps of ${}^*A$ on I are finite. Suppose that I does not witness that A_1 is piecewise syndetic. Then there is an infinite hyperfinite interval $J \subseteq I$ such that $J \cap {}^*A_1 = \emptyset$. It then follows that any gap of ${}^*A_2$ on J must be finite, whence J witnesses that A_2 is piecewise syndetic.

Remark 11.20. We note that neither thickness nor syndeticity are partition regular notions. Indeed, if A is the set of even numbers and B is the set of odd numbers, then neither A nor B is thick but their union certainly is. For syndeticity, let (x_n) be the sequence defined by $x_1 = 1$ and $x_{n+1} := x_n + n$. Set $C := \bigcup_{n \text{ even}} [x_n, x_n + n)$ and $D := \bigcup_{n \text{ odd}} [x_n, x_n + n)$. Then neither C nor D are syndetic but their union is $\mathbb{N}$, a syndetic set.

The following is a nice consequence of the partition regularity of the notion of piecewise syndetic.

Corollary 11.21. *van der Waerden's theorem is equivalent to the statement that piecewise syndetic sets contain arbitrarily long arithmetic progressions.*

Proof. First suppose that van der Waerden's theorem holds and let A be a piecewise syndetic set. Fix $k \in \mathbb{N}$; we wish to show that A contains an arithmetic progression of length k. Take m such that $A + [0,m]$ is thick. Let l be sufficiently large such that when intervals of length I are partitioned into $m+1$ pieces, then there is a monochromatic arithmetic progression of length k. Let $I \subseteq A + [0,m]$ be an interval of length l. Without loss of generality, we may suppose that the left endpoint of I is greater than m. Let c be the coloring of I given by $c(x) :=$ the least $i \in [0,m]$ such that $x \in A + i$. Then there is $i \in [0,m]$ and x, d such that $x, x+d, \ldots, x+(k-1)d \in A+i$. It follows that $(x-i), (x-i)+d, \ldots, (x-i)+(k-1)d \in A$.

Conversely, suppose that piecewise syndetic sets contain arbitrarily long arithmetic progressions. Fix a finite coloring c of the natural numbers. Since being piecewise syndetic is partition regular, some color is piecewise syndetic, whence contains arbitrarily long arithmetic progressions by assumption.

11.3 Working in $\mathbb{Z}$

We now describe what the above densities and structural properties mean in the group $\mathbb{Z}$ as opposed to the semigroup $\mathbb{N}$. Thus, in this section, A now denotes a subset of $\mathbb{Z}$.

It is rather straightforward to define the appropriate notions of density. Indeed, given any sequence (I_n) of intervals in $\mathbb{Z}$ with $\lim_{n\to\infty}|I_n| = \infty$, we define

$$\overline{d}_{(I_n)} := \limsup_{n\to\infty} \delta(A, I_n)$$

and

$$\underline{d}_{(I_n)} := \liminf_{n\to\infty} \delta(A, I_n).$$

When $I_n = [-n, n]$ for each n, we simply write $\overline{d}(A)$ (resp. $\underline{d}(A)$) and speak of the *upper* (resp. *lower*) density of A. Finally, we define the *upper Banach density of* A to be

$$\mathrm{BD}(A) = \lim_{n\to\infty} \max_{x\in\mathbb{N}} \delta(A, [x-n, x+n]).$$

Of course, one must verify that this limit exists, but this is proven in the exact same way as in the case of subsets of $\mathbb{N}$.

Exercise 11.22. Prove that

$$\mathrm{BD}(A) := \max\{\overline{d}_{(I_n)}(A) \ : \ (I_n) \text{ a sequence of intervals with } \lim_{n\to\infty}|I_n| = \infty\}.$$

The notions of thickness and syndeticity for subsets of $\mathbb{Z}$ remains unchanged: A is thick if A contains arbitrarily long intervals and A is syndetic if $\mathbb{Z}\setminus A$ is not thick. Similarly, A is piecewise syndetic if there is a finite set $F \subseteq \mathbb{Z}$ such that $A+F$ is thick. The following lemma is almost immediate:

Lemma 11.23. *A is piecewise syndetic if and only if there is a finite set $F \subseteq \mathbb{Z}$ such that, for every finite $L \subseteq \mathbb{Z}$, we have $\bigcap_{x\in L}(A+F+x) \neq \emptyset$.*

Exercise 11.24. Formulate and verify all of the nonstandard equivalents of the above density and structural notions developed in the previous two sections for subsets of $\mathbb{Z}$.

The following well-known fact about difference sets has a nice nonstandard proof.

Proposition 11.25. *Suppose that $A \subseteq \mathbb{Z}$ is such that $\mathrm{BD}(A) > 0$. Then $A-A$ is syndetic. In fact, if $\mathrm{BD}(A) = r$, then there is a finite set $F \subseteq \mathbb{Z}$ with $|F| \leq \frac{1}{r}$ such that $(A-A)+F = \mathbb{Z}$.*

First, we need a lemma.

Lemma 11.26. *Suppose that $E \subseteq [1,N]$ is such that $\delta(E,N) \approx r$. Then there is a finite $F \subseteq \mathbb{Z}$ with $|F| \leq 1/r$ such that $\mathbb{Z} \subseteq (E-E)+F$.*

Proof. Fix $x_1 \in \mathbb{N}$. If $\mathbb{Z} \subseteq (E-E)+x_1$, then take $F = \{x_1\}$. Otherwise, take $x_2 \notin (E-E)+\{x_1\}$. If $\mathbb{Z} \subseteq (E-E)+\{x_1,x_2\}$, then take $F = \{x_1,x_2\}$. Otherwise, take $x_3 \notin (E-E)+\{x_1,x_2\}$.

Suppose that $x_1,\ldots,x_k$ have been constructed in this fashion. Note that the sets $E+x_i$, for $i=1,\ldots,k$, are pairwise disjoint. Since each $x_i \in \mathbb{Z}$, we have that $\delta((E+x_i),N) \approx r$. It follows that

$$\delta\left(\bigcup_{i=1}^{k}(E+x_i), N\right) = \frac{\sum_{i=1}^{k}|(E+x_i)\cap[1,N]|}{N} \approx kr.$$

It follows that $k \leq \frac{1}{r}$.

Proof (of Proposition 11.25). Set $r := \mathrm{BD}(A)$. Fix N and take $x \in {}^*\mathbb{N}$ such that $\delta({}^*A, [x+1, x+\nu]) \approx r$. Set $E := ({}^*A - x)\cap[1,N]$. Then $\delta(E,N) \approx r$, whence there is finite $F \subseteq \mathbb{Z}$ with $|F| \leq 1/r$ such that $\mathbb{Z} \subseteq (E-E)+F$. It follows that $\mathbb{Z} \subseteq ({}^*A - {}^*A)+F$, whence it follows by transfer that $\mathbb{Z} = (A-A)+F$.

The analog of Proposition 11.10 for $\mathbb{Z}$ is also true:

Proposition 11.27. *If* $\mathrm{BD}(A) > 0$*, then* $\lim_{k\to\infty} \mathrm{BD}(A+[-k,k]) = 1$.

However, for our purposes in Section 13.5, we will need a more precise result. Note that, a priori, for every $\varepsilon > 0$, there is k_ε and infinite hyperfinite interval I_ε such that $\delta(^*A+[-k_\varepsilon,k_\varepsilon],I_\varepsilon) > 1-\varepsilon$. The next proposition tells us that we can take a single interval I to work for each ε. The proof is heavily inspired by the proof of [, Lemma 3.2].

Proposition 11.28. *Suppose that* $\mathrm{BD}(A) > 0$*. Then there is an infinite hyperfinite interval* $I \subseteq \mathbb{Z}$ *such that, for every* $\varepsilon > 0$*, there is* k *for which* $\delta(^*A+[-k,k],I) > 1-\varepsilon$*.*

Proof. Let (I_n) be a sequence of intervals in $\mathbb{Z}$ winessing the Banach density of A and such that, for every k, we have that $\lim_{n\to\infty} \delta(A+[-k,k],I_n)$ exists. (This is possible by a simple diagonalization argument.) Fix N and, for each $\alpha \in {}^*\mathbb{N}$, set $G_\alpha := (^*A+[-\alpha,\alpha]) \cap I_N$. Set $r := \sup_{k\in\mathbb{N}} \mu_{I_N}(G_k)$.

Claim: There is $K > \mathbb{N}$ such that:

(i) For every $l \in \mathbb{Z}$, $\frac{|(l+G_K)\triangle G_K|}{|G_K|} \approx 0$.
(ii) $\frac{|G_K|}{I_N|} \approx r$.

Proof of Claim: For each $l \in \mathbb{Z}$, set X_l to be the set of $\alpha \in {}^*\mathbb{N}$ such that:

(a) $\alpha \geq l$;
(b) For all $x \in \mathbb{Z}$ with $|x| \leq l$, we have $\frac{|(x+G_\alpha)\triangle G_\alpha|}{|G_\alpha|} < \frac{1}{l}$;
(c) $\left|\frac{|G_\alpha|}{I_N|} - r\right| < \frac{1}{l}$.

Since each X_l is internal and unbounded in $\mathbb{N}$, by saturation there is $K \in \bigcap_l X_l$. This K is as desired.

Fix K as in the Claim and set $G := G_K$ and $\mu := \mu_G$. For $k \in \mathbb{N}$, we then have that

$$\delta(^*A+[-k,k],G) = \frac{|(^*A+[-k,k]) \cap I_N|}{|G|} \approx \delta(^*A+[-k,k],I_N) \cdot \frac{1}{r},$$

whence we see that $\delta(^*A+[-k,k],G) \to 1$ as $k \to \infty$.[1]

Now take J to be an infinite hyperfinite interval such that $\frac{|(l+G_K)\triangle G_K|}{|G_K|} \approx 0$ for all $l \in J$; this is possible by . We claim that there is $t \in G$ such that $I := t+J$ is as desired.

For each k, take n_k such that $\delta(^*A+[-n_k,n_k],G) > 1-\frac{1}{k}$; without loss of generality, we may assume that (n_k) is an increasing sequence. Set $B_k := {}^*A+[-n_k,n_k]$ and set $g_k : G \to [0,1]$ to be the $\mathscr{L}_G$-measurable function given by $g_k(t) := \mathrm{st}(\delta(B_k,t+J))$. For each $t \in G$, we have that $(g_k(t))$ is a bounded nondecreasing sequence, whence converges to a limit $g(t)$. By the Dominated Convergence Theorem, we have that $\int_G g(t)d\mu = \lim_{k\to\infty} \int_G g_k(t)d\mu$. Now note that

$$\int_G g_k(t)d\mu \approx \frac{1}{|G|}\sum_{t\in G} \delta(B_k,t+J) = \frac{1}{|I|}\sum_{x\in J} \delta(B_k,x+G) \approx \delta(B_k,G) > 1-\frac{1}{k}.$$

It follows that $\int_G g(t)d\mu = 1$, whence $g(t) = 1$ for some $t \in G$. It is then clear that $I := t+J$ is as desired.

We call I as in the conclusion of Proposition 11.28 *good for A*. One can also prove the previous proposition using a Lebesgue Density Theorem for cut spaces; see [12].

[1] At this point, we may note that G satisfies the conclusion of the proposition except that it is not an interval but instead a *Folner approximation* for $\mathbb{Z}$. While this would suffice for our purposes in Section 13.5, we wanted to avoid having to introduce the theory of Folner approximations and instead opted to work a bit harder to obtain the above cleaner statement.

11.4 Furstenberg's Correspondence Principle

We end this chapter by explaining the nonstandard take on Furstenberg's correspondence principle.

Theorem 11.29 (Furstenberg's Correspondence Principle). *Suppose that $A \subseteq \mathbb{Z}$ is such that $\mathrm{BD}(A) > 0$. Then there is a measure-preserving dynamical system $(X, \mathscr{B}, \nu, T)$ and a measurable set $A_0 \in \mathscr{B}$ such that $\nu(A_0) = \mathrm{BD}(A)$ and such that, for any finite set $F \subseteq \mathbb{Z}$, we have:*

$$\mathrm{BD}\left(\bigcap_{i\in F}(A-i)\right) \geq \nu\left(\bigcap_{i\in F} T^{-i}(A_0)\right).$$

Proof. Fix $I \subseteq {}^*\mathbb{Z}$ witnessing the Banach density of A. It is easy to verify that the hypercycle system (I, Ω, μ, S) introduced in Section 6.6 of Chapter 6 and the set $A_0 := {}^*A \cap I$ are as desired.

Let us mention the ergodic-theoretic fact that Furstenberg proved:

Theorem 11.30 (Furstenberg Multiple Recurrence Theorem). *Suppose that $(X, \mathscr{B}, \nu, T)$ is a measure-preserving dynamical system, $A \in \mathscr{B}$ is such that $\nu(A) > 0$, and $k \in \mathbb{N}$ is given. Then there exists $n \in \mathbb{N}$ such that $\nu(A \cap T^{-n}(A) \cap T^{-2n}(A) \cap \cdots \cap T^{-(k-1)n}(A)) > 0$.*

Notice that the above theorem, coupled with the Furstenberg Correspondence Principle, yields Furstenberg's proof of Szemerédi's Theorem .

Theorem 11.31 (Szemeredi's Theorem). *If $A \subseteq \mathbb{Z}$ is such that $\mathrm{BD}(A) > 0$, then A contains arbitrarily long arithmetic progressions.*

Szemeredi's Theorem is the density version of van der Waerden's theorem and was originally proven by Szemeredi in [42].

Chapter 12
Densities in the remote realm

The material in this chapter is taken from [28].

12.1 Banach density as Shnirelmann density in the remote realm

The title in this chapter refers to looking at copies of $\mathbb{N}$ starting at some infinite element $a \in {}^*\mathbb{N}$ and then connecting some density of the set of points of this copy of $\mathbb{N}$ that lie in the nonstandard extension of a set A and some other density of the original set A itself. In this regard, given $A \subseteq \mathbb{N}$ and $a \in {}^*\mathbb{N}$, we set $\overline{d}({}^*A - a) := \overline{d}(({}^*A - a) \cap \mathbb{N})$ and likewise for other notions of density. We warn the reader that, in general, we do not identify ${}^*A - a$ and $({}^*A - a) \cap \mathbb{N}$ *as sets*, but since we have not defined the density of a subset of ${}^*\mathbb{N}$, our convention should not cause too much confusion.

Lemma 12.1. *Suppose $A \subseteq \mathbb{N}$ is such that there is $a \in {}^*\mathbb{N}$ with $\underline{d}({}^*A - a) \geq r$. Then* $\mathrm{BD}(A) \geq r$.

Proof. Fix $\varepsilon > 0$. Set
$$D := \{\alpha \in {}^*\mathbb{N} \ : \ \delta({}^*A, [a+1, a+\alpha]) \geq r - \varepsilon\}.$$
By the assumption, $D \cap \mathbb{N}$ is unbounded in $\mathbb{N}$. By , there is infinite $N \in D$, whence, by the nonstandard characterization of Banach density, we see that $\mathrm{BD}(A) \geq r - \varepsilon$. Since $\varepsilon > 0$ was arbitrary, the result follows.

The key observation of Renling Jin is that there is a strong converse to the previous lemma.

Proposition 12.2. *Suppose that $A \subseteq \mathbb{N}$ is such that* $\mathrm{BD}(A) = r$. *Let I be an interval of infinite hyperfinite length witnessing the Banach density of A. Then for μ_I-almost all $x \in I$, we have $d({}^*A - x) = r$.*

Proof. Write $I = [H, K]$ and consider the hypercycle system $(I, \mathscr{L}_i, \mu_I, S)$. Let f denote the characteristic function of ${}^*A \cap I$. It follows that, for $x \in I^\# := \bigcap_{n \in \mathbb{N}} [H, K - n]$, we have that
$$\frac{1}{n} \sum_{m=0}^{n-1} f(S^m(x)) = \delta({}^*A, [x, x+n-1]).$$
By the ergodic theorem for hypercycles (Theorem 6.24), there is a $\mathscr{L}_I$-measurable function $\bar{f}$ such that, for μ_I-almost all $x \in I$, we have that
$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} f(S^m(x)) = \bar{f}(x).$$
Since $I^\#$ is a μ_I-conull set, we will thus be finished if we can show that $\bar{f}$ is μ_I-almost everywhere equal to r on $I^\#$.

Towards this end, first note that $\bar{f}(x) \leq r$ for μ_I-almost all $x \in I^\#$. Indeed, if $\bar{f}(x) > r$ for a positive measure set of $x \in I^\#$, then there would be some $x \in I^\#$ with $d({}^*A - x) > r$, whence $\mathrm{BD}(A) > r$ by Lemma 12.1, yielding a contradiction.

Next note that, by the Dominated Convergence Theorem, we have that

$$\int_I \bar{f}(x)d\mu_I = \lim_{n\to\infty}\int_I \frac{1}{n}\sum_{m=0}^{n-1} f(S^m(x))d\mu_I = r,$$

where the last equality follows from the fact that S is measure-preserving and that $\int_I f d\mu_I = \mu_I({}^*A) = r$. By a standard measure theory argument, we have that $\bar{f}(x) = r$ for almost all $x \in I^{\#}$.

Remark 12.3. In the context of the previous proposition, since $\mu_I({}^*A) > 0$, we can conclude that there is $x \in {}^*A$ such that $d({}^*A - x) = r$.

Summarizing what we have seen thus far:

Theorem 12.4. *For $A \subseteq \mathbb{N}$, the following are equivalent:*

1. $\mathrm{BD}(A) \geq r$*;*
2. *for any infinite hyperfinite interval I witnessing the Banach density of A, we have $d({}^*A - x) \geq r$ for μ_I-almost all $x \in I$;*
3. *there is $x \in {}^*\mathbb{N}$ such that $\underline{d}({}^*A - x) \geq r$.*

We now introduce a new notion of density.

Definition 12.5. For $A \subseteq \mathbb{N}$, we define the *Shnirelman density* of A to be

$$\sigma(A) := \inf_{n\geq 1} \delta(A,n).$$

It is clear from the definition that $\underline{d}(A) \geq \sigma(A)$. Note that the Shnirelman density is very sensitive to what happens for "small" n. For example, if $1 \notin A$, then $\sigma(A) = 0$. On the other hand, knowing that $\sigma(A) \geq r$ is a fairly strong assumption and thus there are nice structural results for sets of positive Shnirelman density. We will return to this topic in the next section.

A key idea of Jin was to add one more equivalence to the above theorem, namely that there is $x \in {}^*\mathbb{N}$ such that $\sigma({}^*A - x) \geq r$; in this way, one can prove Banach density parallels of theorems about Shnirelman density. To add this equivalence, one first needs a standard lemma.

Lemma 12.6. *Suppose that $A \subseteq \mathbb{N}$ is such that $\underline{d}(A) = r$. Then for every $\varepsilon > 0$, there is $n_0 \in \mathbb{N}$ such that $\sigma(A - n_0) \geq r - \varepsilon$.*

Proof. Suppose that the lemma is false for a given $\varepsilon > 0$. In particular, $\sigma(A) < r - \varepsilon$, so there is $n_0 \in \mathbb{N}$ such that $\delta(A, n_0) < r - \varepsilon$. Since n_0 does not witness the truth of the lemma, there is $n_1 \in \mathbb{N}$ such that $\delta((A - n_0), n_1) < r - \varepsilon$. Continuing in this way, we find an increasing sequence $n_0 < n_1 < n_2 < \cdots$ such that, for all i, we have $\delta(|A, [n_i + 1, n_i + n_{i+1}]) < r - \varepsilon$. This sequence witnesses that $\underline{d}(A) \leq r - \varepsilon$, yielding a contradiction.

Proposition 12.7. *Suppose that* $\mathrm{BD}(A) \geq r$*. Then there is $x \in {}^*\mathbb{N}$ such that $\sigma({}^*A - x) \geq r$.*

Proof. Take $y \in {}^*\mathbb{N}$ such that $d({}^*A - y) \geq r$. By the previous lemma, for each $n \in \mathbb{N}$, there is $z_n \in {}^*\mathbb{N}$ with $z_n \geq y$ such that $\sigma({}^*A - z_n) \geq r - 1/n$. By , for each $n \in \mathbb{N}$, there is infinite $K_n \in {}^*\mathbb{N}$ such that, for each $m \leq K_n$, we have

$$\delta(({}^*A - z_n), m) \geq r - 1/n.$$

Take infinite $K \in {}^*\mathbb{N}$ such that $K \leq K_n$ for each n. Let

$$D := \{\alpha \in {}^*\mathbb{N} \ : \ (\exists z \in {}^*\mathbb{N})(\forall m \leq K)\delta({}^*A - z), m) \geq r - 1/\alpha\}.$$

Then D is internal and $\mathbb{N} \subseteq D$, whence by there is infinite $N \in D$. Take $x \in {}^*\mathbb{N}$ such that $\delta({}^*A - x), m) \geq r - 1/N$ for all $m \leq N$. In particular, for all $m \in \mathbb{N}$, we have $\delta(({}^*A - x), m) \geq r$, whence this x is as desired.

Theorem 12.4 and Proposition 12.7 immediately yield:

Corollary 12.8. $\mathrm{BD}(A) \geq r$ *if and only if there is $x \in {}^*\mathbb{N}$ such that $\sigma({}^*A - x) \geq r$.*

We end this section with a curious application of Proposition 12.7. We will make more serious use of this technique in the next section.

Proposition 12.9. *Fix $\varepsilon > 0$. Then Szemeredi's Theorem is equivalent to the following (apparently weaker statement): if $A \subseteq \mathbb{N}$ is such that $\sigma(A) \geq 1-\varepsilon$, then A contains arbitrarily long arithmetic progressions.*

Proof. Fix $A \subseteq \mathbb{N}$ with $\mathrm{BD}(A) > 0$; we wish to show that A contains arbitrarily long arithmetic progressions. By Proposition 11.10, there is $k \in \mathbb{N}$ such that $\mathrm{BD}(A+[0,k]) \geq 1-\varepsilon$. If $A+[0,k]$ contains arbitrarily long arithmetic progressions, then by van der Waerden's theorem, there is $i \in [0,k]$ such that $A+i$ contains arbitrarily long arithmetic progressions, whence so does A. It follows that we may assume that $\mathrm{BD}(A) \geq 1-\varepsilon$.

By Proposition 12.7, we have $x \in {}^*\mathbb{N}$ such that $\sigma({}^*A - x) \geq 1-\varepsilon$, whence, by assumption, we have that ${}^*A - x$ contains arbitrarily long arithmetic progressions, and hence so does ${}^*A$. The result now follows from transfer.

12.2 Applications

We use the ideas from the preceding section to derive some Banach density versions of theorems about Shnirelman density. We first recall the following result of Shnirleman (see, for example, [24, page 8]):

Theorem 12.10. *Suppose that $A \subseteq \mathbb{N}_0$ is such that $0 \in A$ and $\sigma(A) > 0$. Then A is a* basis*, that is, there is $h \in \mathbb{N}$ such that $\Sigma_h(A) = \mathbb{N}$.*

Using nonstandard methods, Jin was able to prove a Banach density version of the aforementioned result:

Theorem 12.11. *Suppose that $A \subseteq \mathbb{N}$ is such that $\gcd(A - \min(A)) = 1$ and $\mathrm{BD}(A) > 0$. Then A is a* Banach basis*, that is, there is $h \in \mathbb{N}$ such that $\Sigma_h(A)$ is thick.*

Note that we must assume that $\gcd(A-\min(A)) = 1$, for if $\gcd(A-\min(A)) = c > 1$, then $hA \subseteq \{h\min(A)+nc \ : \ n \in \mathbb{N}\}$, which does not contain arbitrarily long intervals.

Proof (of Theorem 12.11). Suppose $\mathrm{BD}(A) = r$ and $\gcd(A-\min(A)) = 1$. The latter property guarantees the existence of $m \in \mathbb{N}$ such that $\Sigma_m(A-\min(A))$ contains two consecutive numbers, whence $c, c+1 \in \Sigma_m(A)$ for some $c \in \mathbb{N}$. By Proposition 12.7, there is $a \in {}^*\mathbb{N}$ such that $\sigma({}^*A - a+1) \geq r$. In particular, $a \in {}^*A$. Consequently, we have

$$\sigma(\Sigma_{1+m}({}^*A) - a - c) \geq \sigma({}^*A + \{c, c+1\} - a - c) \geq \sigma({}^*A - a + 1) \geq r.$$

Since $0 \in \Sigma_{1+m}({}^*A) - a - c$, Shnirleman's theorem implies that there is n such that $\mathbb{N} \subseteq \Sigma_n(\Sigma_{1+m}({}^*A) - a - c)$. By , there is N such that $[0,N] \subseteq \Sigma_n(\Sigma_{1+m}({}^*A) - a - c)$. Set $h := n(1+m)$, so $[0,N] + n(a+c) \subseteq {}^*(\Sigma_h(A))$. By transfer, $\Sigma_h(A)$ contains arbitrarily long intervals.

With similar methods, one can prove the Banach density analogue of the following theorem of Mann (see, for example, [24, page 5]):

Theorem 12.12. *Given $A, B \subseteq \mathbb{N}_0$ such that $0 \in A \cap B$, we have $\sigma(A+B) \geq \min\{\sigma(A) + \sigma(B), 1\}$.*

Observe that the exact statement of Mann's theorem is false if one replaces Shnirelman density by Banach density. Indeed, if A and B are both the set of even numbers, then $\mathrm{BD}(A+B) = \frac{1}{2}$ but $\mathrm{BD}(A) + \mathrm{BD}(B) = 1$. However, if one replaces $A+B$ by $A+B+\{0,1\}$, the Banach density version of Mann's theorem is true.

Theorem 12.13. *Given $A, B \subseteq \mathbb{N}$, we have $\mathrm{BD}(A+B+\{0,1\}) \geq \min\{\mathrm{BD}(A) + \mathrm{BD}(B), 1\}$.*

The idea behind the proof of Theorem 12.13 is, as before, to reduce to the case of Shnirelman density by replacing the given sets with hyperfinite shifts. In the course of the proof of Theorem 12.13, we will need to use the following fact from additive number theory (see, for example, [24, page 6]):

Theorem 12.14 (Besicovitch's theorem). *Suppose $A,B\subseteq\mathbb{N}$ and $s\in[0,1]$ are such that $1\in A$, $0\in B$, and $|B\cap[1,n]|\geq s(n+1)$ for every $n\in\mathbb{N}$. Then $\sigma(A+B)\geq\min\{\sigma(A)+\sigma(B),1\}$.*

For a proof of Besicovitch's theorem, see, for example, [24, page 6].

Proof (of Theorem 12.13). Set $r:=\mathrm{BD}(A)$ and $s:=\mathrm{BD}(B)$. We can assume, without loss of generality, that $r\leq s\leq 1/2$. By Proposition 12.7, one can find $a\in{}^*A$ and $b\in{}^*B$ such that $\sigma({}^*A-a+1)\geq r$ and $\sigma({}^*B-b+1)\geq s$.

Claim: For every $n\in\mathbb{N}$, one has that $|({}^*B+\{0,1\})\cap[b+1,b+n]|\geq s(n+1)$.
Proof of Claim: Let $[1,k_0]$ be the largest initial segment of $\mathbb{N}$ contained in $({}^*B+\{0,1\}-b)\cap\mathbb{N}$ (if no such k_0 exists, then the claim is clearly true) and let $[1,k_1]$ be the largest initial segment of $\mathbb{N}$ disjoint from $(({}^*B+\{0,1\})-(b+k_0))\cap\mathbb{N}$. We note the following:

- For $1\leq n\leq k_0$, we have that

$$|({}^*B+\{0,1\})\cap[b+1,b+n]|=n\geq(n+1)/2\geq s(n+1).$$

- For $k_0+1\leq n<k_0+k_1$, since $\sigma({}^*B-b+1)\geq s$, we have that

$$\begin{aligned}|({}^*B+\{0,1\})\cap[b+1,b+n]|&\geq|{}^*(B+1)\cap[b+1,b+n]|\\&=|{}^*(B+1)\cap[b+1,b+n+1]|\\&\geq s(n+1).\end{aligned}$$

- For $n\geq k_0+k_1$, since $k_0+k_1+1\in{}^*B$, $k_0+k_1+1\notin{}^*B+1$, and $\sigma({}^*B-b+1)\geq s$, we have that

$$\begin{aligned}|({}^*B+\{0,1\})\cap[b+1,b+n]|&\geq|{}^*(B+1)\cap[b+1,b+n]|+1\\&\geq sn+1\geq s(n+1).\end{aligned}$$

These observations conclude the proof of the claim.

One can now apply Besicovitch's theorem to ${}^*A-a+1$ and ${}^*B+\{0,1\}-b$ (intersected with $\mathbb{N}$) to conclude that

$$\sigma(({}^*A-a+1)+({}^*B+\{0,1\}-b))\geq\min\{\sigma({}^*A-a+1)+s,1\}\geq r+s.$$

Finally, observe that

$$({}^*A-a+1)+({}^*B+\{0,1\}-b)={}^*(A+B+\{0,1\})-(a+b).$$

Hence

$$\mathrm{BD}(A+B+\{0,1\})\geq\sigma({}^*(A+B+\{0,1\})-(a+b))\geq r+s.$$

Chapter 13
Jin's Sumset Theorem

13.1 The statement of Jin's Sumset Theorem and some standard consequences

Definition 13.1. An initial segment U of ${}^*\mathbb{N}_0$ is a *cut* if $\mathrm{U}+\mathrm{U} \subseteq \mathrm{U}$.

Exercise 13.2. If U is a cut, then either U is external or else $\mathrm{U} = {}^*\mathbb{N}$.

Example 13.3.

1. $\mathbb{N}$ is a cut.
2. If N is an infinite element of ${}^*\mathbb{N}$, then $\mathrm{U}_N := \{x \in {}^*\mathbb{N} \ : \ \frac{x}{N} \approx 0\}$ is a cut.

Fix a cut U of ${}^*\mathbb{N}$ and suppose that $\mathrm{U} \subseteq [0,N)$. Given $x,y \in {}^*\mathbb{N}$, we write $x \sim_{\mathrm{U}} y$ if $|x-y| \in \mathrm{U}$; note that $\sim_{\mathrm{U}}$ is an equivalence relation on ${}^*\mathbb{N}$. We let $[x]_{\mathrm{U},N}$, or simply $[x]_N$ if no confusion can arise, denote the equivalence class of x under $\sim_{\mathrm{U}}$ and we let $[0,N)/\mathrm{U}$ denote the set of equivalence classes. We let $\pi_{\mathrm{U}} : [0,N) \to [0,N)/\mathrm{U}$ denote the quotient map. The linear order on $[0,N)$ descends to a linear order on $[0,N)/\mathrm{U}$. Moreover, one can push forward the Loeb measure on $[0,N)$ to a measure on $[0,N)/\mathrm{U}$, which we also refer to as Loeb measure.

Example 13.4. Fix $N \in {}^*\mathbb{N}$ infinite and consider the cut U_N from Example 13.3. Note that the surjection $f : [0,N) \to [0,1]$ given by $f(\beta) := \mathrm{st}(\beta/N)$ descends to a bijection of ordered sets $f : [0,N)/\mathrm{U}_N \to [0,1]$. The discussion in Section 6.3 of Chapter 6 shows that the measure on $[0,1]$ induced by the Loeb measure on $[0,N)/\mathrm{U}_N$ via f is precisely Lebesgue measure.

For any cut U contained in $[0,N)$, the set $[0,N)/\mathrm{U}$ has a natural topology induced from the linear order, whence it makes sense to talk about category notions in $[0,N)/\mathrm{U}$. (This was first considered in [33].) It will be convenient to translate the category notions from $[0,N)/\mathrm{U}$ back to $[0,N]$:

Definition 13.5. $A \subseteq [0,N)$ is U-*nowhere dense* if $\pi_{\mathrm{U}}(A)$ is nowhere dense in $[0,N)/U$. More concretely: A is U-nowhere dense if, given any $a < b$ in $[0,N)$ with $b-a > \mathrm{U}$, there is $[c,d] \subseteq [a,b]$ with $d-c > \mathrm{U}$ such that $[c,d] \subseteq [0,N) \setminus A$. If A is not U-nowhere dense, we say that A is U-*somewhere dense*.

Recall the following famous theorem of Steinhaus:

Theorem 13.6. *If $C,D \subseteq [0,1]$ have positive Lebesgue measure, then $C+D$ contains an interval.*

For $x,y \in [0,N)$, set $x \oplus_N y := x+y \mod N$. For $A,B \subseteq [0,N)$, set

$$A \oplus_N B := \{x \oplus_N y \ : \ x \in A, y \in B\}.$$

In light of Example 13.4, Theorem 13.6 says that whenever $A,B \subseteq [0,N)$ are internal sets of positive Loeb measure, then $A \oplus_N B$ is U_N-somewhere dense. Keisler and Leth asked whether or not this is the case for any cut. Jin answered this positively in [29]:

Theorem 13.7 (Jin's Sumset Theorem). *If* $\mathrm{U} \subseteq [0,N)$ *is a cut and* $A,B \subseteq [0,N)$ *are internal sets with positive Loeb measure, then* $A \oplus_N B$ *is* U*-somewhere dense.*

Exercise 13.8. Prove Theorem 13.6 from Theorem 13.7.

We will prove Theorem 13.7 in the next section. We now prove the following standard corollary of Theorem 13.7, which is often also referred to as Jin's sumset theorem, although this consequence was known to Leth beforehand.

Corollary 13.9. *Suppose that* $A,B \subseteq \mathbb{N}$ *have positive Banach density. Then* $A+B$ *is piecewise syndetic.*

Proof. Set $r := \mathrm{BD}(A)$ and $s := \mathrm{BD}(B)$. Fix $N \in {}^*\mathbb{N}$ infinite and take $x,y \in {}^*\mathbb{N}$ such that

$$\delta({}^*A \cap [x,x+N)|) \approx r, \quad \delta({}^*B \cap [y,y+N)) \approx s.$$

Let $C := {}^*A - x$ and $D := {}^*B - y$, so we may view C and D as internal subsets of $[0,2N)$ of positive Loeb measure. By Jin's theorem applied to the cut $\mathbb{N}$, we have that $C \oplus_{2N} D = C + D$ is $\mathbb{N}$-somewhere dense, that is, there is a hyperfinite interval I such that all gaps of $C+D$ on I have finite length. By , there is $m \in \mathbb{N}$ such that all gaps of $C+D$ on I have length at most m. Therefore, $x+y+I \subseteq {}^*(A+B+[0,m])$. By transfer, for any $k \in \mathbb{N}$, $A+B+[0,m]$ contains an interval of length k, whence $A+B$ is piecewise syndetic.

It is interesting to compare the previous corollary to Proposition 11.25. It is also interesting to point out that Corollary 13.9 can also be used to give an alternative proof of Theorem 12.11. Indeed, suppose $\mathrm{BD}(A) > 0$ and $\gcd(A - \min(A)) = 1$. Then there is $h \in \mathbb{N}$ such that $A+A+[0,h]$ is thick. It follows that $A+A+[x,x+h]$ is thick for all $x \in \mathbb{N}$. As in the proof of Theorem 12.11, take m and consecutive $a, a+1 \in \Sigma_m(A)$. Note that, for all $i = 0,1,\ldots,h$, we have that $ha+i = i(a+1)+(h-i)a \in \Sigma_{hm}(A)$. It follows that $A+A+[ha,ha+h] \subseteq \Sigma_{hm+2}(A)$, whence $\Sigma_{hm+2}(A)$ is thick.

13.2 Jin's proof of the sumset theorem

We now turn to the proof of Theorem 13.7 given in [29]. Suppose, towards a contradiction, that there is a cut U for which the theorem is false. If $H > \mathrm{U}$ and $A,B \subseteq [0,H)$ are internal, we say that (A,B) is (H,U)-*bad* if $\mu_H(A), \mu_H(B) > 0$ and $A \oplus_H B$ is U-nowhere dense. We set

$$r := \sup\{\mu_H(A) \ : \ (A,B) \text{ is } (H,\mathrm{U}) \text{ bad for some } H > \mathrm{U} \text{ and some } B \subseteq [0,H)\}.$$

By assumption, $r > 0$. We fix $\varepsilon > 0$ sufficiently small. We then set

$$s := \sup\{\mu_H(B) \ : (A,B) \text{ is } (H,\mathrm{U})\text{-bad for some } H > \mathrm{U} \text{ and some } A \subseteq [0,H) \text{ with } \mu_H(A) > r - \varepsilon\}.$$

By the definition of r, we have that $s > 0$. Also, by the symmetry of the definition of r, we have that $r \geq s$. The following is slightly less obvious:

Claim 1: $s < \frac{1}{2} + \varepsilon$.
Proof of Claim 1: Suppose, towards a contradiction, that $s \geq \frac{1}{2} + \varepsilon$. We may thus find $H > \mathbb{N}$ and an (H,U)-bad pair (A,B) with $\mu_H(A) > \frac{1}{2}$ and $\mu_H(B) > \frac{1}{2}$. Since addition modulo H is translation invariant, it follows that for any $x \in [0,H)$, we have that $A \cap (x \ominus_H B) \neq \emptyset$, whence $A \oplus_H B = [0,H-1)$, which is a serious contradiction to the fact that $A \oplus_H B$ is U-nowhere dense.

We now fix $\delta > 0$ sufficiently small, $H > \mathrm{U}$ and an (H,U)-bad (A,B) such that $\mu_H(A) > r - \varepsilon$ and $\mu_H(B) > s - \delta$. We will obtain a contradiction by producing $K > \mathrm{U}$ and (K,U)-bad (A',B') such that $\mu_K(A') > r - \varepsilon$ and $\mu_K(B') > s + \delta$, contradicting the definition of s.

We first show that it suffices to find $K > U$ such that $K/H \approx 0$ and such that there are hyperfinite intervals $I,J \subseteq [0,H)$ of length K for which

$$\mathrm{st}\left(\frac{|A\cap I|}{K}\right) > r-\varepsilon \text{ and } \mathrm{st}\left(\frac{|B\cap J|}{K}\right) > s+\delta.$$

Indeed, suppose that $I := [a, a+K)$ and $J := [b, b+K)$ are as above. Let $A' := (A\cap I) - a$ and $B' := (B\cap J) - b$. Then $\mu_K(A') > r-\varepsilon$ and $\mu_K(B') > s+\delta$. It remains to see that (A', B') is (K, U)-bad. Since $A\oplus_H B$ is U-nowhere dense, it is clear that $(A\cap I)\oplus_H (B\cap J)$ is also U-nowhere dense. Since $A'\oplus_H B' = ((A\cap I)\oplus_H (B\cap J))\ominus (a+b)$, we have that $A'\oplus_H B'$ is U-nowhere dense. Since K/H is infinitesimal, we have that $A'\oplus_H B' = A'\oplus_{2K} B'$. It follows that $A'\oplus_K B'$ is the union of two U-nowhere dense subsets of $[0,K)$, whence is also U-nowhere dense, and thus (A', B') is (K, U)-bad, as desired.

We now work towards finding the appropriate K. By the definition of U-nowhere dense, we have, for every $k\in \mathrm{U}$, that $A\oplus_H (B\oplus_H [-k,k])) = (A\oplus_H B)\oplus_H [-k,k]$ is U-nowhere dense. By the definition of s, it follows that $\mu_H(B\oplus_H [-k,k]) \leq s$ for each $k\in \mathrm{U}$. Since U is external and closed under addition, it follows that there is $K > \mathrm{U}$ with K/H infinitesimal such that

$$\frac{|B\oplus_H [-K,K]|}{H} \leq s+\frac{\delta}{2}.$$

We finish by showing that this K is as desired.

Let $\mathscr{I} := \{[iK,(i+1)K) \ : \ 0\leq i\leq H/K-1\}$ be a partition of $[0,H-1)$ into intervals of length K (with a negligible tail omitted). Let $X := \{i\in [0,H/K-1] \ : \ [iK,(i+1)K-1)\cap B = \emptyset\}$.

Claim 2: $\frac{|X|}{|\mathscr{I}|} > \frac{1}{3}$.

Proof of Claim 2: Suppose, towards a contradiction, that $\frac{|X|}{|\mathscr{I}|} \leq \frac{1}{3}$. Fix $i\notin X$ and $x\in [iK,(i+1)K)$. Write $x = iK+j$ with $j\in [0,K-1]$. Since $i\notin X$, there is $l\in [0,K-1)$ such that $iK+l\in B$. It follows that $x = (iK+l)+(j-l)\in B\oplus_H [-K,K]$. Consequently,

$$|B\oplus_H [-K,K]| \geq \sum_{i\notin X} K \geq \frac{2}{3}(H/K-1)\cdot K = \frac{2}{3}H - \frac{2}{3}K,$$

whence

$$\frac{|B\oplus_H [-K,K]|}{H} \geq \frac{2}{3} - \frac{2}{3}\frac{K}{H} \approx \frac{2}{3},$$

which, for sufficiently small ε and δ, contradicts the fact that $\frac{|B\oplus_H[-K,K]|}{H} \leq s+\frac{\delta}{2}$.

Let $\mathscr{I}' := \{[iK,(i+1)K) \ : \ i\notin X\}$. As explained above, the following claim completes the proof of the theorem.

Claim 3: There are $I,J\in \mathscr{I}$ such that

$$\mathrm{st}\left(\frac{|A\cap I|}{K}\right) > r-\varepsilon \text{ and } \mathrm{st}\left(\frac{|B\cap J|}{K}\right) > s+\delta.$$

Proof of Claim 3: We only prove the existence of J; the proof of the existence of I is similar (and easier). Suppose, towards a contradiction, that $\mathrm{st}(\frac{|B\cap J|}{K}) \leq s+\delta$ for all $J\in \mathscr{I}$. We then have

$$s-\delta < \frac{|B\cap [0,H-1)|}{H} = \frac{1}{H}\sum_{J\in \mathscr{I}'} |B\cap [iK,(i+1)K)| \leq \frac{1}{H}\cdot\frac{2}{3}\cdot (H/K)\cdot (s+\delta)K = \frac{2}{3}(s+\delta).$$

If $\delta \leq \frac{s}{5}$, then this yields a contradiction.

13.3 Beiglböck's proof

It is straightforward to verify that Corollary 13.9 is also true for subsets of $\mathbb{Z}$:

Corollary 13.10. *If* $A,B\subseteq \mathbb{Z}$ *are such that* $\mathrm{BD}(A),\mathrm{BD}(B) > 0$*, then* $A+B$ *is piecewise syndetic.*

In this section, we give Beiglböck's ultrafilter proof of Corollary 13.10 appearing in [3]. We first start with some preliminary facts on invariant means on $\mathbb{Z}$.

Definition 13.11. An *invariant mean* on $\mathbb{Z}$ is a linear functional $\ell : B(\mathbb{Z}) \to \mathbb{R}$ that satisfies the following properties:

1. ℓ is positive, that is, $\ell(f) \geq 0$ if $f \geq 0$;
2. $\ell(1) = 1$; and
3. $\ell(k.f) = \ell(f)$ for all $k \in \mathbb{Z}$ and $f \in B(\mathbb{Z})$, where $(k.f)(x) := f(x-k)$.

There are many invariant means on $\mathbb{Z}$:

Exercise 13.12. Suppose that (I_n) is a sequence of intervals in $\mathbb{Z}$ with $|I_n| \to \infty$ as $n \to \infty$. Fix $\mathscr{U} \in \beta\mathbb{N}$. Define, for $f \in B(\mathbb{Z})$, $\ell(f) = \lim_{\mathscr{U}}(\frac{1}{|I_n|}\sum_{x \in I_n} f(x))$. Show that ℓ is an invariant mean on $\mathbb{Z}$.

In fact, we have:

Lemma 13.13. *For every $A \subseteq \mathbb{Z}$, there is an invariant mean ℓ on $\mathbb{Z}$ such that $\ell(1_A) = \mathrm{BD}(A)$.*

Proof. Let (I_n) be a sequence of intervals witnessing the Banach density of A. Fix nonprincipal $\mathscr{U} \in \beta\mathbb{Z}$. Define ℓ as in Exercise 13.12 for these choices of (I_n) and $\mathscr{U}$. It is clear that $\ell(1_A) = \mathrm{BD}(A)$.

Lemma 13.14. *For every invariant mean ℓ on $\mathbb{Z}$, there is a regular Borel probability measure ν on $\beta\mathbb{Z}$ such that $\ell(1_A) = \nu(\overline{A})$ for every $A \subseteq \mathbb{Z}$.*

Proof. Fix a mean ℓ on $\mathbb{Z}$. Since $f \mapsto \beta f$ yields an isomorphism $B(\mathbb{Z}) \cong C(\beta\mathbb{Z})$, the Riesz Representation Theorem yields a regular Borel probability measure ν on $\beta\mathbb{Z}$ such that $\ell(f) = \int_{\beta\mathbb{Z}}(\beta f)d\nu$ for all $f \in B(\mathbb{Z})$. In particular,

$$\ell(1_A) = \int_{\beta\mathbb{Z}} (\beta 1_A) d\nu = \nu(\overline{A}).$$

The following lemma is the key to Beiglböck's proof of Corollary 13.10.

Lemma 13.15. *For any $A, B \subseteq \mathbb{Z}$, there is $\mathscr{U} \in \beta\mathbb{Z}$ such that* $\mathrm{BD}(A \cap (B - \mathscr{U})) \geq \mathrm{BD}(A) \cdot \mathrm{BD}(B)$.

Proof. Fix an invariant mean ℓ on $\mathbb{Z}$ such that $\ell(1_B) = \mathrm{BD}(B)$ and let ν be the associated Borel probability measure on $\beta\mathbb{Z}$. Let (I_n) be a sequence of intervals witnessing the Banach density of A. Define $f_n : \beta\mathbb{Z} \to [0,1]$ by

$$f_n(\mathscr{U}) := \delta((A \cap (B - \mathscr{U}), I_n) = \frac{1}{|I_n|} \sum_{k \in A \cap I_n} 1_{\overline{B-k}}(\mathscr{U}).$$

Set $f(\mathscr{U}) := \limsup_n f_n(\mathscr{U})$ and note that $f(\mathscr{U}) \leq \mathrm{BD}(A \cap (B - \mathscr{U}))$ for all $\mathscr{U} \in \beta\mathbb{Z}$. Fatou's Lemma implies

$$\int_{\beta\mathbb{Z}} f d\nu \geq \limsup_n \int_{\beta\mathbb{Z}} \frac{1}{|I_n|} \sum_{k \in A \cap I_n} 1_{U_{B-k}} d\nu = \limsup_n \frac{1}{|I_n|} \sum_{k \in I_n \cap A} \ell(1_{B-k}).$$

Since ℓ is invariant, the latter term is equal to $\limsup_n \delta(A, I_n) \cdot \ell(1_B) = \mathrm{BD}(A) \cdot \mathrm{BD}(B)$. Thus, we have shown $\int_{\beta\mathbb{Z}} f d\nu \geq \mathrm{BD}(A) \cdot \mathrm{BD}(B)$. In particular, there is some $\mathscr{U} \in \mathbb{Z}$ such that $f(\mathscr{U}) \geq \mathrm{BD}(A) \cdot \mathrm{BD}(B)$, as desired.

Notice that, in the notation of the above proof, $\mu(\mathbb{Z}) = 0$, whence we can take $\mathscr{U}$ as in the conclusion of the lemma to be nonprincipal.

We can now give Beiglböck's proof of Corollary 13.10. Assume that $\mathrm{BD}(A), \mathrm{BD}(B) > 0$. Apply the previous lemma with A replaced by $-A$ (which has the same Banach density), obtaining $\mathscr{U} \in \beta\mathbb{Z}$ such that $C := (-A) \cap (B - \mathscr{U})$ has positive Banach density. By Lemma 11.25, $C - C$ is syndetic; since $C - C \subseteq A + (B - \mathscr{U})$, we have that $A + (B - \mathscr{U})$ is also syndetic.

Suppose $s \in A + (B - \mathscr{U})$. Then for some $a \in A$, $B - (s - a) \in \mathscr{U}$, whence $a + B - s \in \mathscr{U}$ and hence $A + B - s \in \mathscr{U}$. Thus, for any finite set $s_1, \ldots, s_n \in A + (B - \mathscr{U})$, we have $\bigcap_{i=1}^n (A + B - s_i) \in \mathscr{U}$, and, in particular, is nonempty, meaning there is $t \in \mathbb{Z}$ such that $t + \{s_1, \ldots, s_n\} \subseteq A + B$. We claim that this implies that $A + B$ is piecewise syndetic. Indeed, take $F \subseteq \mathbb{Z}$ such that $F + A + (B - \mathscr{U}) = \mathbb{Z}$. We claim that $F + A + B$ contains arbitrarily long intervals. To see this, fix $n \in \mathbb{N}$ and, for $i = 1, \ldots, n$ take $s_i \in A + (B - \mathscr{U})$ such that $i \in F + s_i$. Take $t \in \mathbb{Z}$ such that $t + \{s_1, \ldots, s_n\} \subseteq A + B$. Then $t + [1, n] \subseteq t + F + \{s_1, \ldots, s_n\} \subseteq F + (A + B)$, completing the proof.

13.4 A proof with an explicit bound

A proof of Corollary 13.10 can be given by using a simple counting argument of finite combinatorics in the nonstandard setting. In this way, one also obtains an explicit bound on the number of shifts of the sumset that are needed to produce a thick set.

Lemma 13.16. *Let $C \subseteq [1,n]$ and $D \subseteq [1,m]$ be finite sets of natural numbers. Then there exists $k \leq n$ such that*

$$\frac{|(C-k)\cap D|}{m} \;\geq\; \frac{|C|}{n}\cdot\frac{|D|}{m} \;-\; \frac{|D|}{n}.$$

Proof. If $\chi : [1,n] \to \{0,1\}$ is the characteristic function of C, then for every $d \in D$, we have

$$\frac{1}{n}\cdot\sum_{k=1}^{n}\chi(k+d) \;=\; \frac{|C\cap[1+d,n+d]|}{n} \;=\; \frac{|C|}{n}+\frac{e(d)}{n}$$

where $|e(d)| \leq d$. Then:

$$\frac{1}{n}\cdot\sum_{k=1}^{n}\left(\frac{1}{m}\cdot\sum_{d\in D}\chi(k+d)\right) = \frac{1}{m}\cdot\sum_{d\in D}\left(\frac{1}{n}\cdot\sum_{x=1}^{n}\chi(k+d)\right)$$
$$= \frac{1}{m}\cdot\sum_{d\in D}\frac{|C|}{n} \;+\; \frac{1}{nm}\cdot\sum_{d\in D}e(d) \;=\; \frac{|C|}{n}\cdot\frac{|D|}{m} \;+\; e$$

where

$$|e| \;=\; \left|\frac{1}{nm}\sum_{d\in D}e(d)\right| \;\leq\; \frac{1}{nm}\sum_{d\in D}|e(d)| \;\leq\; \frac{1}{nm}\cdot\sum_{d\in D}d \;\leq\; \frac{1}{nm}\sum_{d\in D}m \;=\; \frac{|D|}{n}.$$

By the *pigeonhole principle*, there must exist at least one number $k \leq n$ such that

$$\frac{|(C-k)\cap D|}{m} \;=\; \frac{|(D+k)\cap C|}{m} \;=\; \frac{1}{m}\cdot\sum_{d\in D}\chi(k+d) \;\geq\; \frac{|C|}{n}\cdot\frac{|D|}{m} \;-\; \frac{|D|}{n}.$$

Theorem 13.17. *Let $A,B \subseteq \mathbb{Z}$ have positive Banach densities $BD(A) = \alpha > 0$ and $BD(B) = \beta > 0$. Then there exists a finite set F with $|F| \leq \frac{1}{\alpha\beta}$ such that $(A+B)+F$ is thick. In particular, $A+B$ is piecewise syndetic.*

Proof. Pick infinite $\nu, N \in {}^*\mathbb{N}$ such that $\nu/N \approx 0$, and pick intervals $[\Omega+1,\Omega+N]$ and $[\Xi+1,\Xi+\nu]$ such that

$$\frac{|{}^*A\cap[\Omega+1,\Omega+N]|}{N} \approx \alpha \quad\text{and}\quad \frac{|(-{}^*B)\cap[\Xi+1,\Xi+\nu]|}{\nu} \approx \beta.$$

By applying the nonstandard version of the previous lemma to the hyperfinite sets $C = ({}^*A-\Omega)\cap[1,N] \subseteq [1,N]$ and $D = (-{}^*B-\Xi)\cap[1,\nu]$, one obtains the existence of a number ζ such that

$$\frac{|(C-\zeta)\cap D|}{\nu} \;\geq\; \frac{|C|}{N}\cdot\frac{|D|}{\nu} \;-\; \frac{|D|}{N} \;\approx\; \alpha\beta.$$

Finally, apply Lemma 11.26 to the internal set $E = (C-\zeta)\cap D \subseteq [1,\nu]$. Since $|E|/\nu \approx \alpha\beta$, there exists a finite $F \subset \mathbb{Z}$ with $|F| \leq \frac{1}{\alpha\beta}$ and such that $\mathbb{Z} \subseteq (E-E)+F$, and hence, by *overflow*, $I \subseteq (E-E)+F$ for some infinite interval I. Since $E \subseteq {}^*A-\Omega$ and $E \subset -{}^*B-\Xi$, it follows that ${}^*(A+B+F) = {}^*A+{}^*B+F$ includes the infinite interval $I+\Omega+\Xi+\zeta$, and hence it is thick.

13.5 Quantitative strengthenings

We end this chapter by proving some technical strengthenings of Corollary 13.10. Indeed, in light of Lemma 11.23, the following theorem can be viewed as a "quantitative" strengthening of Corollary 13.10:

Theorem 13.18. *Suppose that (I_n) is a sequence of intervals with $|I_n| \to \infty$ as $n \to \infty$. Suppose that $A, B \subseteq \mathbb{Z}$ and $\mathrm{BD}(B) > 0$. Then:*

1. *If $\overline{d}_{(I_n)}(A) \geq r$, then there is a finite set $F \subseteq \mathbb{Z}$ such that, for every finite set $L \subseteq \mathbb{Z}$, we have*

$$\overline{d}_{(I_n)}\left(\bigcap_{x\in L}(A+B+F+x)\right) \geq r.$$

2. *If $\underline{d}_{(I_n)}(A) \geq r$, then for every $\varepsilon > 0$, there is a finite set $F \subseteq \mathbb{Z}$ such that, for every finite set $L \subseteq \mathbb{Z}$, we have*

$$\underline{d}_{(I_n)}\left(\bigcap_{x\in L}(A+B+F+x)\right) \geq r-\varepsilon.$$

In connection with item (2) of the previous theorem, it will turn out that F depends only on B and ε (but not on A or (I_n)). Moreover, item (2) is false if $r-\varepsilon$ is replaced by r; see [11].

In order to prove Theorem 13.18, we need a preparatory counting lemma.

Lemma 13.19. *Suppose that (I_n) is a sequence of intervals in $\mathbb{Z}$ such that $|I_n| \to \infty$ as $n \to \infty$. Further suppose I is an infinite hyperfinite interval in ${}^*\mathbb{Z}$ and $A \subseteq \mathbb{Z}$.*

1. *If $\overline{d}_{(I_n)}(A) \geq r$, then there is $N > \mathbb{N}$ such that*

$$\delta({}^*A, I_N) \gtrsim r \quad \textit{and} \quad \frac{1}{|I_N|}\sum_{x\in I_N}\delta(x-({}^*A\cap I_N), I) \gtrsim r. \quad (\dagger)$$

2. *If $\underline{d}_{(I_n)}(A) > r$, then there is $N_0 > \mathbb{N}$ such that $(\dagger)$ holds for all $N \geq N_0$.*

Proof. For (1), first apply transfer to the statement "for every finite interval $J \subseteq \mathbb{Z}$ and every natural number k, there exists $n \geq k$ such that

$$\delta(A, I_n) > r - 2^{-k} \quad \text{and} \quad \frac{1}{|I_n|}\sum_{x\in J}|(I_n - x)\triangle I_n| < 2^{-k}."$$

Fix $K > \mathbb{N}$ and let N be the result of applying the transferred statement to I and K. Set $C = {}^*A \cap I_N$ and let χ_C denote the characteristic function of C. We have

$$\begin{aligned}
\frac{1}{|I_N|}\sum_{x\in I_N}\delta((x-C),Y) &= \frac{1}{|I_N|}\sum_{x\in I_N}\frac{1}{|I|}\sum_{y\in I}\chi_C(x-y)\\
&= \frac{1}{|I|}\sum_{y\in I}\frac{|C\cap(I_N-y)|}{|I_N|}\\
&\geq \frac{|C|}{|I_N|} - \sum_{y\in I}\frac{|(I_N-y)\triangle I_N|}{|I_N|}\\
&\approx r.
\end{aligned}$$

For (2), apply transfer to the statement "for every finite interval $J \subseteq \mathbb{Z}$ and every natural number k, there exists $n_0 \geq k$ such that, for all $n \geq n_0$,

$$\delta(A, I_n) > r - 2^{-n_0} \quad \text{and} \quad \frac{1}{|I_n|}\sum_{x\in J}|(I_n - x)\triangle I_n| < 2^{-n_0}."$$

Once again, fix $K > \mathbb{N}$ and let N_0 be the result of applying the transferred statement to I and K. As above, this N_0 is as desired.

Proof (of Theorem 13.18). Fix an infinite hyperfinite interval I that is good for B. (See Proposition 11.28.)

For (1), assume that $\overline{d}_{(I_n)}(A) \geq r$. Let N be as in part (1) of Lemma 13.19 applied to I and A. Once again, set $C := {}^*A \cap I_N$. Consider the μ_{I_N}-measurable function

$$f(x) = \mathrm{st}(\delta(x - C, I)).$$

By Lemma 6.18, we have that

$$\int_{I_N} f d\mu_{I_N} = \mathrm{st}\left(\frac{1}{|I_N|}\sum_{x\in I_N}\delta(x-C,I)\right) \geq r,$$

whence there is some standard $s > 0$ such that $\mu_{I_N}(\{x \in I_N : f(x) \geq 2s\}) \geq r$. Setting $\Gamma = \{x \in I_N : \delta(x - C, I) \geq s\}$, we have that $\mu_{I_N}(\Gamma) \geq r$. Since I is good for B, we may take a finite subset F of $\mathbb{Z}$ such that

$$\delta({}^*(B+F), I) > 1 - \frac{s}{2}.$$

Fix $x \in \mathbb{Z}$. Since I is infinite, we have that

$$\delta({}^*(B+F+x), I) = \delta({}^*(B+F), (I-x)) \approx \delta({}^*(B+F), I),$$

whence $\delta({}^*(B+F+x), I) > 1 - s$. Thus, for any $y \in \Gamma$, we have that $(y - C) \cap {}^*(B+F+x) \neq \emptyset$. In particular, if L is a finite subset of $\mathbb{Z}$, then $\Gamma \subseteq {}^*\left(\bigcap_{x\in L} A+B+F+x\right)$. Therefore

$$\overline{d}_{(I_n)}\left(\bigcap_{x\in L} A+B+F+x\right) \geq \mu_{I_N}\left({}^*(\bigcap_{x\in L} A+B+F+x)\right) \geq \mu_{I_N}(\Gamma) \geq r.$$

This establishes (1).

Towards (2), note that we may suppose that $\underline{d}_{(I_n)}(A) > r$. Fix $N_0 > \mathbb{N}$ as in part (2) of Lemma 13.19 applied to I and A. Fix $N \geq N_0$ and standard $\varepsilon > 0$ with $\varepsilon < r$. Set

$$\Lambda := \{x \in I_N : \delta((x - C), I) \geq \varepsilon\}$$

and observe that $\frac{|\Lambda|}{|I_N|} > r - \varepsilon$. Since I is good for B, we may fix a finite subset F of $\mathbb{Z}$ such that

$$\delta({}^*(B+F), I) > 1 - \frac{\varepsilon}{2}.$$

Fix $x \in \mathbb{Z}$. Since I is infinite, arguing as in the proof of part (1), we conclude that

$$\delta^*(B+F+x), I) > 1 - \varepsilon.$$

Fix $L \subseteq \mathbb{Z}$ finite. As in the proof of part (1), it follows that $\Lambda \subseteq {}^*\left(\bigcap_{x\in L} A+B+F+x\right)$ whence

$$\delta\left({}^*\left(\bigcap_{x\in L} A+B+F+x\right), I_N\right) \geq \frac{|\Lambda|}{|I_N|} > r - \varepsilon.$$

Since the previous inequality held for every $N \geq N_0$, by transfer we can conclude that there is n_0 such that, for all $n \geq n_0$, we have

$$\delta\left(\left(\bigcap_{x\in L} A+B+F+x\right), I_n\right) \geq r - \varepsilon,$$

whence it follows that

$$\underline{d}_{(I_n)}\left(\bigcap_{x\in L} A+B+F+x\right) \geq r - \varepsilon.$$

We remark that the original proof of Theorem 13.18 given in [12] used a Lebesgue Density Theorem for the cut spaces $[0,H]/\mathrm{U}$. Indeed, one can give a nice proof of Theorem 13.6 using the standard Lebesgue density theorem and Example 13.4 suggested that perhaps a general Lebesgue density theorem holds for cut spaces. Once this was established, the fact that one has many density points was used to strengthen the sumset theorem in the above manner. The proof given in this section follows [11], which actually works for all countable amenable groups rather than just $\mathbb{Z}$; other than the fact that Proposition 11.10 is more difficult to prove for amenable groups than it is for $\mathbb{Z}$, there is not much added difficulty in generalizing to the amenable situation. We should also mention that the amenable group version of Corollary 13.10 was first proven by Beiglböck, Bergelson, and Fish in [4].

Chapter 14
Sumset configurations in sets of positive density

14.1 Erdős' conjecture

Just as Szemeredi's theorem is a "density" version of van der Waerden's theorem, it is natural to wonder if the density version of Hindman's theorem is true, namely: does every set of positive density contain an FS set? It is clear that the answer to this question is: no! Indeed, the set of odd numbers has positive density, but does not even contain $\mathrm{PS}(B)$ for any infinite set B. Here, $\mathrm{PS}(B) := \{b+b' \ : \ b,b' \in B,\ b \neq b'\}$. This example is easily fixed if we allow ourselves to translate the original set, so Erdős conjectured that this was the only obstruction to a weak density version of Hindman's theorem, namely: if $A \subseteq \mathbb{N}$ has positive density, then there is $t \in \mathbb{N}$ and infinite $B \subseteq A$ such that $t + \mathrm{PS}(B) \subseteq A$. Straus provided a counterexample to this conjecture (see [15]), whence Erdős changed his conjecture to the following, which we often refer to as Erdős' sumset conjecture (see [39] and [16, page 85]):

Conjecture 14.1. Suppose that $A \subseteq \mathbb{N}$ is such that $\underline{d}(A) > 0$. Then there exist infinite sets B and C such that $B + C \subseteq A$.

We should remark that no counterexample to Erdős' conjecture is known if one merely assumes positive Banach density as opposed to positive lower density. The first progress on Erdős' conjecture was made by Nathanson in [39], where he proved that if $\overline{d}(A) > 0$, then, for any n, there are $B, C \subseteq \mathbb{N}$ with B infinite and $|C| \geq n$ such that $B + C \subseteq A$. Nathanson's result will be a byproduct of our methods in the next section, although Nathanson's technique allows one to obtain B with positive upper density (while ours do not).

It will be convenient to give a name to sets satisfying the conclusion of Erdős' conjecture.

Definition 14.2. We say that $A \subseteq \mathbb{N}$ has the *sumset property* if there are infinite sets $B, C \subseteq \mathbb{N}$ such that $B + C \subseteq A$.

Many sets that are structurally large have the sumset property as indicated by the following proposition. While this result follows from standard results in the literature, we prefer to give the following elegant argument of Leth.

Proposition 14.3. *If A is piecewise syndetic, then A has the sumset property. More precisely, there is an infinite set $B \subseteq \mathbb{N}$ and $k \in \mathbb{N}$ such that* $\mathrm{PS}(B) - k \subseteq A$.

Proof. Since A is piecewise syndetic, there exists m and an interval $[a,b]$ in ${}^*\mathbb{N}$ with a and $b-a$ infinite such that ${}^*A$ has no gaps of size larger than m on $[a,b]$. Set $L := ({}^*A - a) \cap \mathbb{N}$, so that $a + L \subseteq {}^*A$. Let l be the first element in ${}^*L$ greater than or equal to a. Set $k := l - a$. Since L contains no gaps of size larger than m, we know that $0 \leq k \leq m$. We now have:

$$l - k + L \subseteq {}^*A \text{ and } l \in {}^*L.$$

Take $b_0 \in L$ arbitrary. Assume now that $b_0 < b_1 < \cdots < b_n \in L$ have been chosen so that $b_i + b_j - k \in A$ for $1 \leq i < j \leq n$. Since the statement "there is $l \in {}^*L$ such that $l > b_n$ and $l - k + b_i \in {}^*A$ for $i = 1, \ldots, n$" is true, by transfer there is $b_{n+1} \in L$ such that $b_{n+1} > b_n$ and $b_i + b_{n-1} - k \in A$ for $i = 1, \ldots, n$. The set $B := \{b_0, b_1, b_2, \ldots\}$ defined this way is as desired.

We next establish a nonstandard reformulation of the sumset property. We will actually need the following more general statement:

Proposition 14.4. *Given $A \subseteq \mathbb{N}$ and $k \in \mathbb{Z}$, the following are equivalent:*

1. *there exists $B = \{b_1 < b_2 < \cdots\}$ and $C = \{c_1 < c_2 < \cdots\}$ such that $b_i + c_j \in A$ for $i \leq j$ and $b_i + c_j \in A + k$ for $i > j$;*
2. *there exist nonprincipal ultrafilters $\mathscr{U}$ and $\mathscr{V}$ on $\mathbb{N}$ such that $A \in \mathscr{U} \oplus \mathscr{V}$ and $A + k \in \mathscr{V} \oplus \mathscr{U}$;*
3. *there exist infinite $\beta, \gamma \in {}^*\mathbb{N}$ such that $\beta + {}^*\gamma \in {}^{**}A$ and $\gamma +^* \beta \in {}^{**}A + k$.*

Proof. First suppose that (1) holds as witnessed by B and C. By assumption, the collection of sets

$$\{B\} \cup \{A - c \ : \ c \in C\}$$

has the finite intersection property with the Frechet filter, whence there is a nonprincipal ultrafilter $\mathscr{U}$ on $\mathbb{N}$ extending this family. Likewise, there is a nonprincipal ultrafilter $\mathscr{V}$ on $\mathbb{N}$ extending the family $\{C - k\} \cup \{A - b \ : \ b \in B\}$. These $\mathscr{U}$ and $\mathscr{V}$ are as desired.

Next, given (2), take $\beta, \gamma \in {}^*\mathbb{N}$ such that $\mathscr{U} = \mathscr{U}_\beta$ and $\mathscr{V} = \mathscr{U}_\gamma$. These β and γ are as desired.

Finally, suppose that $\beta, \gamma \in {}^*\mathbb{N}$ are as in (3). We define $B = \{b_1 < b_2 < b_3 < \cdots\}$ and $C = \{c_1 < c_2 < c_3 < \cdots\}$ recursively as follows. Suppose that b_i and c_j for $i, j = 1, \ldots, n$ have been constructed so that, for all i, j we have:

- $b_i + c_j \in A$ if $i \leq j$;
- $b_i + c_j \in A + k$ if $i > j$;
- $b_i + \gamma \in {}^*A$;
- $c_j + \beta \in {}^*A + k$.

Applying transfer to the statement "there is $x \in {}^*\mathbb{N}$ such that $x + c_j \in {}^*A + k$ for $j = 1, \ldots, n$ and $x > b_n$ and $x + {}^*\gamma \in {}^{**}A$" (which is witnessed by β), we get $b_{n+1} \in \mathbb{N}$ such that $b_{n+1} > b_n$, $b_{n+1} + c_j \in A + k$ for $j = 1, \ldots, n$ and for which $b_{n+1} + \gamma \in {}^*A$. Next, apply transfer to the statement "there is $y \in {}^*\mathbb{N}$ such that $b_i + y \in {}^*A$ for $i = 1, \ldots, n+1$ and $y > c_n$ and $y + {}^*\beta \in {}^{**}A + k$" (which is witnessed by γ), we get $c_{n+1} \in \mathbb{N}$ such that $c_{n+1} > c_n$ and for which $b_i + c_{n+1} \in A$ for $i = 1, \ldots, n+1$ and for which $c_{n+1} + \beta \in {}^*A$. This completes the recursive construction.

Taking $k = 0$ in the previous proposition yields a nonstandard reformulation of the sumset property.

Corollary 14.5. *Given $A \subseteq \mathbb{N}$, the following are equivalent:*

1. *A has the sumset property;*
2. *there exist nonprincipal ultrafilters $\mathscr{U}$ and $\mathscr{V}$ on $\mathbb{N}$ such that $A \in (\mathscr{U} \oplus \mathscr{V}) \cap (\mathscr{V} \oplus \mathscr{U})$;*
3. *there exist infinite $\xi, \eta \in {}^*\mathbb{N}$ such that $\xi + {}^*\eta, \eta +^* \xi \in {}^{**}A$.*

14.2 A 1-shift version of Erdős' conjecture

The main result of this chapter, due to Di Nasso, Goldbring, Jin, Leth, Lupini, and Mahlburg [13], is that a set of large Banach density satisfies the conclusion of Erdős' conjecture.

Theorem 14.6. *If $\mathrm{BD}(A) > \frac{1}{2}$, then A has the sumset property.*

Before proving Theorem 14.6, let us show how it, together with Corollary 14.5, implies that sets of positive Banach density satisfy a "1-shift" version of Erdős' conjecture. This implication was first proven in [13] using Ramsey's theorem, whereas we give a proof here using the techniques described at the end of the previous section.

Corollary 14.7. *Suppose that $\mathrm{BD}(A) > 0$. Then there exists $B = \{b_1 < b_2 < \cdots\}$, $C = \{c_1 < c_2 < \cdots\}$, and $k \in \mathbb{N}$ such that $b_i + c_j \in A$ for $i \leq j$ and $b_i + c_j \in A + k$ for $i > j$.*

Proof. By Proposition 11.10, we may fix $n \in \mathbb{N}$ such that $\mathrm{BD}(A + [-n, n]) > \frac{1}{2}$. By Theorem 14.6 and Corollary 14.5, we may take infinite $\beta, \gamma \in {}^*\mathbb{N}$ such that $\beta + {}^*\gamma, \gamma +^* \beta \in {}^{**}A + [-n, n]$. Take $i, j \in [-n, n]$ such that $\beta + {}^*\gamma \in {}^{**}A + i$ and $\gamma +^* \beta \in {}^{**}A + j$. Without loss of generality, $i < j$. Set $k := j - i$. Then $\beta + {}^*(\gamma - i) \in {}^{**}A$ and $(\gamma - i) + {}^*\beta \in {}^{**}A + k$, whence the conclusion holds by Proposition 14.4.

In order to prove Theorem 14.6, we need one technical lemma:

Lemma 14.8. *Suppose that* $\mathrm{BD}(A) = r > 0$. *Suppose further that* (I_n) *is a sequence of intervals with witnessing the Banach density of* A. *Then there is* $L \subseteq \mathbb{N}$ *satisfying:*

1. $\limsup_{n\to\infty} \frac{|L\cap I_n|}{|I_n|} \geq r$;
2. *for all finite* $F \subseteq L$, $A \cap \bigcap_{x\in F}(A - x)$ *is infinite*

Proof. First, we note that it suffices to find L satisfying (1) and

(2') there is $x_0 \in {}^*A \setminus A$ such that $x_0 + L \subseteq {}^*A$.

Indeed, given finite $F \subseteq L$ and $K \subseteq \mathbb{N}$, x_0 witnesses the truth of "there exists $x \in {}^*\mathbb{N}$ such that $x + F \subseteq {}^*A$ and $x \notin K$" whence, by transfer, such an x can be found in $\mathbb{N}$, establishing (2).

In the rest of the proof, we fix infinite $H \in {}^*\mathbb{N}$ and let μ denote Loeb measure on I_H. In addition, for any $\alpha \in {}^*\mathbb{N}$ and hyperfinite $X \subseteq {}^*\mathbb{N}$, we set $d_\alpha(X) := \frac{|X|}{|I_\alpha|}$. Finally, we fix $\varepsilon \in (0, \frac{1}{2})$.

Next we remark that it suffices to find a sequence $X_1, X_2, \ldots$ of internal subsets of I_H and an increasing sequence $n_1 < n_2 < \cdots$ of natural numbers such that, for each i, we have:

(i) $\mu(X_i) \geq 1 - \varepsilon^i$ and,
(ii) for each $x \in X_i$, we have $d_{n_i}({}^*A \cap (x + I_{n_i})) \geq r - \frac{1}{i}$.

Indeed, suppose that this has been accomplished and set $X := \bigcap_i X_i$. Then X is Loeb measurable and $\mu(X) > 0$. Fix $y_0 \in X \setminus \mathbb{N}$ arbitrary and set x_0 to be the minimum element of ${}^*A$ that is greater than or equal to y_0; note that $x_0 - y_0 \in \mathbb{N}$ since $y_0 \in X$. Set $L := ({}^*A - x_0) \cap \mathbb{N}$; note that (2') is trivially satisfied. To see that (1) holds, note that

$$\limsup_{i\to\infty} d_{n_i}(L \cap I_{n_i}) = \limsup_{i\to\infty} d_{n_i}({}^*A \cap (x_0 + I_{n_i})) = \limsup_{i\to\infty} d_{n_i}({}^*A \cap (y_0 + I_{n_i})) \geq r,$$

where the last inequality follows from the fact that $y_0 \in X$.

Thus, to finish the lemma, it suffices to construct the sequences (X_i) and (n_i). Suppose that $X_1, \ldots, X_{i-1}$ and $n_1 < \cdots < n_{i-1}$ have been constructed satisfying the conditions above. For $\alpha \in {}^*\mathbb{N}$, set

$$Y_\alpha := \{x \in I_H \ : \ d_\alpha({}^*A \cap (x + I_m)) \geq r - \frac{1}{i}\}.$$

Set $Z := \{\alpha \in {}^*\mathbb{N} \ : \ n_{i-1} < \alpha \text{ and } d_H(Y_\alpha) > 1 - \varepsilon^i\}$. Note that Z is internal. It will be enough to show that Z contains all sufficiently small infinite elements of ${}^*\mathbb{N}$, for then, by underflow, there is $n_i \in Z \cap \mathbb{N}$. Setting $X_i := Y_{n_i}$, these choices of X_i and n_i will be as desired.

We now work towards proving that Z contains all sufficiently small infinite elements of ${}^*\mathbb{N}$. First, we remark that we may assume, without loss of generality, that the sequences $(|I_n|)$ and (b_n) are increasing, where b_n denotes the right endpoint of I_n. Fix $K \in {}^*\mathbb{N} \setminus \mathbb{N}$ such that $2b_K/|I_H| \approx 0$. We finish the proof of the lemma by proving that $K \in Z$, which we claim follows from the following two facts:

(a) for all $x \in I_H$, $\mathrm{st}(d_K({}^*A \cap (x + I_K))) \leq r$;
(b) $\frac{1}{|I_H|}\sum_{x\in I_H} d_K({}^*A \cap (x + I_K)) \approx r$.

To see that these facts imply that $K \in Z$, for $x \in I_H$, set $f(x) := d_K({}^*A \cap (x + I_K))$. It is enough to show that $f(x) \approx r$ for μ-almost all $x \in I_H$. Given n, let $A_n := \{x \in I_H \ : \ f(x) < r - \frac{1}{n}\}$. Suppose, towards a contradiction, that $\mu(A_n) = s > 0$. By (a), we may fix a positive infinitesimal η such that $f(x) \leq r + \eta$ for all $x \in I_H$. We then have

$$\frac{1}{|I_H|}\sum_{x\in I_H} f(x) = \frac{1}{|I_H|}\left[\sum_{x\in A_n} f(x) + \sum_{x\notin A_n} f(x)\right] < s(r - \frac{1}{n}) + (1-s)(r+\eta).$$

Since the right-hand side of the above display is appreciably less than s, we get a contradiction to (b).

It remains to establish (a) and (b). (a) follows immediately from the fact that $\mathrm{BD}(A) = r$. To see (b), we first observe that

$$\frac{1}{|I_H|}\sum_{x\in I_H} d_K({}^*A\cap(x+I_K)) = \frac{1}{|I_K|}\sum_{y\in I_K}\frac{1}{|I_H|}\sum_{x\in I_H}\chi_{{}^*A}(x+y).$$

Fix $y\in I_K$. Since $|\sum_{x\in I_H}\chi_{{}^*A}(x+y)-|{}^*A\cap I_H||\leq 2y\leq 2b_K$, we have that

$$\left|\frac{1}{|I_H|}\sum_{x\in I_H}\chi_{{}^*A}(x+y)-d_H({}^*A)\right|\approx 0.$$

Since a hyperfinite average of infinitesimals is infinitesimal, we see that

$$\frac{1}{|I_H|}\sum_{x\in I_H} d_K({}^*A\cap(x+I_K))\approx\frac{1}{|I_K|}\sum_{y\in I_K} d_H({}^*A)\approx r,$$

establishing (b).

Proof (of Theorem 14.6). Set $r:=\mathrm{BD}(A)$. Let (I_n) witness the Banach density of A and let $L:=(l_n)$ be as in the previous lemma. We may then define an increasing sequence $D:=(d_n)$ contained in A such that $l_i+d_n\in A$ for $i\leq n$.[1] Now take N such that $\mu_{I_N}({}^*L)\geq r$. Note also that $\mu_{I_N}({}^*A-d_n)\geq r$ for any n. Since $r>1/2$, for any n we have that $\mu_{I_N}({}^*L\cap({}^*A-d_n))\geq 2r-1>0$. By a standard measure theory fact, by passing to a subsequence of D if necessary, we may assume that, for each n, we have that $\mu_{I_N}({}^*L\cap\bigcap_{i\leq n}({}^*A-d_i))>0$. In particular, for every n, we have that $L\cap\bigcap_{i\leq n}(A-d_i)$ is infinite.

We may now conclude as follow. Fix $b_1\in L$ arbitrary and take $c_1\in D$ such that $b_1+c_1\in A$. Now assume that $b_1<\cdots<b_n$ and $c_1<\cdots<c_n$ are taken from L and D respectively such that $b_i+c_j\in A$ for all $i,j=1,\ldots,n$. By assumption, we may find $b_{n+1}\in L\cap\bigcap_{i\leq n}(A-c_i)$ with $b_{n+1}>b_n$ and then we may take $c_{n+1}\in D$ such that $b_i+c_{n+1}\in A$ for $i=1,\ldots,n+1$.

14.3 A weak density version of Folkman's theorem

At the beginning of this chapter, we discussed the fact that the density version of Hindman's theorem is false. In fact, the odd numbers also show that the density version of Folkman's theorem is also false. (Recall that Folkman's theorem stated that for any finite coloring of $\mathbb{N}$, there are arbitrarily large finite sets G such that $\mathrm{FS}(G)$ are monochromatic.) However, we can use Lemma 14.8 to prove a weak density version of Folkman's theorem. Indeed, the proof of Lemma 14.8 yields the following:

Lemma 14.9. *Suppose that $A\subseteq\mathbb{N}$ is such that $\mathrm{BD}(A)\geq r$. Then there is $\alpha\in{}^*A\setminus A$ such that $\mathrm{BD}(A-\alpha)\geq r$.*

One should compare the previous lemma with Beiglbock's Lemma 13.15. Indeed, a special case of (the nonstandard formulation of) Lemma 13.15 yields $\alpha\in{}^*\mathbb{N}\setminus\mathbb{N}$ such that $\mathrm{BD}({}^*A-\alpha)\geq\mathrm{BD}(A)$; the previous lemma is stronger in that it allows us to find $\alpha\in{}^*A$. We can now prove the aforementioned weak version of a density Folkman theorem.

Theorem 14.10. *Fix $k\in\mathbb{N}$ and suppose $A\subseteq\mathbb{N}$ is such that $\mathrm{BD}(A)>0$. Then there exist increasing sequences $(x_n^{(i)})$ for $i=0,1,2,\ldots,k$ such that, for any i and any $n_i\leq n_{i+1}\leq\cdots\leq n_k$, we have $x_{n_i}^{(i)}+x_{n_{i+1}}^{(i+1)}+\cdots+x_{n_k}^{(k)}\in A$.*

The reason we think of the previous theorem as a weak density version of Folkman's theorem is that if all of the sequences were identical, then we would in particular have a set of size k all of whose finite sums belong to A.

Proof (of Theorem 14.10). Set $A=A^{(k)}$. Repeatedly applying Lemma 14.9, one can define, for $i=0,1,\ldots,k$, subsets $A^{(i)}$ of $\mathbb{N}$ and $\alpha_i\in{}^*A^{(i)}$ such that $A^{(i)}+\alpha_{i+1}\subseteq{}^*A^{(i+1)}$for all $i<k$. We then define the sequences $(x_n^{(i)})$ for $i=0,1,2,\ldots,k$ and finite subsets $A_n^{(i)}$ of $A^{(i)}$ so that:

- for $i=0,1,\ldots,k$ and any n, we have $x_n^{(i)}\in A_n^{(i)}$,

[1] Notice that at this point we already have obtained Nathanson's result mentioned in the previous section: if we set $B:=\{d_n,d_{n+1},\ldots\}$ and $C:=\{l_1,\ldots,l_n\}$, then $B+C\subseteq A$.

- for $i = 0, 1, \dots, k$ and any $n \leq m$, we have $A_n^{(i)} \subseteq A_m^{(i)}$, and
- for $i = 0, 1, \dots, k-1$ and any $n \leq m$, we have $A_n^{(i)} + x_m^{(i+1)} \subseteq A_m^{(i+1)}$.

It is clear that the sequences $(x_n^{(i)})$ defined in this manner satisfy the conclusion of the theorem. Suppose that the sequences $(x_n^{(i)})$ and $A_n^{(i)}$ have been defined for $n < m$. We now define $x_m^{(i)}$ and $A_m^{(i)}$ by recursion for $i = 0, 1, \dots, k$. We set $x_m^{(0)}$ to be any member of $A^{(0)}$ larger than $x_{m-1}^{(0)}$ and set $A_m^{(0)} := A_{m-1}^{(0)} \cup \{x_{m-1}^{(0)}\}$. Supposing that the construction has been carried out up through $i < k$, by transfer of the fact that $A_m^{(i)} + \alpha^{(i+1)} \subseteq {}^*A^{(i+1)}$, we can find $x_m^{(i+1)} \in A^{(i+1)}$ larger than $x_{m-1}^{(i+1)}$ such that $A_m^{(i)} + x_m^{(i+1)} \subseteq A^{(i+1)}$. We then define $A_m^{(i+1)} := A_{m-1}^{(i+1)} \cup (A_m^{(i)} + x_m^{(i+1)})$. This completes the recursive construction and the proof of the theorem.

The usual compactness argument gives a finitary version:

Corollary 14.11. *Suppose that $k \in \mathbb{N}$ and $\varepsilon > 0$ are given. Then there exists m such that for any interval I of length at least m and any subset A of I such that $|A| > \varepsilon |I|$, there exist $(x_n^{(i)})$ for $i, n \in \{0, 1, \dots, k\}$ such that $x_{n_i}^{(i)} + x_{n_{i+1}}^{(i+1)} + \cdots + x_{n_{\ell-1}}^{(k)} \in A$ for any $i = 0, 1, \dots, k$ and any $0 \leq n_i \leq n_{i+1} \leq \cdots \leq n_{\ell-1} \leq k$.*

Chapter 15
Near arithmetic progressions in sparse sets

15.1 The main theorem

Szemeredi's theorem says that relatively dense sets contain arithmetic progressions. The purpose of this chapter is to present a result of Leth from [34] which shows that certain sparse sets contain "near" arithmetic progressions. Our first task is to make precise what "near" means in the previous sentence.

Definition 15.1. Fix $w \in \mathbb{N}_0$ and $t, d \in \mathbb{N}$.[1] A (t,d,w)*-progression* is a set of the form

$$\mathrm{B}(b,t,d,w) := \bigcup_{i=0}^{t-1} [b+id, b+id+w].$$

By a *block progression* we mean a (t,d,w)-progression for some t,d,w.

Note that a $(t,d,0)$-progression is the same thing as a t-term arithmetic progression with difference d.

Definition 15.2. If $A \subseteq \mathbb{N}$, we say that *A nearly contains a* (t,d,w)*-progression* if there is a (t,d,w)-progression $\mathrm{B}(b,t,d,w)$ such that $A \cap [b+id, b+id+w] \neq \emptyset$ for each $i = 1, \ldots, t-1$.

Thus, if A nearly contains a $(t,d,0)$-progression, then A actually contains a t-term arithmetic progression. Consequently, when A nearly contains a (t,d,w)-progression with "small" w, then this says that A is "close" to containing an arithmetic progression. The main result of this chapter allows us to conclude that even relatively sparse sets with a certain amount of density regularity nearly contain block progressions satisfying a further homogeneity assumption that we now describe.

Definition 15.3. Suppose that $A \subseteq \mathbb{N}$, I is an interval in $\mathbb{N}$, and $0 < s < 1$. We say that *A nearly contains a* (t,d,w)*-progression in I with homogeneity s* if there is some $\mathrm{B}(b,t,d,w)$ contained in I such that the following two conditions hold for all $i, j = 0, 1, \ldots, t-1$:

(i) $\delta(A, [b+id, b+id+w]) \geq (1-s)\delta(A, I)$
(ii) $\delta(A, [b+id, b+id+w]) \geq (1-s)\delta(A, [b+jd, b+jd+w])$.

Thus, for small s, we see that A meets each block in a density that is roughly the same throughout and that is roughly the same as on the entire interval.

The density regularity condition roughly requires that on sufficiently large subintervals of I, the density does not increase too rapidly. Here is the precise formulation:

Definition 15.4. Suppose that $I \subseteq \mathbb{N}$ is an interval, $r \in \mathbb{R}^{>1}$, and $m \in \mathbb{N}$. We say that $A \subseteq I$ has the (m,r)*-density property on I* if, whenever $J \subseteq I$ is an interval with $|J|/|I| \geq 1/m$, then $\delta(A,J) \leq r\delta(A,I)$.

[1] In this chapter, we deviate somewhat from our conventions so as to match up with the notation from [34].

Of course, given any $m \in \mathbb{N}$ and $A \subseteq I$, there is $r \in \mathbb{R}^{>1}$ such that A has the (m,r)-density property on I. The notion becomes interesting when we think of r as fixed.

Given a hyperfinite interval $I \subseteq {}^*\mathbb{N}$, $r \in {}^*\mathbb{R}^{>1}$ and $M \in {}^*\mathbb{N}$, we say that an internal set $A \subseteq I$ has the *internal (M,r)-density property on I* if the conclusion of the definition above holds for internal subintervals J of I.

Lemma 15.5. *Suppose that $A \subseteq [1,N]$ is an internal set with the internal (M,r)-density property for some $M > \mathbb{N}$. Let $f : [0,1] \to [0,1]$ be the (standard) function given by*

$$f(x) := \mathrm{st}\left(\frac{|A \cap [1,xN]|}{|A \cap [1,N]|}\right).$$

Then f is a Lipschitz function with Lipschitz constant r.

Proof. Fix $x < y$ in $[0,1]$. Write $x := \mathrm{st}(K/N)$ and $y := \mathrm{st}(L/N)$. Since $y - x \neq 0$, we have that $\frac{L-K}{N}$ is not infinitesimal; in particular, $\frac{L-K}{N} > 1/M$. Since A has the (M,r)-density property on $[1,N]$, we have that $\delta(A,[K,L]) \leq r\delta(A,[1,N])$. Thus, it follows that

$$f(y) - f(x) = \mathrm{st}\left(\frac{|A \cap [K,L]|}{|A \cap [1,N]|}\right) = \mathrm{st}\left(\delta(A,[K,L]\frac{L-K}{|A \cap [1,N]|}\right) \leq r\,\mathrm{st}\left(\frac{L-K}{N}\right) = r(y-x).$$

Here is the main result of this section:

Theorem 15.6 (Leth). *Fix functions $g,h : \mathbb{R}^{>0} \to \mathbb{R}^{>0}$ such that h is increasing and $g(x) \to \infty$ as $x \to \infty$. Fix also $s > 0$, $r > 1$, and $j,t \in \mathbb{N}$. Then there is $m = m(g,h,s,r,t,j) \in \mathbb{N}$ such that, for all $n > m$, whenever I is an interval of length n and $A \subseteq I$ is nonempty and has the (m,r)-density property on I, then A contains a (t,d,w)-almost progression with homogeneity s such that $w/d < h(d/n)$ and $1/g(m) < d/n < 1/j$.*

Roughly speaking, if A has sufficient density regularity, then A contains an almost-progression with "small" w (small compared to the distance of the progression).

The proof of the theorem relies on the following standard lemma; see [34, Lemma 1].

Lemma 15.7. *Suppose that $E \subseteq \mathbb{R}$ has positive Lebesgue measure and $t \in \mathbb{N}$. Then there is $v > 0$ such that, for all $0 < u < v$, there is an arithmetic progression in E of length t and difference u.*

We stress that in the previous lemma, u and v are real numbers.

Proof (of Theorem 15.6). Fix g,h,s,r,j,t as in the statement of Theorem 15.6. We show that the conclusion holds for all infinite M, whence by underflow there exists $m \in \mathbb{N}$ as desired. Thus, we fix $M > \mathbb{N}$ and consider $N > M$, an interval $I \subseteq {}^*\mathbb{N}$ of length N, and a hyperfinite subset $A \subseteq I$ that has the internal (M,r)-density property on I. Without loss of generality, we may assume that $I = [1,N]$. Suppose that we can find $B,D,W \in {}^*\mathbb{N}$ and standard $c > 0$ such that $[B, B+(t-1)D+W] \subseteq [1,N]$ and, for all $i = 0,1,\ldots,t-1$, we have:

$$\delta(A,[1,N])(c - \frac{s}{2}) \leq \delta(A,[B+iD,B+iD+W]) \leq \delta(A,[1,N])(c + \frac{s}{4}). \quad (\dagger)$$

We claim that A nearly contains the internal (t,D,W)- progression $\mathrm{B}(B,t,D,W)$ with homogeneity s. Indeed, item (i) of Definition 15.3 is clear. For item (ii), observe that

$$\delta(A,[B+iD,B+iD+W]) \geq \delta(A,[1,N])(c - \frac{s}{2}) \geq \delta(A,[B+jD,B+jD+W])(\frac{c-\frac{s}{2}}{c+\frac{s}{4}})$$

and note that $\frac{c-\frac{s}{2}}{c+\frac{s}{4}} > 1 - s$. Thus, it suffices to find B,D,W,c satisfying $(\dagger)$ and for which $W/D < h(D/N)$ and $1/g(M) < D/N < 1/j$.

Let f be defined as in the statement of Lemma 15.5. Set $b := \mathrm{st}(B/N)$, $d := \mathrm{st}(D/N)$, and $w := \mathrm{st}(W/N)$. Assume that $w \neq 0$. Then we have that

$$\mathrm{st}\left(\frac{\delta(A,[B+iD,B+iD+W])}{\delta(A,[1,N])}\right)=\frac{f(b+id+w)-f(b+id)}{w}.$$

We thus want to find B,D,W and c satisfying

$$c-\frac{s}{2}<\frac{f(b+id+w)-f(b+id)}{w}<c+\frac{s}{4}. \quad (\dagger\dagger)$$

Now the middle term in $(\dagger\dagger)$ looks like a difference quotient and the idea is to show that one can bound $f'(b+id)$ for $i=0,1,\ldots,t-1$. Indeed, by Lemma 15.5, f is Lipshitz, whence it is absolutely continuous. In particular, by the Fundamental Theorem of Calculus, f is differentiable almost everywhere and $f(x)=\int_0^x f'(u)du$. Since $f(0)=0$ and $f(1)=1$, it follows that $\{x\in[0,1] \ : \ f'(x)\geq(1-\frac{s}{4})\}$ has positive measure. In particular, there is $c>1$ such that

$$E:=\{x\in[0,1] \ : \ c-\frac{s}{4}\leq f'(x)\leq c\}$$

has positive measure. By Lemma 15.7, there is $b\in E$ and $0<u<1/j$ such that $b,b+u,b+2u,\ldots,b+(t-1)u\in E$. Take $B,D\in[1,N]$ such that $b=\mathrm{st}(B/N)$ and $u=\mathrm{st}(D/N)$. Note that $g(M)$ is infinite and D/N is noninfinitesimal, so $1/g(M)<D/N<1/j$. It remains to choose W. Since f is differentiable on E, there is $w>0$ sufficiently small so that for all $i=0,1,\ldots,t-1$, we have $|f'(b+id)-\frac{f(b+id+w)-f(b+id)}{w}|<\frac{s}{4}$. For this w, $(\dagger\dagger)$ clearly holds; we now take W such that $w=\mathrm{st}(W/N)$. Since $h(D/N)$ is nonfinitesimal (as D/N is noninfinitesimal), if w is chosen sufficiently small, then $W/D<h(D/N)$.

Theorem 15.6 implies a very weak form of Szemeredi's theorem.

Corollary 15.8. *Suppose that* $\mathrm{BD}(A)>0$. *Suppose that* g,h,s,t,j *are as in the hypothesis of Theorem 15.6. Then for* n *sufficiently large, there is an interval* I *of length* n *such that* $A\cap I$ *contains a* (t,s,d)*-almost progression in* I *with* $w/d<h(d/n)$ *and* $1/g(m)<d/n<1/j$.

Proof. Let $r\in\mathbb{R}^{>1}$ be such that $\mathrm{BD}(A)>1/r$. Let $m:=m(g,h,s,r,t,j)$ as in the conclusion of Theorem 15.6. Let $n>m$ and take an interval I of length n such that $\delta(A,I)>1/r$. It remains to observe that $A\cap I$ has the (m,r)-density property on I.

15.2 Connection to the Erdős-Turán conjecture

Leth's original motivation was the following conjecture of Erdős and Turán from [17]:

Conjecture 15.9 (Erdős-Turán). Suppose that $A=(a_n)$ is a subset of $\mathbb{N}$ such that $\sum 1/a_n$ diverges. Then A contains arbitrarily long arithmetic progressions.

Leth first observed the following standard fact about the densities of sequences satisfying the hypotheses of the Erdős-Turán conjecture.

Lemma 15.10. *Suppose that* $A=(a_n)$ *is enumerated in increasing order and is such that* $\sum 1/a_n$ *diverges. Then, for arbitrarily large* n, *one has* $\delta(A,n)>1/(\log n)^2$.

Proof. We argue by contrapositive. Suppose that $\delta(A,n])\leq 1/(\log n)^2$ for all $n\geq n_0\geq 4$. We first show that this implies that $a_n\geq\frac{1}{2}n(\log n)^2$ for all $n>n_0$. Suppose otherwise and fix $n\geq n_0$. Then $|A\cap[1,\frac{1}{2}n(\log n)^2]|\geq n$. On the other hand, by our standing assumption, we have that

$$|A\cap[1,\frac{1}{2}n(\log n)^2])\leq\frac{1/2n(\log n)^2}{(\log((1/2n(\log n))^2}\leq\frac{1}{2}n,$$

yielding the desired contradiction.

Since $a_n\geq\frac{1}{2}n(\log n)^2$ eventually, we have that

$$\sum \frac{1}{a_n} \leq \sum \frac{2}{n(\log n)^2},$$

whence $\sum \frac{1}{a_n}$, converges.

The truth of the following conjecture, together with the theorem that follows it, would imply that, for sets satisfying the density condition in the previous lemma, the existence of almost arithmetic progressions implies the existence of arithmetic progressions.

Conjecture 15.11 (Leth). Fix $t \in \mathbb{N}$ and $c > 0$. Then there is $n_0 := n_0(t,c)$ such that, for all $n \geq n_0$, whenever $A \subseteq \mathbb{N}$ is such that $\delta(A,n) > 1/(c\log n)^{2\log\log n}$, then A nearly contains a (t,d,w)-progression on $[1,n]$ with $w/d < d/n$ where d is a power of 2.

We should remark that requiring that d be a power of 2 is not much of an extra requirement. Indeed, our proof of Theorem 15.6 shows that one can take d there to be a power of 2. For any t and c, we let $L(t,c)$ be the statement that the conclusion of the previous conjecture holds for the given t and c. We let $L(t)$ be the statement that $L(t,c)$ holds for all $c > 0$.

Theorem 15.12. *Suppose that $L(t)$ is true for a given $t \in \mathbb{N}$. Further suppose that $A \subseteq \mathbb{N}$ is such that there is $c > 0$ for which, for arbitrarily large n, one has $\delta(A,n) > c/(\log n)^2$. Then A contains an arithmetic progression of length t.*

Before we prove this theorem, we state the following standard combinatorial fact, whose proof we leave as an exercise to the reader (alternatively, this is proven in [34, Proposition 1]).

Proposition 15.13. *Let $m,n \in \mathbb{N}$ be such that $m < n$, let $A \subseteq \mathbb{N}$, and let I be an interval of length n. Then there is an interval $J \subseteq I$ of length m such that $\delta(A,J) > \delta(A,I)/2$.*

Proof (of Theorem 15.12). For reasons that will become apparent later in the proof, we will need to work with the set $2A$ rather than A. Note that $2A$ satisfies the hypothesis of the theorem for a different constant $c' > 0$.

By overflow , we may find $M > \mathbb{N}$ such that $\delta(^*(2A),M) > \frac{c'}{(\log M)^2}$. Take $L > \mathbb{N}$ such that $2^{2^L} \leq M < 2^{2^{L+1}}$ and set $N := 2^{2^L}$. If we apply Proposition 15.13 to any $n \leq N$ and $I = [1,N]$, we can find an interval $I_n \subseteq [1,M]$ of length n such that

$$|^*(2A) \cap I_n| > \frac{c'M}{2(\log M)^2} \geq \frac{c'M}{2(\log 2^{2^{L+1}})^2} = \frac{c'/8}{(\log N)^2}.$$

For $1 \leq k \leq L$, write $I_{2^{2^k}} = [x_k, y_k]$.

We will now construct an internal set $B \subseteq [1,N]$ such that $\delta(B,N) > \frac{1}{(c''\log N)^{2\log\log N}}$, where $c'' := \sqrt{8/c'}$. Since we are assuming that $L(t)$ holds, by transfer we will be able to find an internal (t,d,w)-progression nearly inside of B with $w/d < d/N$ and w and d both powers of 2. The construction of B will allow us to conclude that $^*(2A)$ contains a t-termed arithmetic progression of difference d, whence so does $2A$ by transfer, and thus so does A.

Set $B_0 := [1,N]$ and, for the sake of describing the following recursive construction, view B_0 as the union of two subintervals of length $N/2 = 2^{2^L-1} = 2^{2^L-2^0}$; we refer to these subintervals of B_0 as *blocks*. Now divide each block in B_0 into $2 = 2^{2^0}$ intervals of length $2^{2^L-2^0}/2^{2^0} = 2^{2^L-2^1}$ and, for each $0 \leq j < 2^{2^0}$, we place the j^{th} subblock of each block in B_0 into B_1 if and only if $x_0 + j \in {}^*2A$.

Now divide each block in B_1 into 2^{2^1} intervals of length $2^{2^L-2^1}/2^{2^1} = 2^{2^L-2^2}$ and, for each $0 \leq j < 2^{2^1}$, we place the j^{th} subblock of each block in B_1 into B_2 if and only if $x_1 + j \in {}^*2A$.

We continue recursively in this manner. Thus, having constructed the hyperfinite set B_k, which is a union of blocks of length $2^{2^L-2^k}$, we break each block of B_k into 2^{2^k} many intervals of length $2^{2^L-2^k}/2^{2^k} = 2^{2^L-2^{k+1}}$ and we place the j^{th} subblock of each block in B_k into B_{k+1} if and only if $x_k + j \in {}^*2A$.

We set $B := B_L$. Since $|B_{k+1}|/|B_k| > \frac{c'/8}{(\log N)^2}$ for each $0 \leq k < L$, it follows that

$$|B| > \frac{(c'/8)^L N}{(\log N)^{2L}} = \frac{N}{(c''\log N)^{2\log\log N}}.$$

By applying transfer to $L(t)$, we have that B nearly contains an internal (t,d,w)-progression $\mathrm{B}(b,t,d,w)$ contained in $[1,N]$ such that $w/d < d/N$ and d is a power of 2. Take k such that $2^{2^L-2^{k+1}} \leq d < 2^{2^L-2^k}$. Note that this implies that $2^{2^L-2^{k+1}} \mid d$. Also, we have

$$w < (d/N)\cdot d < (2^{-2^k})2^{2^L-2^k} = 2^{2^L-2^{k+1}}.$$

We now note that $\mathrm{B}(b,t,d,w)$ must be contained in a single block C of B_k. Indeed, since $d \mid 2^{2^L-2^k}$ and $w \mid 2^{2^L-2^{k+1}}$, we have $d+w < (\frac{1}{2}+\frac{1}{2^{2^k}})(2^{2^L-2^k})$, whence the fact that $[b,b+w]$ and $[b+d,b+d+w]$ both intersect B_k would imply that $[x_{k-1},y_{k-1}]$ contains consecutive elements of ${}^*2A$, which is clearly a contradiction.

Now write $d = m\cdot 2^{2^L-2^{k+1}}$. Take $0 \leq j < 2^{2^k}$ so that $[b,b+w]$ intersects B_{k+1} in the j^{th} subblock of C so $x_k + j \in {}^*2A$. Since $[b+d,b+d+w]\cap B_{k+1} \neq \emptyset$, we have that at least one of $x_k+j+(m-1)$, x_k+j+m, or $x_k+j+(m+1)$ belong to ${}^*(2A)$. However, since x_k+j and m are both even, it follows that we must have $x_k+j+m \in {}^*(2A)$. Continuing in this matter, we see that $x_k+j+im \in {}^*2A$ for all $i=0,1,\ldots,t-1$. It follows by transfer that $2A$ contains a t-term arithmetic progression, whence so does A.

Putting everything together, we have:

Corollary 15.14. *The Erdős-Turán conjecture follows from Leth's Conjecture.*

Leth used Theorem 15.6 to prove the following theorem, which is similar in spirit to Conjecture 15.6, except that it allows sparser sequences but in turn obtains almost progressions with weaker smallness properties relating d and w.

Theorem 15.15. *Suppose that $s>0$ and $t \in \mathbb{N}^{>2}$ are gven. Further suppose that h is as in Theorem 15.6. Let $A \subseteq \mathbb{N}$ be such that, for all $\varepsilon > 0$, we have $\delta(A,n) > 1/n^{\varepsilon}$ for sufficiently large n. Then for sufficiently large n, A nearly contains an (t,d,w)-progression on $[1,n]$ of homogeneity s with $w/d < h(\log d/\log n)$, where d is a power of 2.*

Proof. Suppose that the conclusion is false. Then there is N such that ${}^*A$ does not nearly contain any internal (t,d,w)-progression on $[1,N]$ of homogeneity s with $w/d < h(\log d/\log N)$. It suffices to show that there is $\varepsilon > 0$ such that $\delta({}^*A,N) < 1/N^{\varepsilon}$. Let m be as in the conclusion of Theorem 15.6 with $r=2$ and $g(x)=x$ (and h as given in the assumptions of the current theorem).

Claim: If $I \subseteq [1,N]$ is a hyperfinite interval with $|I| > \sqrt{N}$, then ${}^*A$ does *not* have the $(m,2)$-density property on I.

We will return to the proof of the claim in a moment. We first see how the claim allows us to complete the proof of the theorem. Let $K > \mathbb{N}$ be the maximal $k \in {}^*\mathbb{N}$ such that $m^{2k} \leq N$, so $m^{2K} \leq N < m^{2K+2}$. We construct, by internal induction, for $i=0,1,\ldots,K$, a descending chain of hyperfinite subintervals (I_i) of I of length m^{2K-i} as follows. By Proposition 15.13, we may take I_0 to be any hyperfinite subinterval of I of length m^{2K} such that $\delta({}^*A,I_0) \geq \delta({}^*A,N)/2$. Suppose that $i<K$ and I_i has been constructed such that $|I_i| = m^{2k-i}$. Since ${}^*A$ does not have the $(m,2)$ density property on I_i, there is a subinterval I_{i+1} of length $|I_i|/m^{2k-i-1}$ with $\delta({}^*A,I_{i+1}) \geq 2\delta({}^*A,I_i)$. Notice now that I_K is a hyperfinite interval of length $m^K \leq \sqrt{N} < m^{K+1}$ and $\delta({}^*A,I_K) \geq 2^K\delta({}^*A,I_0)$. It follows that

$$\delta({}^*A,N) \leq 2\delta(A,I_0) \leq 2^{-(K-1)}\delta(A,I_K) \leq 2^{-(K-1)}.$$

It follows that

$$|A\cap[1,N]| \leq 2^{-(K-1)}N \leq 2^{-(K-1)}m^{2K+2} = m^{2K+2-(K-1)\frac{\log 2}{\log m}} = (m^{2K})^{1-z}.$$

if we set $z := \frac{(K-1)\log 2}{2K\log m} - \frac{1}{K}$. If we set $\varepsilon := \mathrm{st}(z/2) = \frac{\log 2}{4\log m}$, then it follows that $|A\cap[1,N]| \leq N^{1-\varepsilon}$, whence this ε is as desired.

We now prove the claim. Suppose, towards a contradiction, that $I \subseteq [1,N]$ is a hyperfinite interval with $|I| > \sqrt{N}$ and is such that ${}^*A$ does have the $(m,2)$-density property on I. By the choice of m, ${}^*A$ nearly contains an internal (t,d,w)-almost progression of homogeneity s with $w/d < h(d/|I|)$ and $d > |I|/m > \sqrt{N}/m$. Notice now that $\mathrm{st}\left(\frac{\log d}{\log N}\right) \geq \mathrm{st}\left(\frac{1/2\log N - \log m}{\log N}\right) = \frac{1}{2}$. Note that we trivially have that $d/|I| < 1/t$, whence $d/|I| < \log d/\log N$; since h is increasing, we have that $w/d < h(\log d/\log N)$, contradicting the choice of N. This proves the claim and the theorem.

In [35, Theorem 3], Leth shows that one cannot replace $(\log d)/(\log n)$ with d/n in the previous theorem.

Chapter 16
The interval measure property

The material in this chapter comes from the paper [36] although many of the proofs appearing below, communicated to us by Leth, are simpler than those appearing in the aforementioned article.

16.1 IM sets

Let $I := [y,z]$ be an infinite, hyperfinite interval. Set $\mathrm{st}_I := \mathrm{st}_{[y,z]} : I \to [0,1]$ to be the map $\mathrm{st}_I(a) := \mathrm{st}(\frac{a-y}{z-y})$. For $A \subseteq {}^*\mathbb{N}$ internal, we set $\mathrm{st}_I(A) := \mathrm{st}_I(A \cap I)$. We recall that $\mathrm{st}_I(A)$ is a closed subset of $[0,1]$ and we may thus consider $\lambda_I(A) := \lambda(\mathrm{st}_I(A))$, where λ is Lebesgue measure on $[0,1]$.

We also consider the quantity $g_A(I) := \frac{d-c}{|I|}$, where $[c,d] \subseteq I$ is maximal so that $[c,d] \cap A = \emptyset$.

The main concern of this subsection is to compare the notions of making $g_A(I)$ small (an internal notion) and making $\lambda_I(A)$ large (an external notion). There is always a connection in one direction:

Lemma 16.1. *If $\lambda_I(A) > 1-\varepsilon$, then $g_A(I) < \varepsilon$.*

Proof. Suppose that $g_A(I) \geq \varepsilon$, whence there is $[c,d] \subseteq I$ such that $[c,d] \cap A = \emptyset$ and $\frac{d-c}{|I|} \geq \varepsilon$. It follows that, for any $\delta > 0$, we have $(\mathrm{st}_I(c) + \delta, \mathrm{st}_I(d) - \delta) \cap \mathrm{st}_I(A) = \emptyset$, whence

$$\lambda_I(A) \leq 1 - \left(\mathrm{st}\left(\frac{d-c}{|I|}\right) - 2\delta\right) \leq 1 - \varepsilon + 2\delta.$$

Letting $\delta \to 0$ yields the desired result.

We now consider sets where there is also a relationship in the other direction.

Definition 16.2. We say that A has the *interval-measure property* (or *IM property*) on I if for every $\varepsilon > 0$, there is $\delta > 0$ such that, for all infinite $J \subseteq I$ with $g_A(J) \leq \delta$, we have $\lambda_J(A) \geq 1-\varepsilon$.

If A has the IM property on I, we let $\delta(A,I,\varepsilon)$ denote the supremum of the δ's that witness the conclusion of the definition for the given ε.

We now seek to establish nice properties of sets with the IM property. We first establish a kind of partition regularity theorem. For convenience, let us say that A has the *enhanced IM property on I* if it has the IM property on I and $\lambda_I(A) > 0$.[1]

Theorem 16.3. *Suppose that A has the enhanced IM property on I. Further suppose that $A \cap I = B_1 \cup \cdots B_n$ with each B_i internal. Then there is i and infinite $J \subseteq I$ such that B_i has the enhanced IM property on J.*

[1] This terminology does not appear in the original article of Leth.

Proof. We prove the theorem by induction on n. The result is clear for $n = 1$. Now suppose that the result is true for $n-1$ and suppose $A \cap I = B_1 \cup \cdots B_n$ with each B_i internal. If there is an i and infinite $J \subseteq I$ such that $B_i \cap J = \emptyset$ and $\lambda_J(A) > 0$, then we are done by induction. We may thus assume that whenever $\lambda_J(A) > 0$, then each $B_i \cap J \neq \emptyset$. We claim that this implies that each of the B_i have the IM property on I. Since there must be an i such that $\lambda_I(B_i) > 0$, for such an i it follows that B_i has the enhanced IM property on I.

Fix i and set $B := B_i$. Suppose that $J \subseteq I$ is infinite, $\varepsilon > 0$, and $g_B(J) \leq \delta(A, I, \varepsilon)$; we show that $\lambda_J(B) \geq 1-\varepsilon$. Since $g_A(J) \leq g_B(J) \leq \delta(A, I, \varepsilon)$, we have that $\lambda_J(A) \geq 1-\varepsilon$. Suppose that $[r,s] \subseteq [0,1] \setminus \mathrm{st}_J(B)$. Then $r = \mathrm{st}_J(x)$ and $s = \mathrm{st}_J(y)$ with $\frac{y-x}{|J|} \approx s-r$ and $B \cap [x,y] = \emptyset$. By our standing assumption, this implies that $\lambda_{[x,y]}(A) = 0$, whence it follows that $\lambda_J(A \cap [x,y]) = 0$. It follows that $\lambda_J(B) = \lambda_J(A) \geq 1-\varepsilon$, as desired.

An important standard tool in the study of sets with the IM property is the *Lebesgue density theorem*. Recall that for a measurable set $E \subseteq [0,1]$, a point $r \in E$ is a *(one-sided) point of density of* E if

$$\lim_{s \to r^+} \frac{\mu(E \cap [r,s])}{s-r} = 1.$$

The Lebesgue density theorem asserts that almost every point of E is a density point of E.

If A has the IM property on an interval I and we have a subinterval of I on which A has small gap ratio, then by applying the IM property, the Lebesgue density theorem, and Lemma 16.1, we can find a smaller, but appreciably sized, subinterval on which A once again has small gap ratio. Roughly speaking, one can iterate this procedure until one finds a *finite* subinterval of I on which A has small gap ratio; the finiteness of the subinterval will be crucial for applications. Here is a precise formulation:

Theorem 16.4. *Suppose that $A_1, \ldots, A_n$ are internal sets that satisfy the IM property on $I_1, \ldots, I_n$ respectively. Fix $\varepsilon > 0$ such that $\varepsilon < \frac{1}{n}$. Take $\delta > 0$ with $\delta < \min_{i=1,\ldots,n} \delta(A_i, I_i, \varepsilon)$. Then there is $w \in \mathbb{N}$ such that whenever*

$$[a_i, a_i + b] \subseteq I_i \text{ and } g_{A_i}([a_i, a_i+b]) \leq \delta \text{ for all } i = 1, \ldots, n \quad (\dagger)$$

then there is $c \in {}^*\mathbb{N}$ and $b' \leq w$ with $c + b' \leq b$ such that

$$g_{A_i}([a_i + c, a_i + c + b']) \leq \delta \text{ for all } i = 1, \ldots, n. \quad (\dagger\dagger)$$

Proof. Let $A_1, \ldots, A_n$, $I_1, \ldots, I_n$, ε and δ be as in the statement of the theorem. The entire proof rests on the following:

Claim: Whenever $b > \mathbb{N}$ is such that ($\dagger$) is true for some $a_1, \ldots, a_n$, then there is $b' < b$ with $\frac{b'}{b} \not\approx 0$ and $c \in {}^*\mathbb{N}$ with $c + b' \leq b$ such that ($\dagger\dagger$) holds.

Let us assume that the claim is true and finish the proof of the theorem. Let $N := \min_{i=1,\ldots,n} |I_i|$. Define an internal function $f : [0,N] \to [0,N]$ by $f(b) =$ the maximal $b' < b$ witnessing the truth of ($\dagger\dagger$) if $b > \mathbb{N}$ and there are $a_1, \ldots, a_n$ witnessing the truth of ($\dagger$) for b; otherwise, let $f(b) = b$. By the claim, if $b > \mathbb{N}$, then $f(b) < b$ and $\frac{f(b)}{b} \not\approx 0$. By saturation, there is $\varepsilon > 0$ such that if $b > \mathbb{N}$, then $\frac{f(b)}{b} \geq \varepsilon$. Let $g : {}^*\mathbb{N} \times [0,N] \to [0,N]$ be the unique internal function satisfying $g(0,b) = b$ and $g(n+1,b) = f(g(n,b))$ for all $n \in {}^*\mathbb{N}$ and all $b \in [0,N]$. It must be the case that, for all $b \in [0,N]$, the function $n \mapsto g(n,b)$ is eventually constant. Let $h : [0,N] \to [0,N]$ be the internal function defined by $h(b) =$ the eventual value of $g(n,b)$. It must then be the case that $h(b) \in \mathbb{N}$ for all $b \in [0,N]$, whence there is $w \in \mathbb{N}$ such that $h(b) \leq w$ for all $b \in [0,N]$; this w is as desired.

Thus, to finish the proof of the theorem, it suffices to prove the claim.

Proof of Claim: Suppose that $a_1, \ldots, a_n$ and $b > \mathbb{N}$ are as in the claim. Let $J_i := [a_i, a_i + b]$ for $i = 1, \ldots, n$. By assumption, $\lambda_{J_i}(A_i) \geq 1 - \frac{1}{n}$, whence $\lambda(\bigcap_{i=1}^n \mathrm{st}_{J_i}(A_i)) > 0$. Let r be a point of density for $\bigcap_{i=1}^n \mathrm{st}_{J_i}(A_i)$. Thus, there is $s < 1-r$ such that

$$\lambda\left(\left(\bigcap_{i=1}^n \mathrm{st}_{J_i}(A_i)\right) \cap [r, r+s]\right) \geq (1-\delta)s.$$

Set $c := \lfloor r \cdot b \rfloor$ and $b' := \lfloor s \cdot b \rfloor$. Then $c + b' \leq b$ and, by Lemma 16.1, we have

$$g_{A_i}([a_i+c,a_i+c+b']) \leq \delta \text{ for all } i=1,\ldots,n.$$

Let us record a corollary of the *proof* of the previous theorem.

Corollary 16.5. *If A has the IM property on I, then there is $w \in \mathbb{N}$ and a descending hyperfinite sequence $I = I_0, I_1, \ldots, I_K$ of hyperfinite subintervals of I such that:*

- $|I_K| \leq w$;
- $\frac{|I_{k+1}|}{|I_k|} \geq \frac{1}{w}$;
- *whenever I_k is infinite, we have $\lambda_{I_k}(A) > 0$.*

A special case of Theorem 16.4 is worth singling out:

Corollary 16.6. *Let $A_1,\ldots,A_n$, $I_1,\ldots,I_n$, ε, and δ be as in Theorem 16.4. Then there is $c \in \mathbb{N}$ such that, whenever $[a_i,a_i+b]$ satisfies (†), then there is $c \in {}^*\mathbb{N}$ such that*

$$A_i \cap [a_i+c,a_i+c+w] \neq \emptyset \text{ for all } i=1,\ldots,n.$$

Definition 16.7. For any (not necessarily internal) $A \subseteq {}^*\mathbb{N}$, we set

$$D(A) := \{n \in \mathbb{N} \ : \ n = a-a' \text{ for infinitely many pairs } a,a' \in A\}.$$

The following corollary will be important for our standard application in the next section.

Corollary 16.8. *Suppose that A has the enchanced IM property on I. Then $D(A)$ is syndetic.*

Proof. Let $w \in \mathbb{N}$ be as in Corollary 16.6 for $A_1 = A_2 = A$ and $I_1 = I_2 = I$. It suffices to show that: for all $m \in \mathbb{N}$, we have $(A-A)\cap[m-w,m+w]$ is infinite.

Let (r_n) be a collection of distinct points of density of $\mathrm{st}_I(A)$. Then by Lemma 16.1 and overflow, there are pairwise disjoint infinite subintervals $J_n := [a_n,b_n] \subseteq I$ such that $g_A(J_n) \approx 0$. Note also that $g_A(J_n+m) \approx 0$. Thus, by the choice of w, for each n, there is $c_n \in {}^*\mathbb{N}$ such that

$$A\cap[a_n+c_n,a_n+c_n+w], A\cap[a_n+m+c_n,a_n+m+c_n+w] \neq \emptyset.$$

If $x_n \in A\cap[a_n+c_n,a_n+c_n+w]$ and $y_n \in A\cap[a_n+m+c_n,a_n+m+c_n+w]$, then $y_n - x_n \in (A-A)\cap[m-w,m+w]$. By construction, the pairs (x_n,y_n) are all distinct.

16.2 SIM sets

We now seek to extract the standard content of the previous section.

Definition 16.9. $A \subseteq \mathbb{N}$ has the *standard interval-measure property* (or *SIM property*) if:

- ${}^*A$ has the IM property on every infinite hyperfinite interval;
- ${}^*A$ has the enchanced IM property on some infinite hyperfinite interval.

Example 16.10. Let $A = \bigcup_n I_n$, where each I_n is an interval, $|I_n| \to \infty$ as $n \to \infty$, and there is $k \in \mathbb{N}$ such that the distance between consecutive I_n's is at most k. Then A has the SIM property.

We now reformulate the definition of SIM set using only standard notions. (Although recasting the SIM property in completely standard terms is not terribly illuminating, it is the polite thing to do.) First, note that one can define $g_A(I)$ for standard $A \subseteq \mathbb{N}$ and standard finite intervals $I \subseteq \mathbb{N}$ in the exact same manner. Now, for $A \subseteq \mathbb{N}$ and $0 < \delta < \varepsilon < 1$, define the function $F_{\delta,\varepsilon,A} : \mathbb{N} \to \mathbb{N}$ as follows. First, if $g_A(I) > \delta$ for every $I \subseteq \mathbb{N}$ of length $\geq n$, set $F_{\delta,\varepsilon,A}(n) = 0$. Otherwise, set $F_{\delta,\varepsilon,A}(n) =$ the minimum k such that there is an interval $I \subseteq \mathbb{N}$ of length $\geq n$ such that $g_A(I) \leq \delta$ and there are subintervals $I_1,\ldots,I_k \subseteq I$ with $I_i \cap A = \emptyset$ for all $i=1,\ldots,k$ and $\sum_{i=1}^k |I_i| \geq \varepsilon|I|$.

Theorem 16.11. *A has the SIM property if and only if: for all $\varepsilon > 0$, there is $\delta > 0$ such that* $\lim_{n\to\infty} F_{\delta,\varepsilon,A}(n) = \infty$.

Proof. First suppose that there is $\varepsilon > 0$ such that $\liminf_{n\to\infty} F_{\delta,\varepsilon,A}(n) < \infty$ for all $\delta > 0$; we show that A does not have the SIM property. Towards this end, we may suppose that $\lambda_I(^*A) > 0$ for some infinite hyperfinite interval I and show that $^*A$ does not have the IM property on some infinite interval. Fix $0 < \delta < \varepsilon$. By the Lebesgue density theorem and Lemma 16.1, we have that $g_{^*A}(J) \leq \delta$ for some infinite subinterval $J \subseteq I$. By transfer, there are intervals $J_n \subseteq \mathbb{N}$ of length $\geq n$ such that $g_A(J_n) \leq \delta$, whence $0 < \liminf_{n\to\infty} F_{\delta,\varepsilon,A}(n)$ for all $0 < \delta < \varepsilon$. For every $k \geq 1$, set $m_k := 1 + \liminf_{n\to\infty} F_{\frac{1}{k},\varepsilon,A}(n)$. Consequently, for every $n \in \mathbb{N}$, there are intervals $I_{1,n}, \ldots, I_{n,n}$ of length $\geq n$ such that, for each $k = 1, \ldots, n$, $g_A(I_{k,n}) \leq \frac{1}{k}$ and the sum of the lengths of m_k many gaps of A in $I_{k,n}$ is at least $\varepsilon \cdot |I_{k,n}|$. Set $I_n := I_{1,n} \cup \cdots \cup I_{n,n}$. By overflow, there is an infinite, hyperfinite interval I that contains infinite subintervals I_k such that $g_{^*A}(I_k) \leq \frac{1}{k}$ and yet the sum the lengths of of m_k many gaps of A on I_k have size at least $\varepsilon |I_k|$. It follows that $^*A$ does not have the IM property on I.

Now suppose that for all $\varepsilon > 0$, there is $\delta > 0$ such that $\lim_{n\to\infty} F_{\delta,\varepsilon,A}(n) = \infty$ and that I is an infinite, hyperfinite interval such that $g_{^*A}(I) \leq \delta$. By transfer, it follows that no finite number of gaps of $^*A$ on I have size at least $\varepsilon \cdot |I|$. Since $\operatorname{st}_I(^*A)$ is closed, we have that $\lambda_I(^*A) \geq 1 - \varepsilon$. Consequently, A has the IM property on any infinite, hyperfinite interval. Since, by transfer, there is an infinite, hyperfinite interval I with $g_{^*A}(I) \leq \delta$, this also shows that $^*A$ has the enhanced IM property on this I. Consequently, A has the SIM property.

The next lemma shows that the SIM property is not simply a measure of "largeness" as this property is not preserved by taking supersets.

Lemma 16.12. *Suppose that $A \subseteq \mathbb{N}$ is not syndetic. Then there is $B \supseteq A$ such that B does not have the SIM property.*

Proof. For each n, let $x_n \in \mathbb{N}$ be such that $[x_n, x_n + n^2] \cap A = \emptyset$. Let

$$B := A \cup \bigcup_n \{x_n + kn \ : \ k = 0, 1, \ldots n\}.$$

Fix $\varepsilon > 0$. Take $m \in \mathbb{N}$ such that $m > \frac{1}{\varepsilon}$ and take $N > \mathbb{N}$. Set $I := [x_N, x_N + mN]$. Indeed, $g_{^*B}(I) = \frac{N}{mN} < \frac{1}{\varepsilon}$ while

$$\operatorname{st}_I(^*B) = \left\{ \operatorname{st}\left(\frac{kN}{mN}\right) \ : \ k = 0, \ldots, m \right\} = \left\{ \frac{k}{m} \ : \ k = 0, \ldots, m \right\}$$

is finite and thus has measure 0. It follows that $^*B$ does not have the IM property on I, whence B does not have the SIM property.

The previous lemma also demonstrates that one should seek structural properties of a set which ensure that it contains a set with the SIM property. Here is an example:

Lemma 16.13. *If B is piecewise syndetic, then there is $A \subseteq B$ with the SIM property.*

Proof. For simplicity, assume that B is thick; the argument in general is similar, just notationally more messy. Let $A := \bigcup_n I_n$, with I_n intervals contained in B, $|I_n| \to \infty$ as $n \to \infty$, and such that, setting g_n to be the length in between I_n and I_{n+1}, we have $g_{n+1} \geq n g_n$ for all n. We claim that A has the SIM property. It is clear that $\lambda_I(^*A) > 0$ for some infinite hyperfinite interval I; indeed, $\lambda_{I_N}(^*A) = 1$ for $N > \mathbb{N}$. Now suppose that I is an infinite hyperfinite interval; we claim that $^*A$ has the IM property on I as witnessed by $\delta = \varepsilon$. Suppose that J is an infinite subinterval of I such that $g_{^*A}(J) \leq \varepsilon$. Suppose that $I_n, \ldots, I_{M+1}$ is a maximal collection of intervals from $^*A$ intersecting J. Since $\frac{g_M}{|J|} \leq \varepsilon$, for $k = N, \ldots, M-1$, we have $\frac{g_k}{|J|} = \frac{g_k}{g_M} \cdot \frac{g_M}{|J|} \approx 0$, whence the intervals $I_n, \ldots, I_M$ merge when one applies st_J. It follows that $\lambda_J(^*A) \geq 1 - \varepsilon$.

In connection with the previous result, the following question seems to be the most lingering open question about sets that contain subsets with the SIM property:

Question 16.14. Does every set of positive Banach density contain a subset with the SIM property?

The next result shows that many sets do *not* have the SIM property.

Proposition 16.15. *Suppose that $A = (a_n)$ is a subset of $\mathbb{N}$ written in increasing order. Suppose that $\lim_{n\to\infty}(a_{n+1} - a_n) = \infty$. Then A does not have the SIM property.*

Proof. Suppose that A has the SIM property. Take I such that $\lambda_I({}^*A) > 0$. Then by the proof of Corollary 16.8, we can find $x, y \in {}^*A \setminus A$ such that $x < y$ and $y - x \le 2w$. Then, by transfer, there are arbitrarily large $m, n \in \mathbb{N}$ with $0 < m - n \le 2w$. It follows that $\lim_{n\to\infty}(a_{n+1} - a_n) \neq \infty$.

We end this section with a result concerning a structural property of sets with the SIM property. A direct consequence of Corollary 16.8 is the following:

Corollary 16.16. *If A has the SIM property, then $D(A)$ is syndetic.*

Leth's original main motivation for studying the IM property was a generalization of the previous corollary. Stewart and Tijdeman [41] proved that, given $A_1, \ldots, A_n \subseteq \mathbb{N}$ with $\mathrm{BD}(A_i) > 0$ for all $i = 1, \ldots, n$, one has $D(A_1) \cap \cdots \cap D(A_n)$ is syndetic. Leth proved the corresponding statement for sets with the SIM property:

Theorem 16.17. *If $A_1, \ldots, A_n \subseteq \mathbb{N}$ all have the SIM property, then $D(A_1) \cap \cdots \cap D(A_n)$ is syndetic.*

Proof. We break the proof up into pieces.

Claim 1: There are infinite hyperfinite intervals $I_1, \ldots, I_n$, all of which have the same length, such that $\lambda_{I_i}({}^*A_i) = 1$ for all $i = 1, \ldots, n$.
Proof of Claim 1: By the definition of the SIM property and Corollary 16.5, we may find infinite, hyperfinite intervals $J_1, \ldots, J_n$ whose length ratios are all finite and for which $\mathrm{st}_{J_i}({}^*A_i) > 0$ for $i = 1, \ldots, n$. By taking points of density in each of these intervals, for any $\varepsilon > 0$, we may find *equally sized* subintervals J_i' of J_i such that $\lambda_{J_i'}({}^*A_i) \ge 1 - \varepsilon$, whence $g_{{}^*A_i}(J_i') \le \varepsilon$. Since this latter condition is internal, by saturation, we may find equally sized subintervals I_i of J_i such that each $g_{{}^*A_i}(I_i) \approx 0$, whence, by the fact that ${}^*A_i$ has the IM property on J_i, we have $\lambda_{I_i}({}^*A_i) = 1$.

We now apply Corollary 16.6 to $A_1, \ldots, A_n$, $I_1, \ldots, I_n$ and $\varepsilon := \frac{1}{n+1}$. Let $w \in \mathbb{N}$ be as in the conclusion of that corollary. Write $I_i := [x_i, y_i]$ and for $i = 1, \ldots, n$, set $d_i := x_i - x_1$. We then set

$$B := \{a \in {}^*A_1 \cap I_1 \ : \ {}^*A_i \cap [a + d_i - w, a + d_i + 2w] \neq \emptyset \text{ for all } i = 1, \ldots, n\}.$$

Claim 2: Suppose that $J \subseteq I_1$ is infinite and r is a point of density of

$$\bigcap_{i=1}^{n} \mathrm{st}_{J+d_i}({}^*A_i).$$

Then $r \in \mathrm{st}_J(B)$.
Proof of Claim 2: By a (hopefully) by now familiar Lebesgue density and overflow argument, there is an infinite hyperfinite interval $[u, v] \subseteq J$ such that $\mathrm{st}_J(u) = \mathrm{st}_J(v) = r$ and

$$g_{{}^*A_i}([u + d_i, v + d_i]) \approx 0 \text{ for all } i = 1, \ldots, n.$$

This allows us to find $c \in {}^*\mathbb{N}$ such that $u + d_i + c + w \le v_i$ and ${}^*A_i \cap [u + d_i + c, u + d_i + c + w] \neq \emptyset$ for $i = 1, \ldots, n$. Take $a \in {}^*A_1 \cap [u + c, u + c + w]$, say $a = u + c + j$ for $j \in [0, w]$. It follows that

$${}^*A_i \cap [a + d_i - j, a + d_i + j + w] \neq \emptyset \text{ for all } i = 1, \ldots, n$$

whence $a \in B$. Since $u \le u + c \le a \le u + c + w \le v$, we have that $\mathrm{st}_J(a) = \mathrm{st}_J(v) = r$, whence $r \in \mathrm{st}_J(B)$, as desired.

Claim 3: B has the enhanced IM property on I_1.

Proof of Claim 3: Taking $J = I_1$ in Claim 2 shows that $\lambda_{I_1}(B) = 1$. We now show that B has the IM property on I_1. Fix $\varepsilon > 0$. Let $\delta = \min_{i=1,\ldots,n} \delta({}^*A_i, I_i, \frac{\varepsilon}{n})$. Suppose $J \subseteq I_1$ is such that $g_B(J) \leq \delta$. Then $g_{{}^*A_i}(J + d_i) \leq \delta$, whence

$$\lambda\left(\bigcap_{i=1}^{n} \mathrm{st}_{J+d_i}({}^*A_i)\right) \geq 1 - \varepsilon.$$

By Claim 2, we have $\lambda_J(B) \geq 1 - \varepsilon$, as desired.

For $-w \leq k_1, \ldots, k_n \leq 2w$, set

$$B_{(k_1,\ldots,k_n)} := \{b \in B \,:\, b + d_i + k_i \in {}^*A_i \text{ for all } i = 1, \ldots, n\}.$$

By the definition of B, we have that B is the union of these sets. Since B has the enhanced IM property on I_1, by Theorem 16.3, there is such a tuple $(k_1, \ldots, k_n)$ and an infinite $J \subseteq I_1$ such that $B' := B_{(k_1,\ldots,k_n)}$ has the enhanced IM property on J. By Corollary 16.8, $D(B')$ is syndetic. Since $B' - B' \subseteq \bigcap_{i=1}^{n}({}^*A_i - {}^*A_i)$, by transfer we have that $D(A_1) \cap \cdots \cap D(A_n)$ is syndetic.

Part IV
Other topics

Chapter 17
Triangle removal and Szemeredi regularity

17.1 Triangle removal lemma

The material in this section was not proven first by nonstandard methods. However, the nonstandard perspective makes the proofs quite elegant. We closely follow [43].

Suppose that $G = (V,E)$ is a finite graph. We define the *edge density* of G to be the quantity

$$e(G) := \frac{|E|}{|V \times V|}$$

and the *triangle density* of G to be the quantity

$$t(G) := \frac{|\{(x,y,z) \in V \times V \times V \ : \ (x,y),(y,z),(x,z) \in E\}|}{|V \times V \times V|}.$$

Theorem 17.1 (Triangle removal lemma). *For every $\varepsilon > 0$, there is a $\delta > 0$ such that, whenever $G = (V,E)$ is a finite graph with $t(G) \le \delta$, then there is a subgraph $G' = (V,E')$ of G that is triangle-free (so $t(G') = 0$) and such that $e(G \setminus G') \le \varepsilon$.*

In short, the triangle removal lemma says that if the triangle density of a graph is small, then one can remove a few number of edges to get one that is actually triangle-free. We first show how the Triangle Removal Lemma can be used to prove Roth's theorem, which was a precursor to Szemeredi's theorem.

Theorem 17.2 (Roth's theorem). *For all $\varepsilon > 0$, there is $n_0 \in \mathbb{N}$ such that, for all $n \ge n_0$ and all $A \subseteq [1,n]$, if $\delta(A,n) \ge \varepsilon$, then A contains a 3-term arithmetic progression.*

Proof. Fix n and form a tripartite graph $G = G(A,n)$ with vertex set $V = V_1 \cup V_2 \cup V_3$, where each V_i is a disjoint copy of $[1,3n]$. If $(v,w) \in (V_1 \times V_2) \cup (V_2 \times V_3)$, we declare $(v,w) \in E \Leftrightarrow w - v \in A$. If $(v,w) \in V_1 \times V_3$, then we declare $(v,w) \in E \Leftrightarrow (w - v) \in 2A$. Note then that if (v_1,v_2,v_3) is a triangle in G, then setting $a := v_2 - v_1$, $b := v_3 - v_2$, and $c := \frac{1}{2}(v_3 - v_1)$, we have that $a,b,c \in A$ and $a - c = c - b$. If this latter quantity is nonzero, then $\{a,b,c\}$ forms a 3-term arithmetic progression in A.

Motivated by the discussion in the previous paragraph, let us call a triangle $\{v_1,v_2,v_3\}$ in G *trivial* if $v_2 - v_1 = v_3 - v_2 = \frac{1}{2}(v_3 - v_1)$. Thus, we aim to show that, for n sufficiently large, if $\delta(A,n) \ge 1 - \varepsilon$, then $G(A,n)$ has a nontrivial triangle. If $a \in A$ and $k \in [1,n]$, then $(k, k+a, k+2a)$ is a trivial triangle in G. Since trivial triangles clearly do not share any edges, one would have to remove at least $3 \cdot |A| \cdot n \ge 3\varepsilon n^2$ many edges of G in order to obtain a triangle-free subgraph of G. Thus, if $\delta > 0$ corresponds to 3ε in the triangle removal lemma, then we can conclude that $t(G) \ge \delta$, that is, there are at least $27\delta n^3$ many triangles in G. Since the number of trivial triangles is at most $|A| \cdot (3n) \le 3n^2$, we see that G must have a nontrivial triangle fi n is sufficiently large.

We now turn to the proof of the triangle removal lemma. The basic idea is that if the triangle removal lemma were false, then by a now familiar compactness/overflow argument, we will get a contradiction to some nonstandard triangle removal lemma. Here is the precise version of such a lemma:

Theorem 17.3 (Nonstandard triangle removal lemma). *Suppose that V is a nonempty hyperfinite set and $E_{12}, E_{23}, E_{13} \in \mathscr{L}_{V\times V}$ are such that*

$$\int_{V\times V\times V} 1_{E_{12}}(u,v)1_{E_{23}}(v,w)1_{E_{13}}(u,v)d\mu(u,v,w) = 0. \quad (\dagger)$$

Then for every $\varepsilon > 0$ and $(i,j) \in \{(1,2),(2,3),(1,3)\}$, there are hyperinite $F_{ij} \subseteq V \times V$ such that $\mu_{V\times V}(E_{ij} \setminus F_{ij}) < \varepsilon$ and

$$1_{F_{12}}(u,v)1_{F_{23}}(v,w)1_{F_{13}}(u,v) = 0 \text{ for all } (u,v,w) \in V\times V\times V. \quad (\dagger\dagger)$$

Proposition 17.4. *The nonstandard triangle removal lemma implies the triangle removal lemma.*

Proof. Suppose that the triangle removal lemma is false. Then there is $\varepsilon > 0$ such that, for all $n \in \mathbb{N}$, there is a finite graph $G_n = (V_n, E_n)$ for which $t(G_n) \le \frac{1}{n}$ and yet there does not exist a triangle-free subgraph $G' = (V_n, E_n')$ with $|E_n \setminus E_n'| \le \varepsilon |V_n|^2$. Note that it follows that $|V_n| \to \infty$ as $n \to \infty$. By , there is an infinite hyperfinite graph $G = (V,E)$ such that $t(G) \approx 0$, whence $(\dagger)$ holds, and yet there does not exist a triangle-free hyperfinite subgraph $G' = (V,E')$ with $|E \setminus E'| \le \varepsilon |V|^2$. We claim that this latter statement yields a counterexample to the nonstandard triangle removal lemma. Indeed, if the nonstandard triangle removal held, then there would be hyperfinite $F_{ij} \subseteq V \times V$ such that $\mu_{V\times V}(E \setminus F_{ij}) < \frac{\varepsilon}{6}$ and for which $(\dagger\dagger)$ held. If one then sets $E' := E \cap \bigcap_{ij}(F_{ij} \cap F_{ij}^{-1})$, then $G' = (V,E')$ is a hyperfinite subgraph of G that is triangle-free and $\mu(E \setminus E') < \varepsilon$, yielding the desired contradiction.[1]

It might look like the nonstandard triangle removal lemma is stated in a level of generality that is more than what is needed for we have $E_{12} = E_{23} = E_{13} = E$. However, in the course of proving the lemma, we will come to appreciate this added level of generality of the statement.

Lemma 17.5. *Suppose that $f \in L^2(\mathscr{L}_{V\times V})$ is orthogonal to $L^2(\mathscr{L}_V \otimes \mathscr{L}_V)$. Then for any $g, h \in L^2(\mathscr{L}_{V\times V})$, we have*

$$\int_{V\times V\times V} f(x,y)g(y,z)h(x,z)d\mu_{V\times V\times V}(x,y,z) = 0.$$

Proof. Fix $z \in V$. Let $g_z : V \to \mathbb{R}$ be given by $g_z(y) := g(y,z)$. Likewise, define $h_z(x) := h(x,z)$. Note then that $g_z \cdot h_z \in L^2(\mathscr{L}_V \otimes \mathscr{L}_V)$. It follows that

$$\int_{V\times V} f(x,y)g(y,z)h(x,z)d\mu_{V\times V}(x,y) = \int_{V\times V} f(x,y)g_z(y)h_z(x)d\mu_{V\times V}(x,y) = 0.$$

By Theorem 6.21, we have that

$$\int_{V\times V\times V} f(x,y)g(y,z)h(x,z)d\mu_{V\times V\times V}(x,y,z) = \int_V \left[\int_{V\times V} f(x,y)g(y,z)h(x,z)d\mu_{V\times V}(x,y)\right] d\mu_V(z) = 0.$$

Proof (of Theorem 17.3). We first show that we can assume that each E_{ij} belongs to $\mathscr{L}_V \otimes \mathscr{L}_V$. Indeed, let $f_{ij} := \mathbb{E}[1_{E_{ij}} | \mathscr{L}_V \otimes \mathscr{L}_V]$.[2] Then by three applications of the previous lemma, we have

$$\int_{V\times V\times V} f_{12}f_{23}f_{13}d\mu_{V\times V\times V} = \int_{V\times V\times V} f_{12}f_{23}1_{13}d\mu_{V\times V\times V} = \int_{V\times V\times V} f_{12}1_{23}1_{13}d\mu_{V\times V\times V} = \int_{V\times V\times V} 1_{12}1_{23}1_{13}d\mu_{V\times V\times V} = 0. \quad (*)$$

Let $G_{ij} := \{(u,v) \in V\times V \ : \ f_{ij}(u,v) \ge \frac{\varepsilon}{2}\}$. Observe that each G_{ij} belongs to $\mathscr{L}_V \otimes \mathscr{L}_V$ and

$$\mu(E_{ij} \setminus G_{ij}) = \int_{V\times V} 1_{E_{ij}}(1 - 1_{G_{ij}})d\mu_{V\times V} = \int_{V\times V} f_{ij}(1 - 1_{G_{ij}})d\mu_{V\times V} \le \frac{\varepsilon}{2}.$$

By $(*)$ we have

$$\int_{V\times V} 1_{G_{12}}1_{G_{23}}1_{G_{13}}d\mu_{V\times V} = 0.$$

[1] Given a binary relation R on a set X, we write R^{-1} for the binary relation on X given by $(x,y) \in R^{-1}$ if and only if $(y,x) \in R$.

[2] Here, for $f \in L^2(\mathscr{L}_{V\times V})$, $\mathbb{E}[f|\mathscr{L}_V \otimes \mathscr{L}_V]$ denotes the conditional expectation of f onto the subspace $L^2(\mathscr{L}_V \otimes \mathscr{L}_V)$.

Thus, if the nonstandard triangle removal lemma is true for sets belonging to $\mathscr{L}_V \otimes \mathscr{L}_V$, we can find hyperfinite $F_{ij} \subseteq V \times V$ such that $\mu(G_{ij} \setminus F_{ij}) < \frac{\varepsilon}{2}$ and such that $(\dagger\dagger)$ holds. Since $\mu(E_{ij} \setminus F_{ij}) < \varepsilon$, the F_{ij} are as desired.

Thus, we may now assume that each E_{ij} belongs to $\mathscr{L}_V \otimes \mathscr{L}_V$. Consequently, there are elementary sets H_{ij} such that $\mu(E_{ij} \triangle H_{ij}) < \frac{\varepsilon}{6}$. By considering the boolean algebra generated by the sides of the boxes appearing in the description of H_{ij}, we obtain a partition $V = V_1 \sqcup \cdots \sqcup V_n$ of V into finitely many hyperfinite subsets of V such that each H_{ij} is a union of boxes of the form $V_k \times V_l$ for $k, l \in \{1, \ldots, n\}$. Let

$$F_{ij} := \bigcup \{V_k \times V_l \ : \ V_k \times V_l \subseteq H_{ij}, \ \mu(V_k \times V_l) > 0, \text{ and } \mu(E_{ij} \cap (V_k \times V_l)) > \frac{2}{3}\mu(V_k \times V_l)\}.$$

Clearly each F_{ij} is hyperfinite. Note that

$$\mu(H_{ij} \setminus F_{ij}) = \mu((H_{ij} \setminus F_{ij}) \cap E_{ij}) + \mu((H_{ij} \setminus F_{ij}) \setminus E_{ij}) \leq \frac{2}{3}\mu(H_{ij} \setminus F_{ij}) + \frac{\varepsilon}{6},$$

whence $\mu(H_{ij} \setminus F_{ij}) \leq \frac{\varepsilon}{2}$ and thus $\mu(E_{ij} \setminus F_{ij}) \leq \frac{\varepsilon}{6} + \frac{\varepsilon}{2} < \varepsilon$. It remains to show that $(\dagger\dagger)$ holds. Towards a contradiction, suppose that (u, v, w) witnesses that $(\dagger\dagger)$ is false. Take $k, l, m \in \{1, \ldots, n\}$ such that $u \in V_k$, $v \in V_l$, and $w \in V_m$. Since $(u, v) \in F_{12}$, we have that $\mu(E_{12} \cap (V_k \times V_l)) > \frac{2}{3}\mu(V_k \times V_l)$. Consequently, $\mu(E_{12} \times V_m) > \frac{2}{3}\mu(V_k \times V_l \times V_m)$. Similarly, we have that $\mu(E_{23} \times V_k), \mu(E_{13} \times V_l) > \frac{2}{3}\mu(V_k \times V_l \times V_m)$. Thus, by elementary probability considerations, it follows that

$$\int_{V \times V} 1_{E_{12}} 1_{E_{23}} 1_{E_{13}} d\mu_{V \times V} > 0,$$

contradicting $(\dagger)$.

We should note that one can prove Szemeredi's theorem in the style of this chapter by first proving an appropriate removal lemma called the *Hypergraph removal lemma* and then coding arithmetic progressions by an appropriate hypergraph generalization of the argument given above. For more details, see [21].

17.2 Szemerédi Regularity Lemma

Suppose that (V, E) is a finite graph. For two nonempty subsets X, Y of V, we define the *density of arrows between X and Y* to be the quantity

$$d(X,Y) := \delta(E, X \times Y) = \frac{|E \cap (X \times Y)|}{|X||Y|}.$$

For example, if every element of X is connected to every element of Y by an edge, then $d(X, Y) = 1$. Fix $\varepsilon \in \mathbb{R}^{>0}$. We say that X and Y as above are *ε-pseudorandom* if whenever $A \subseteq X$ and $B \subseteq Y$ are such that $|A| \geq \varepsilon|X|$ and $|B| \geq \varepsilon|Y|$, then $|d(A,B) - d(X,Y)| < \varepsilon$. In other words, as long as A and B contain at least an ε proportion of the elements of X and Y respectively, then $d(A, B)$ is essentially the same as $d(X, Y)$, so the edges between X and Y are distributed in a sort of random fashion.

If $X = \{x\}$ and $Y = \{y\}$ are singletons, then clearly X and Y are ε-pseudorandom for any ε. Thus, any finite graph can trivially be partitioned into a finite number of ε-pseudorandom pairs by partitioning the graph into singletons. Szemerédi's Regularity Lemma essentially says that one can do much better in the sense that there is a constant $C(\varepsilon)$ such that any finite graph has an "ε-pseudorandom partition" into at most $C(\varepsilon)$ pieces. Unfortunately, the previous sentence is not entirely accurate as there is a bit of error that we need to account for.

Suppose that $V_1, \ldots, V_m$ is a partition of V into m pieces. Set

$$R := \{(i, j) \mid 1 \leq i, j \leq m, \ \ V_i \text{ and } V_j \text{ are } \varepsilon\text{-pseudorandom}\}.$$

We say that the partition is *ε-regular* if $\sum_{(i,j)\in R} \frac{|V_i||V_j|}{|V|^2} > (1 - \varepsilon)$. This says that, in some sense, almost all of the pairs of points are in ε-pseudorandom pairs. We can now state:

Theorem 17.6 (Szemerédi's Regularity Lemma). *For any $\varepsilon \in \mathbb{R}^{>0}$, there is a constant $C(\varepsilon)$ such that any graph (V,E) admits an ε-regular partition into $m \leq C(\varepsilon)$ pieces.*

As in the previous section, the regularity lemma is equivalent to a nonstandard version of the lemma. We leave the proof of the equivalence as an exercise to the reader.

Proposition 17.7. *Szemerédi's Regularity Lemma is equivalent to the following statement: for any ε and any hyperfinite graph (V,E), there is a* finite *partition $V_1,\ldots,V_m$ of V into internal sets and a subset $R \subseteq \{1,\ldots,m\}^2$ such that:*

- *for $(i,j) \in R$, V_i and V_j are* internally *ε-pseudorandom: for all* internal *$A \subseteq V_i$ and $B \subseteq V_j$ with $|A| \geq \varepsilon|V_i|$ and $|B| \geq \varepsilon|V_j|$, we have $|d(A,B) - d(V_i,V_j)| < \varepsilon$; and*
- *$\sum_{(i,j)\in R} \frac{|V_i||V_j|}{|V|^2} > (1-\varepsilon)$.*

We will now prove the above nonstandard equivalent of the Szemerédi Regularity Lemma. Fix ε and a hyperfinite graph (V,E). Set $f := \mathbb{E}[1_E | \mathscr{L}_V \otimes \mathscr{L}_V]$. The following calculation will prove useful: Suppose that $A,B \subseteq V$ are internal and $\frac{|A|}{|V|}$ and $\frac{|B|}{|V|}$ are noninfinitesimal. Then (♣):

$$\begin{aligned}\int_{A\times B} f d(\mu_V \otimes \mu_V) &= \int_{A\times B} 1_E d\mu_{V\times V} \quad \text{by the definition of } f\\ &= \operatorname{st}\left(\frac{|E\cap(A\times B)|}{|V|^2}\right)\\ &= \operatorname{st}\left(\frac{|E\cap(A\times B)|}{|A||B|}\right)\operatorname{st}\left(\frac{|A||B|}{|V|^2}\right)\\ &= \operatorname{st}(d(A,B))\operatorname{st}\left(\frac{|A||B|}{|V|^2}\right).\end{aligned}$$

Fix $r \in \mathbb{R}^{>0}$, to be determined later. Now, since f is $\mu_V \otimes \mu_V$-integrable, there is a $\mu_V \otimes \mu_V$-simple function $g \leq f$ such that $\int (f-g)d(\mu_V \otimes \mu_V) < r$. Set $C := \{\omega \in V\times V \mid f(\omega) - g(\omega) \geq \sqrt{r}\} \in s_V \otimes s_V$. Then $(\mu_V \otimes \mu_V)(C) < \sqrt{r}$, for otherwise

$$\int (f-g)d(\mu_V\otimes\mu_V) \geq \int_C (f-g)d(\mu_V\otimes\mu_V) \geq \int_C \sqrt{r}d(\mu_V\otimes\mu_V) \geq \sqrt{r}\sqrt{r} = r.$$

By Fact 6.10, there is an elementary set $D \in s_V \otimes s_V$ that is a finite, disjoint union of rectangles of the form $V' \times V''$, with $V', V'' \subseteq V$ internal sets, such that $C \subseteq D$ and $(\mu_V\otimes\mu_V)(D) < \sqrt{r}$. In a similar way, we may assume that the level sets of g (that is, the sets on which g takes constant values) are elementary sets (Exercise). We now take a finite partition $V_1,\ldots,V_m$ of V into internal sets such that g and 1_D are constant on each rectangle $V_i \times V_j$. For ease of notation, set d_{ij} to be the constant value of g on $V_i \times V_j$.

Claim: If $\mu_V(V_i), \mu_V(V_j) \neq 0$ and $(V_i \times V_j) \cap D = \emptyset$, then V_i and V_j are internally $2\sqrt{r}$-pseudorandom.

Proof of Claim: Since $C \subseteq D$, we have that $(V_i \times V_j) \cap C = \emptyset$, whence

$$d_{ij} \leq f(\omega) < d_{ij} + \sqrt{r} \text{ for } \omega \in V_i \times V_j. \quad (♣♣).$$

Now suppose that $A \subseteq V_i$ and $B \subseteq V_j$ are such that $|A| \geq 2\sqrt{r}|V_i|$ and $|B| \geq 2\sqrt{r}|V_j|$. In particular, $\frac{|A|}{|V_i|}$ and $\frac{|B|}{|V_j|}$ are noninfinitesimal. Since $\mu_V(V_i), \mu_V(V_j) > 0$, it follows that $\frac{|A|}{|V|}$ and $\frac{|B|}{|V|}$ are noninfinitesimal and the calculation (♣) applies. Integrating the inequalities (♣♣) on $A \times B$ yields:

$$d_{ij}\operatorname{st}\left(\frac{|A||B|}{|V|^2}\right) \leq \operatorname{st}(d(A,B))\operatorname{st}\left(\frac{|A||B|}{|V|^2}\right) < (d_{ij}+\sqrt{r})\operatorname{st}\left(\frac{|A||B|}{|V|^2}\right).$$

We thus get:

$$|d(A,B)-d(V_i,V_j)| \leq |d(A,B)-d_{ij}| + |d(V_i,V_j)-d_{ij}| < 2\sqrt{r}.$$

By the Claim, we see that we should choose $r < (\frac{\varepsilon}{2})^2$, so V_i and V_j are internally ε-pseudorandom when V_i and V_j are non-null and satisfy $(V_i \times V_j) \cap D = \emptyset$. It remains to observe that the ε-pseudorandom pairs almost cover all pairs of vertices. Let $R := \{(i,j) \mid V_i \text{ and } V_j \text{ are } \varepsilon\text{-pseudorandom}\}$. Then

$$\begin{aligned}
\mathrm{st}\left(\sum_{(i,j)\in R} \frac{|V_i||V_j|}{|V|^2}\right) &= \mu_{V\times V}\left(\bigcup_{(i,j)\in R}(V_i \times V_j)\right) \\
&\geq \mu_{V\times V}((V\times V)\setminus D) \\
&> 1-\sqrt{r} \\
&> 1-\varepsilon.
\end{aligned}$$

This finishes the proof of the Claim and the proof of the Szemerédi Regularity Lemma.

Chapter 18
Approximate groups

In this chapter, we describe a recent application of nonstandard methods to multiplicative combinatorics, namely to the structure theorem for finite approximate groups. The general story is much more complicated than the rest of the material in this book and there are already several good sources for the complete story (see [7] or [46]), so we content ourselves to a summary of some of the main ideas. Our presentation will be similar to the presentation from [46].

One important convention will be important to keep in mind. In this chapter, we follow the custom in the literature of writing, for X a subset of a group G and $n \in \mathbb{N}$, $X^n := \{x_1 \cdots x_n \ : \ x_1, \ldots, x_n \in X\}$ (so X^n does *not* mean the n-fold Cartesian power of X).

18.1 Statement of definitions and the main theorem

In this chapter, $(G, \cdot)$ denotes an arbitrary group and $K \in \mathbb{R}^{\geq 1}$. (Although using K for a real number clashes with the notation used throughout the rest of this book, it is standard in the area.) By a *symmetric* subset of G, we mean a set that contains the identity of G and is closed under taking inverse.

Definition 18.1. $X \subseteq G$ is a *K-approximate group* if X is symmetric and X^2 can be covered by at most K left translates of X, that is, there are $g_1, \ldots, g_m \in G$ with $m \leq K$ such that $X^2 \subseteq \bigcup_{i=1}^m g_i X$.

Example 18.2.

1. A 1-approximate subgroup of G is simply a subgroup of G.
2. If $X \subseteq G$ is finite, then X is a $|X|$-approximate subgroup of G.

The second example highlights that, in order to try to study the general structure of finite K-approximate groups, one should think of K as fixed and "small" and then try to classify the finite K-approximate groups X, where X has cardinality much larger than K.

Exercise 18.3. Suppose that $(G, +)$ is an abelian group. For distinct $v_1 \ldots, v_r \in G$ and (not necessarily distinct) $N_1, \ldots, N_r \in \mathbb{N}$, set

$$P(\mathbf{v}, \mathbf{N}) := \{a_1 v_1 + \cdots + a_r v_r \ : \ a_i \in \mathbb{Z}, \ |a_i| \leq N_i\}.$$

Show that $P(\mathbf{v}, \mathbf{N})$ is a 2^r-approximate subgroup of G.

The approximate subgroups appearing in the previous exercise are called *symmetric generalized arithmetic progressions* and the number r of generators is called the *rank* of the progression. The *Freiman Theorem for abelian groups* (due to due to Freiman [18] for $\mathbb{Z}$ and to Green and Rusza [23] for a general abelian group) says that approximate subgroups of abelian groups are "controlled" by symmetric generalized arithmetic progressions:

Theorem 18.4. *There are constants r_K, C_K such that the following hold: Suppose that G is an abelian group and $A \subseteq G$ is a finite K-approximate group. Then there is a finite subgroup H of G and a symmetric generalized arithmetic progression $P \subseteq G/H$ such that P has rank at most r_K, $\pi^{-1}(P) \subseteq \Sigma_4(A)$, and $|P| \geq C_K \cdot \frac{|A|}{|H|}$.*

Here, $\pi : G \to G/H$ is the quotient map. For a while it was an open question as to whether there was a version of the Freiman theorem that held for finite approximate subgroups of arbitrary groups. Following a breakthrough by Hrushovski [26], Breuillard, Green, and Tao [7] were able to prove the following general structure theorem for approximate groups.

Theorem 18.5. *There are constants r_K, s_K, C_K such that the following hold: Suppose that G is a group and $A \subseteq G$ is a finite K-approximate group. Then there is a finite subgroup $H \subseteq G$, a* noncommutative progression *of rank at most r_K whose generators generate a nilpotent group of step at most s_K such that $\pi^{-1}(P) \subseteq A^4$ and $|P| \geq C_K \cdot \frac{|A|}{|H|}$.*

Here, $\pi : G \to G/H$ is once again the quotient map. To understand this theorem, we should explain the notion of noncommutative progression.

Suppose that G is a group, $v_1, \ldots, v_r \in G$ are distinct, and $N_1, \ldots, N_r > 0$ are (not necessarily distinct) natural numbers. The noncommutative progression generated by $v_1, \ldots, v_r$ with dimensions $N_1, \ldots, N_r$ is the set of words on the alphabet $\{v_1, v_1^{-1}, \ldots, v_r, v_r^{-1}\}$ such that the total number of occurrences of v_i and v_i^{-1} is at most N_i for each $i = 1, \ldots, r$; as before, r is called the rank of the progression. In general, noncommutative progressions need not be approximate groups (think free groups). However, if $v_1, \ldots, v_r$ generate a nilpotent subgroup of G of step s, then for $N_1, \ldots, N_r$ sufficiently large, the noncommutative progression is in fact a K-approximate group for K depending only on r and s. (See, for example, [43, Chapter 12].)

18.2 A special case: approximate groups of finite exponent

To illustrate some of the main ideas of the proof of the Breuillard-Green-Tao theorem, we prove a special case due to Hrushovski [26]:

Theorem 18.6. *Suppose that $X \subseteq G$ is a finite K-approximate group. Assume that X^2 has exponent e, that is, for every $x \in X^2$, we have $x^e = 1$. Then X^4 contains a subgroup H of $\langle X \rangle$ such that X can be covered by L left cosets of H, where L is a constant depending only on K and e.*

Here, $\langle X \rangle$ denotes the subgroup of G generated by X. Surprisingly, this theorem follows from the simple observation that the only connected Lie group which has an identity neighborhood of finite exponent is the trivial Lie group consisting of a single point. But how do continuous objects such as Lie groups arise in proving a theorem about finite objects like finite approximaite groups? The key insight of Hrushovski is that ultraproducts of finite K-approximate groups are naturally "modeled" in a precise sense by second countable, locally compact groups and that, using a classical theorem of Yamabe, this model can be perturbed to a Lie model.

More precisely, for each $i \in \mathbb{N}$, suppose that $X_i \subseteq G_i$ is a finite K-approximate group. We set $X := \prod_{\mathscr{U}} X_i$, which, by transfer, is a hyperfinite K-approximate subgroup of $G := \prod_{\mathscr{U}} G_i$. In the rest of this chapter, unless specified otherwise, X and G will denote these aforementioned ultraproducts. By a *monadic* subset of G we mean a countable intersection of internal subsets of G. Also, $\langle X \rangle$ denotes the subgroup of G generated by X.

Theorem 18.7. *There is a monadic subset $o(X)$ of X^4 such that $o(X)$ is a normal subgroup of $\langle X \rangle$ such that the quotient $\mathscr{G} := \langle X \rangle / o(X)$ has the structure of a second countable, locally compact group. Moreover, letting $\pi : \langle X \rangle \to \mathscr{G}$ denote the quotient map, we have:*

1. *The quotient $\langle X \rangle / o(X)$ is* bounded*, meaning that for all internal sets $A, B \subseteq \langle X \rangle$ with $o(X) \subseteq A$, finitely many left translates of A cover B.*
2. *$Y \subseteq \mathscr{G}$ is compact if and only if $\pi^{-1}(Y)$ is monadic; in particular, $\pi(X)$ is compact.*
3. *If $Y \subseteq G$ is internal and contains $o(X)$, then Y contains $\pi^{-1}(U)$ for some open neighborhood of the identity in $\mathscr{G}$.*
4. *$\pi(X^2)$ is a compact neighborhoods of the identity in $\mathscr{G}$.*

Let us momentarily assume that Theorem 18.7 holds and see how it is used to prove Theorem 18.6. As usual, we first prove a nonstandard version of the desired result.

Theorem 18.8. *Suppose that $X \subseteq G$ is a hyperfinite K-approximate group such that X^2 has exponent e. Then X^4 contains an internal subgroup H of G such that $o(X) \subseteq H$.*

Proof. Let U be an open neighborhood of the identity in $\mathscr{G}$ with $\pi^{-1}(U) \subseteq X^4$ such that U is contained in $\pi(X^2)$, whence U has exponent e. By the Gleason-Yamabe theorem [48], there is an open subgroup $\mathscr{G}'$ of $\mathscr{G}$ and normal $N \trianglelefteq \mathscr{G}'$ with $N \subseteq U$ such that $\mathscr{H} := \mathscr{G}'/N$ is a connected Lie group. Let $Y := X \cap \pi^{-1}(\mathscr{G}')$ and let $\rho : \langle Y \rangle \to \mathscr{H}$ be the composition of π with the quotient map $\mathscr{G}' \to \mathscr{H}$. Since $\mathscr{G}'$ is clopen in $\mathscr{G}$, $\pi^{-1}(\mathscr{G}')$ is both monadic and co-monadic (the complement of a monadic, also known as *galactic*), whence internal by saturation; it follows that Y is also internal. Since the image of $U \cap \mathscr{G}'$ in $\mathscr{H}$ is also open, it follows that $\mathscr{H}$ is a connected Lie group with an identity neighborhood of finite exponent. We conclude that $\mathscr{H}$ is trivial, whence $\ker(\rho) = Y = \langle Y \rangle$ is the desired internal subgroup of G contained in X^4.

Remark 18.9. The passage from $\mathscr{G}$ to the Lie subquotient $\mathscr{G}'/N$ is called the *Hrushovski Lie Model Theorem*. More precisely, [7] abstracts the important properties of the quotient map $\pi : \langle X \rangle \to \mathscr{G}$ and calls any group morphism onto a second countable, locally compact group satisfying these properties a *good model*. In the proof of Theorem 18.8, we actually showed that the good model $\pi : \langle X \rangle \to \mathscr{G}$ can be replaced by a good model $\rho : \langle Y \rangle \to \mathscr{H}$ onto a connected Lie group. One can show that Y is also an approximate group (in fact, it is a K^6-approximate group) that is closely related to the original approximate group X, whence the Hrushovski Lie model theorem allows one to study ultraproducts of K-approximate groups by working with the connected Lie groups that model them. For example, the proof of Theorem 18.4 actually proceeds by induction on the dimension of the corresponding Lie model. To be fair, the proof of Theorem 18.4 actually requires the use of *local Lie groups* and, in particular, uses the local version of Yamabe's theorem, whose first proof used nonstandard analysis [20].

Proof (of Theorem 18.6). Suppose, towards a contradiction, that the theorem is false. For each L, let G_L be a group and $X_L \subseteq G_L$ a finite K-approximate group such that X_L^2 has exponent e and yet, for any finite subgroup H of $\langle X_L \rangle$ contained in X_L^4, we have that X_L is not covered by L cosets of H. Let $X := \prod_{\mathscr{U}} X_L$ and $G := \prod_{\mathscr{U}} G_L$. By transfer, X is a K-approximate subgroup of G such that X^2 has exponent e. By Theorem 18.8, X^4 contains an internal subgroup $H \supseteq o(X)$ of $\langle X \rangle$. Without loss of generality, we may write $H := \prod_{\mathscr{U}} H_L$ with H_L a subgroup of G_L contained in X_L^4. Since the quotient is bounded by Theorem 18.7, there is $M \in \mathbb{N}$ such that M left translates of H cover X^4. Thus, for $\mathscr{U}$-almost all L, M left translates of H_L cover X_L; taking $L > M$ yields the desired contradiction.

We now turn to the proof of Theorem 18.7. Hrushovski's original proof used some fairly sophisticated model theory. A key insight of Breuillard-Green-Tao was that a proof that relied only on fairly elementary combinatorics and nonstandard methods could be given. The following result is the combinatorial core of their proof. It, and the easy lemma after it, do not follow the convention that X is a hyperfinite K-approximate group.

Theorem 18.10 (Sanders-Croot-Sisask). *Given K and $\delta > 0$, there is $\varepsilon > 0$ so that the following holds: Suppose that X is a finite K-approximate subgroup of G. Suppose that $Y \subseteq X$ is symmetric and $|Y| \geq \delta|X|$. Then there is a symmetric $E \subseteq G$ such that $|E| \geq \varepsilon|X|$ and $(E^{16})^X \subseteq Y^4$.*

Lemma 18.11. *Let $X \subseteq G$ be a finite K-approximate group and $S \subseteq G$ symmetric such that $S^4 \subseteq X^4$ and $|S| \geq c|X|$ for some $c > 0$. Then X^4 can be covered by K^7/c left cosets of S^2.*

We now return to our assumption that X is a hyperfinite K-approximate subgroup of G.

Proposition 18.12. *There is a descending sequence*

$$X^4 =: X_0 \supseteq X_1 \supseteq X_2 \supseteq \cdots \supseteq X_n \supseteq \cdots$$

of internal, symmetric subsets of G such that:

(i) $X_{n+1}^2 \subseteq X_n$;
(ii) $X_{n+1}^X \subseteq X_n$;
(iii) X^4 *is covered by finitely many left cosets of* X_n.

Proof. Suppose that $Y \subseteq G$ is internal, symmetric, $Y^4 \subseteq X^4$, and X^4 can be covered by finitely many left cosets of Y. We define a new set $\tilde{Y}$ with these same properties. First, take $\delta > 0$ such that $|Y| \geq \delta |X^4|$; such δ exists since X^4 can be covered by finitely many left cosets of Y. By the transfer of Theorem 18.10, there is an internal, symmetric $S \subseteq Y^4$ such that $|S| \geq \varepsilon |X^4|$ and $(S^{16})^X \subseteq Y^4$. Let $\tilde{Y} := S^2$. Note that $\tilde{Y}$ has the desired properties, the last of which follows from the preceding lemma.

We now define a sequence $Y_0, Y_1, Y_2, \ldots,$ of internal subsets of X^4 satisfying the above properties by setting $Y_0 := X$ and $Y_{n+1} := \tilde{Y}_n$. Finally, setting $X_n := Y_n^4$ yields the desired sequence.

Proof (of Theorem 18.7). Take (X_n) as guaranteed by Proposition 18.12. We set $o(X) := \bigcap_n X_n$, a monadic subset of X^4. It is clear from (i) and (ii) that $o(X)$ is a normal subgroup of $\langle X \rangle$. We can topologize $\langle X \rangle$ by declaring, for $a \in \langle X \rangle$, $\{aX_n \ : \ n \in \mathbb{N}\}$ to be a neighborhood base for a. The resulting space is not Hausdorff, but it is clear that the quotient space $\langle X \rangle / o(X)$ is precisely the separation of $\langle X \rangle$. It is straightforward to check that the resulting space is separable and yields a group topology on $\mathscr{G}$. Now one uses the boundedness property (proven in the next paragraph) to show that $\mathscr{G}$ is locally compact; see [46] for details.

To show that it is bounded, suppose that $A, B \subseteq \langle X \rangle$ are such that $o(X) \subseteq A$. We need finitely many left cosets of A to cover B. Take n such that $X_n \subseteq A$ and take m such that $B \subseteq (X^4)^m$. Since X^4 is a K^4-approximate group, $(X^4)^m \subseteq E \cdot X^4$ for some finite E. By (iii), we have that $X^4 \subseteq F \cdot X_n$ for some finite F. It follows that $B \subseteq EFA$, as desired.

The proof that $Y \subseteq \mathscr{G}$ is compact if and only if $\pi^{-1}(Y)$ is monadic is an exercise left to the reader (or, once again, one can consult [46]). To see the moreover part, note that

$$\pi^{-1}(\pi(X)) = \{x \in \langle X \rangle \ : \ \text{there is } y \in X \text{ such that } x^{-1}y \in \bigcap_n X_n\}.$$

In particular, $\pi^{-1}(\pi(X)) \subseteq X^5$ and, by saturation, we actually have

$$\pi^{-1}(\pi(X)) = \{x \in X^5 \ : \ \text{for all } n \text{ there is } y \in X \text{ such that } x^{-1}y \in X_n\}.$$

From this description of $\pi^{-1}(\pi(X))$, we see that it is monadic, whence $\pi(X)$ is compact.

To prove (3), suppose that Y is an internal subset of G containing $o(X)$. Take n such that $X_n \subseteq Y$. Thus, $\pi^{-1}(\pi(X_{n+1})) \subseteq X_n \subseteq Y$ and $\pi(X_{n+1})$ is open in $\mathscr{G}$.

Finally, to see that $\pi(X^2)$ is a neighborhood of the identity in $\mathscr{G}$, first observe that since X^4 is covered by finitely many left cosets of X, the neighborhood $\pi(X^4)$ of the identity is covered by finitely many left cosets of the compact set $\pi(X)$, whence $\pi(X)$ has nonempty interior and thus $\pi(X^2) = \pi(X) \cdot \pi(X)^{-1}$ is a neighborhood of the identity in $\mathscr{G}$.

Part V
Appendix

Appendix A
Foundations of nonstandard analysis

A.1 Foundations

In this appendix we will revise all the basic notions and principles that we presented in Chapter 2 and put them on firm foundations. As it is customary in the foundations of mathematics, we will work in a set-theoretic framework as formalized by Zermelo-Fraenkel set theory with choice ZFC. Since the purpose of this book is not a foundational one, we will only outline the main arguments, and then give precise bibliographic references where the interested reader can find all proofs worked out in detail.

A.1.1 Mathematical universes and superstructures

Let us start with the notion of a mathematical universe, which formalizes the idea of a sufficiently large collection of mathematical objects that contains all that one needs when applying nonstandard methods.

Definition A.1. A *universe* $\mathbb{U}$ is a nonempty collection of "mathematical objects" that satisfies the following properties:

1. The numerical sets $\mathbb{N},\mathbb{Z},\mathbb{Q},\mathbb{R},\mathbb{C}\in\mathbb{U}$;
2. If $a_1,\ldots,a_k\in\mathbb{U}$ then also the tuple $\{a_1,\ldots,a_k\}$ and the ordered tuple $(a_1,\ldots,a_k)$ belong to $\mathbb{U}$;
3. If the family of sets $\mathscr{F}\in\mathbb{U}$ then also its union $\bigcup\mathscr{F}=\bigcup_{F\in\mathscr{F}}F\in\mathbb{U}$;
4. If the sets $A,B\in\mathbb{U}$ then also the *Cartesian product* $A\times B$, the *powerset* $\mathscr{P}(A)=\{A'\mid A'\subseteq A\}$, and the *function set* $\mathrm{Fun}(A,B)=\{f\mid f:A\to B\}$ belong to $\mathbb{U}$;
5. $\mathbb{U}$ is *transitive*, that is, $a\in A\in\mathbb{U}\Rightarrow a\in\mathbb{U}$.

Notice that a universe $\mathbb{U}$ is necessarily closed under subsets; indeed if $A'\subseteq A\in\mathscr{U}$, then $A'\in\mathscr{P}(A)\in\mathbb{U}$, and hence $A'\in\mathbb{U}$, by transitivity. Thus, if the sets $A,B\in\mathbb{U}$ then also the the *intersection* $A\cap B$ and the *set-difference* $A\setminus B$ belong to $\mathbb{U}$; moreover, by combining properties 2 and 3, one obtains that also the *union* $A\cup B=\bigcup\{A,B\}\in\mathbb{U}$.

Remark A.2. It is a well-known fact that all "mathematical objects" used in the ordinary practice of mathematics, including numbers, sets, functions, relations, ordered tuples, and Cartesian products, can all be coded as sets. Recall that, in ZFC, an ordered pair (a,b) is defined as the so-called *Kuratowski pair* $\{\{a\},\{a,b\}\}$; in fact, it is easily shown that by adopting that definition one has the characterizing property that $(a,b)=(a',b')$ if and only if $a=a'$ and $b=b'$. Ordered tuples are defined inductively by letting $(a_1,\ldots,a_k,a_{k+1})=((a_1,\ldots,a_k),a_{k+1})$. A binary relation R is defined as a set of ordered pairs; so, the notion of a relation is identified with the set of pairs that that satisfy it. A function f is a relation such that every element a in the domain is in relation with a unique element b of the range, denoted $b=f(a)$; so, the notion of a function is identified with its graph. As for numbers, the natural numbers $\mathbb{N}_0$ of ZFC are defined as the set of *von Neumann naturals*: $0=\emptyset$ and, recursively, $n+1=n\cup\{n\}$, so that each natural number $n=\{0,1,\ldots,n-1\}$ is identified with the set of its predecessors; the integers $\mathbb{Z}$ are then defined as a suitable quotient of $\mathbb{N}\times\mathbb{N}$, and the rationals $\mathbb{Q}$ as a suitable quotient of $\mathbb{Z}\times\mathbb{Z}$; the real numbers $\mathbb{R}$ are usually defined suitable sets of rational numbers, namely the *Dedekind cuts*; the complex numbers $\mathbb{C} = \mathbb{R} \times \mathbb{R}$ are defined as ordered pairs of real numbers, where the pair (a,b) is denoted $a+ib$. (See, *e.g.*, [25].)

We remark that the above definitions are instrumental if one works within axiomatic set theory, where all notions must be reduced to the sole notion of a set; however, in the ordinary practice of mathematics, one can safely take the ordered tuples, the relations, the functions, and the natural numbers as primitive objects of a different nature with respect to sets.

For convenience, in the following we will consider *atoms*, that is, primitive objects that are not sets.[1] A notion of a universe that is convenient to our purposes is the following.

Definition A.3. Let X be a set of atoms. The *superstructure* over X is the union $\mathbb{V}(X) := \bigcup_{n\in\mathbb{N}_0} V_n(X)$, where $V_0(X) = X$, and, recursively, $V_{n+1}(X) = V_n(X) \cup \mathscr{P}(V_n(X))$.

Proposition A.4. *Let X be a set of atoms that includes (a copy of) $\mathbb{N}$. Then the superstructure $\mathbb{V}(X)$ is a universe in the sense of Definition A.1.*[2]

Proof. See [8, §4.4].

Remark A.5. In set theory, one considers the universe $\mathbf{V} = \bigcup_\gamma V_\gamma$ given by the union of all levels of the so-called *von Neumann cumulative hierarchy*, which is defined by transfinite recursion on the class of all ordinals by letting $V_0 = \emptyset$, $V_{\gamma+1} = \mathscr{P}(V_\gamma)$, and $V_\lambda = \bigcup_{\gamma<\lambda} V_\gamma$ if λ is a limit ordinal. Basically, the Regularity axiom was introduced in set theory to show that the above class $\mathbf{V}$ is the universal class of all sets.

Instead, the superstructures are defined by only taking the finite levels $V_n(X)$ constructed over a given set of atoms X. The main motivation for that restriction is that if one goes beyond the finite levels and allow the first infinite ordinal ω to belong to the domain of the star map, then ${}^*\omega$ would contain $\in$-descending chains $\xi \ni \xi-1 \ni \xi-2 \ni \ldots$ for every $\xi \in {}^*\omega \setminus \omega$, contradicting the Regularity axiom. Since $V_\omega = \bigcup_{n\in\omega} V_n$ would not be suitable, as it only contains finite sets, one takes an infinite set of atoms X as the ground level $V_0(X) = X$, so as to enclose (a copy of the) natural numbers in the universe.

However, we remark that if one drops the Regularity Axiom from the axiomatics ZFC, and replace it with a suitable Anti-Foundation Axiom (such as Boffa's *superuniversality axiom*), then one can construct star maps $* : \mathbf{V} \to \mathbf{V}$ from the universe all sets into itself that satisfies the *transfer principle* and κ-saturation for any given cardinal κ. (This is to be contrasted with the well-known result by Kunen about the impossibility in ZFC of non-trivial elementary extensions $j : \mathbf{V} \to \mathbf{V}$.) This kind of foundational issues are the subject matter of the so-called *nonstandard set theory* (see Remark A.11).

A.1.2 Bounded quantifier formulas

In this section we formalize the notion of "elementary property" by means of suitable formulas. It is a well-known fact that virtually all properties of mathematical objects can be described within first-order logic; in particular, one can reduce to the language of set theory grounded on the usual logic symbols plus the sole membership relation symbol. Here is the "alphabet" of our language.[3]

- *Variables*: $x, y, z, \ldots, x_1, x_2, \ldots$;
- *Logical Connectives*: $\neg$ (negation "not"); $\wedge$ (conjunction "and"); $\vee$ (disjunction "or"); $\Rightarrow$ (implication "if ... then"); $\Leftrightarrow$ (double implication "if and only if");
- *Quantifiers*: $\exists$ (existential quantifier "there exists"); $\forall$ (universal quantifier "for all");
- *Equality symbol* $=$;
- *Membership symbol* $\in$.

[1] The existence of atoms is disproved by the axioms of ZFC, where all existing objects are sets; however, axiomatic theories are easily formalized that allow a proper class of atoms. For instance, one can consider a suitably modified versions of ZFC where a unary predicate $A(x)$ for "x is an atom" is added to the language, and where the axiom of extensionality is restricted to non-atoms.

[2] Clearly, the transitivity property "$a \in A \in \mathbb{V}(X) \Rightarrow a \in \mathbb{V}(X)$" applies provided $A \notin X$.

[3] To be precise, also parentheses " (" and ") " should be included among the symbols of our alphabet.

Definition A.6. An *elementary formula* σ is a finite string of symbols in the above alphabet where it is specified a set of *free variables* $FV(\sigma)$ and a set of *bound variables* $BV(\sigma)$, according to the following rules.

- *Atomic formulas*. If x and y are variables then "$(x = y)$" and "$(x \in y)$" are elementary formulas, named *atomic formulas*, where $FV(x = y) = FV(x \in y) = \{x, y\}$ and $BV(x = y) = BV(x \in y) = \emptyset$;
- *Restricted quantifiers*. If σ is an elementary formula, $x \in FV(\sigma)$ and $y \notin BV(\sigma)$, then "$(\forall x \in y)\ \sigma$" is an elementary formulas where $FV((\forall x \in y)\ \sigma) = (FV(\sigma) \setminus \{x\}) \cup \{y\}$ and $BV((\forall x \in y)\ \sigma) = BV(\sigma) \cup \{y\}$; and similarly with the elementary formula "$(\exists x \in y)\ \sigma$" obtained by applying the existential quantifier;
- *Negation*. If σ is an elementary formula then $(\neg\sigma)$ is an elementary formula where $FV(\neg\sigma) = FV(\sigma)$ and $BV(\neg\sigma) = BV(\sigma)$;
- *Binary connectives*. If σ and τ are elementary formulas where $FV(\sigma) \cap BV(\tau) = FV(\tau) \cap BV(\sigma) = \emptyset$, then "$(\sigma \wedge \tau)$" is an elementary formula where $FV(\sigma \wedge \tau) = FV(\sigma) \cup FV(\tau)$ and $BV(\sigma \wedge \tau) = BV(\sigma \wedge \tau)$; and similarly with the elementary formulas $(\sigma \vee \tau)$, $(\sigma \Rightarrow \tau)$, and $(\sigma \Leftrightarrow \tau)$ obtained by applying the connectives $\vee$, $\Rightarrow$, and $\Leftrightarrow$, respectively.

According to the above, every elementary formula is built from *atomic formulas* (and this justifies the name "atomic"). in that an arbitrary elementary formula is obtained from atomic formulas by finitely many iterations of restricted quantifiers, negations, and binary connectives, in whatever order. Only quantifiers produces bound variables, and in fact the bound variables are those that are quantified. Notice that a variable can be quantified only if it is free in the given formula, that is, it actually appears and it has been not quantified already.

It is worth stressing that quantifications are only permitted in the *restricted forms* $(\forall x \in y)$ or $(\exists x \in y)$, where the "scope" of the quantified variable x is "restricted" by another variable y. To avoid potential ambiguities, we required that the "bounding" variable y does not appear bound itself in the given formula.

As it is customary in the practice, to simplify notation we will adopt natural short-hands. For instance, we will write "$x \neq y$" to mean "$\neg(x = y)$" and "$x \notin y$" to mean "$\neg(x \in y)$"; we will write "$\forall x_1, \ldots, x_k \in y\ \sigma$" to mean "$(\forall x_1 \in y) \ldots (\forall x_k \in y)\ \sigma$", and similarly with existential quantifiers. Moreover, we will use parentheses informally, and omit some of them whenever confusion is unlikely. So, we may write "$\forall x \in y\ \sigma$" instead of "$(\forall x \in y)\ \sigma$; or "$\sigma \wedge \tau$" instead of "$(\sigma \wedge \tau)$"; and so forth.

Another usual agreement is that negation $\neg$ binds more strongly than conjunctions $\wedge$ and disjunctions $\vee$, which in turn bind more strongly than implications $\Rightarrow$ and double implications $\Leftrightarrow$. So, we may write "$\neg\sigma \wedge \tau$" to mean "$((\neg\sigma) \wedge \tau)$"; or "$\neg\sigma \vee \tau \Rightarrow \upsilon$" to mean "$(((\neg\sigma) \vee \tau) \Rightarrow \upsilon)$"; or "$\sigma \Rightarrow \tau \vee \upsilon$" to mean "$(\sigma \Rightarrow (\tau \vee \upsilon))$".

When writing $\sigma(x_1, \ldots, x_k)$ we will mean that $x_1, \ldots, x_k$ are all and only the free variables that appear in the formula σ. The intuition is that the truth or falsity of a formula depends only on the values given to its *free variables*, whereas *bound variables* can be renamed without changing the meaning of a formula.

Definition A.7. A property of mathematical objects $A_1, \ldots, A_k$ is *expressed in elementary form* if it is written down by taking an elementary formula $\sigma(x_1, \ldots, x_k)$, and by replacing all occurrences of each free variable x_i by A_i. In this case we denote

$$\sigma(A_1, \ldots, A_k),$$

and we will refer to objects $A_1, \ldots, A_k$ as *constants* or *parameters*.[4] By a slight abuse, sometimes we will simply say *elementary property* to mean "property expressed in elementary form".

The motivation of our definition is the well-known fact that virtually *all* properties considered in mathematics can be formulated in elementary form. Below is a list of examples that include the fundamental ones. As an exercise, the reader can easily write down by him- or herself any other mathematical property that comes to his or her mind, in elementary form.

Example A.8. Each property is followed by one of its possible expressions in elementary form.[5]

1. "$A \subseteq B$": $(\forall x \in A)(x \in B)$;
2. $C = A \cup B$: $(A \subseteq C) \wedge (B \subseteq C) \wedge (\forall x \in C)(x \in A \vee x \in B)$;
3. $C = A \cap B$: $(C \subseteq A) \wedge (\forall x \in A)(x \in B \Leftrightarrow x \in C)$;

---

[4] In order to make sense, it is implicitly assumed that in every quantification $(\forall x \in A_i)$ and $(\exists x \in A_i)$, the object A_i is a *set* (not an atom).

[5] For simplicity, in each item we use short-hands for properties that have been already considered in previous items.

4. $C = A \setminus B$: $(C \subseteq A) \wedge (\forall x \in A)(x \in C \Leftrightarrow x \notin B)$;
5. $C = \{a_1, \dots, a_k\}$: $(a_1 \in C) \wedge \dots \wedge (a_k \in C) \wedge (\forall x \in C)(x = a_1 \vee \dots \vee x = a_k)$;
6. $\{a_1, \dots, a_k\} \in C$: $(\exists x \in C)(x = \{a_1, \dots, a_k\})$;
7. $C = (a,b)$: $C = \{\{a\},\{a,b\}\}$;[6]
8. $C = (a_1, \dots, a_k)$ with $k \geq 3$: Inductively, $C = ((a_1, \dots, a_{k-1}), a_k)$;
9. $(a_1, \dots, a_k) \in C$: $(\exists x \in C)(x = (a_1, \dots, a_k))$;
10. $C = A_1 \times \dots \times A_k$: $(\forall x_1 \in A_1) \dots (\forall x_k \in A_k)((a_1, \dots, a_k) \in C) \wedge (\forall z \in C)(\exists x_1 \in A_1) \dots (\exists x_k \in A_k)(z = (x_1, \dots, x_k))$;
11. R is a k-place relation on A: $(\forall z \in R)(\exists x_1, \dots, x_k \in A)(z = (x_1, \dots, x_k))$;
12. $f : A \to B$: $(f \subseteq A \times B) \wedge (\forall a \in A)(\exists b \in B)((a,b) \in f) \wedge (\forall a, a' \in A)(\forall b \in B)((a,b),(a',b) \in f \Rightarrow a = a')$;
13. $f(a_1, \dots, a_k) = b$: $((a_1, \dots, a_k), b) = (a_1, \dots, a_k, b) \in f$;
14. $x < y$ in $\mathbb{R}$: $(x,y) \in R$, where $R \subset \mathbb{R} \times \mathbb{R}$ is the order relation on $\mathbb{R}$.

It is worth remarking that a same property may be expressed both in an elementary form and in a non-elementary form. The typical examples involve the powerset operation.

Example A.9. "$\mathscr{P}(A) = B$" is trivially an elementary property of constants $\mathscr{P}(A)$ and B, but *cannot* be formulated as an elementary property of constants A and B. In fact, while the inclusion "$B \subseteq \mathscr{P}(A)$" is formalized in elementary form by "$(\forall x \in B)(\forall y \in x)(y \in A)$", the other inclusion $\mathscr{P}(A) \subseteq B$ does not admit any elementary formulation with A and B as constants. The point here is that quantifications over subsets "$(\forall x \subseteq A)(x \in B)$" are not allowed by our rules.

A.1.3 Los' Theorem

Bounded ultrapowers of superstructures and Los Theorem. Superstructure models of nonstandard analysis as star maps $* : \mathbb{V}(X) \to \mathbb{V}(Y)$ that satisfy the *transfer principle*.

By using a procedure known in logic as "induction on the complexity of formulas", one proves that the equivalences $P(A_1, \dots, A_k, f_1, \dots, f_h) \Leftrightarrow P({}^*A_1, \dots, {}^*A_k, {}^*f_1, \dots, {}^*f_h)$ hold for all elementary properties P, sets A_i, and functions f_j.

Refer to §4.4 of Chang-Keisler's "Model Theory".

DO IT.

A.1.4 Models that allow iterated hyper-extensions

In applications, we needed iterated hyper-extensions, but in the usual frameworks of nonstandard analysis currently found in the literature, such extensions cannot be accommodated directly. To this end, one needs to construct a different standard universe each time, that contains the previous nonstandard universe. A neat way to overcome this problem is to consider a superstructure model of nonstandard analysis where the standard and the nonstandard universe coincide. Clearly, in this case a hyper-extension also belongs to the standard universe, and so one can apply the star map to it.[7]

Theorem A.10. *Let κ, μ be infinite cardinals. Then there exist base sets $X_0 \subset X$ of cardinality μ^κ and star maps $* : \mathbb{V}(X) \to \mathbb{V}(X)$ such that:*

1. *(a copy of) the real numbers $\mathbb{R} \subset Y$;*
2. *${}^*x = x$ for every $x \in X_0$, and hence ${}^*r = r$ for every $r \in \mathbb{R}$;*
3. *${}^*X = X$;*
4. transfer principle. *For every bounded quantifier formula $\varphi(x_1, \dots, x_n)$ and for every $a_1, \dots, a_n \in \mathbb{V}(X)$:*

$$\varphi(a_1, \dots, a_n) \Longleftrightarrow \varphi({}^*a_1, \dots, {}^*a_n);$$

[6] Recall that ordered pairs $(a,b) = \{\{a\},\{a,b\}\}$ were defined as *Kuratowski pairs*.

[7] We remark that the notion of "iterated hyper-image" " does not even make sense in Nelson's *Internal Set Theory* IST, and in the subsequent axiomatic theory elaborated upon that approach.

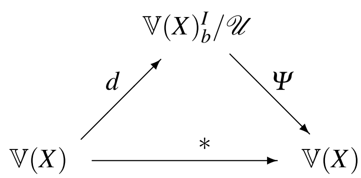
5. The κ^+-saturation principle holds.

Proof. Since $\mu^\kappa \geq \mathfrak{c}$ has at least the size of the *continuum*, we can pick a base set X of cardinality μ^κ that contains (a copy of) the real numbers $\mathbb{R}$, and such that the relative complement $X_0 = X \setminus \mathbb{R}$ has cardinality μ^κ. For every $x \in \kappa$, let $\langle x \rangle = \{a \in \mathrm{Fin}(\kappa) \mid i \in a\}$ be the set of all finite parts of κ that contains x. It is readily seen that the family $\{\langle x \rangle \mid x \in \kappa\}$ has the *finite intersection property*, and so it can be extended to an ultrafilter $\mathscr{U}$ on $I = \mathrm{Fin}(\kappa)$. We now inductively define maps $\Psi_n : V_n(X)^I/\mathscr{U} \to V_n(X)$ as follows.

Since $\mu^\kappa = |X| \leq |X^I/\mathscr{U}| \leq |X|^{|I|} = (\mu^\kappa)^\kappa = \mu^\kappa$, we have $|X| = |X^I/\mathscr{U}|$ and we can pick a bijection $\Psi_0 : X \to X^I/\mathscr{U}$ with the property that $\Psi_0(x) = [c_x]_\mathscr{U}$ for every $x \in X_0$. At the inductive step, let $f : I \to V_{n+1}(X)$ be given. If $f(i) \in V_n(X)$ $\mathscr{U}$-a.e., let $\Psi_{n+1}([f]_\mathscr{U}) = \Psi_n([f]_\mathscr{U})$; and if $f(i) \notin V_n(X)$ $\mathscr{U}$-a.e., that is, if $f(i) \in \mathscr{P}(V_n(X))$ $\mathscr{U}$-a.e., define

$$\Psi_{n+1}([f]_\mathscr{U}) \;=\; \{\Psi_n([g]_\mathscr{U}) \mid g(i) \in f(i)\ \mathscr{U}\text{-a.e.}\}\,.$$

By gluing together the above functions Ψ_n, we obtain a map $\Psi : \mathbb{V}(X)^I_b/\mathscr{U} \to \mathbb{V}(X)$ from the bounded ultrapower of our superstructure into the superstructure itself. Finally, define the star map $* : \mathbb{V}(X) \to \mathbb{V}(X)$ as the composition $\Psi \circ d$, where d is the diagonal embedding:



By the definition of Ψ_0, for every $x \in X_0$ we have that ${}^*x = \Psi(d(x)) = \Psi_0([c_x]_\mathscr{U}) = x$. Moreover, the map $*$ satisfies the *transfer principle* for bounded quantifier formulas, as one can show by using the same arguments as in [8, Theorem 4.4.5]. In brief, the diagonal embedding d preserves the bounded quantifier formulas by Los' Theorem; moreover it is easily verified from the definition that also Ψ preserves the bounded quantifier formulas. Finally, the range of Ψ is a transitive subset of $\mathbb{V}(X)$, and bounded quantifier formulas are preserved under transitive submodels.

In the literature of nonstandard analysis, one usually finds triples $\langle *, \mathbb{V}(X), \mathbb{V}(Y)\rangle$ where $\mathbb{V}(X)$ is called the *standard universe*, $\mathbb{V}(Y)$ the *nonstandard universe*, and the map $* : \mathbb{V}(X) \to \mathbb{V}(Y)$ satisfies the *transfer principle* and ${}^*X = Y$. Such triples are named *superstructure models of nonstandard analysis* (see, *e.g.*, [8, §4.4]). Typically, one takes (a copy of) the real numbers $\mathbb{R}$ as X, and considers triples $\langle *, \mathbb{V}(\mathbb{R}), \mathbb{V}({}^*\mathbb{R})\rangle$. Let us remark that in the above theorem we took $Y = X$, and considered a single universe $\mathbb{V}(X)$.

Remark A.11. The so-called *nonstandard set theories* study suitable adjustements of the usual axiomatic set theory where also the methods of nonstandard analysis are incorporated in their full generality. The most common approach in nonstandard set theories is the so-called *internal viewpoint* as initially proposed independently by E. Nelson XXX and K. Hrbacek XXX, where one includes in the language a unary relation symbol `st` for "standard object". The underlying universe is then given by the internal sets, and the standard objects are those internal elements that are hyper-extensions. As a consequence, external sets do not belong to the universe, and can only be considered indirectly, similarly as proper classes are treated in ZFC as extensions of formulas.

An alternative *external viewpoint*, closer to the superstructure approach, is to postulate a suitably modified version of a Zermelo-Fraenkel theory ZFC, plus the properties of an *elementary embedding* for a star map $* : \mathbb{S} \to \mathbb{I}$ from the sub-universe $\mathbb{S}$ of "standard" objects into the sub-universe $\mathbb{I}$ of "internal" objects. Of course, to this end one needs to include in the language a new function symbol $*$ for the star map.

We remark that if one replaces the *regularity axiom* by a suitable *anti-foundation principle*, then one can actually construct *bounded elementary embeddings* $* : \mathbb{V} \to \mathbb{V}$ defined on the whole universe into itself, thus providing a foundational framework for iterated hyper-extensions that generalizes the superstructure models that we have seen in this section. (See [])

References

1. Sergio Albeverio, Raphael Høegh-Krohn, Jens Erik Fenstad, and Tom Lindstrøm. *Nonstandard methods in stochastic analysis and mathematical physics*, volume 122 of *Pure and Applied Mathematics*. Academic Press, Inc., Orlando, FL, 1986.
2. Leif O. Arkeryd, Nigel J. Cutland, and C. Ward Henson, editors. *Nonstandard analysis*, volume 493 of *NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences*. Kluwer Academic Publishers Group, Dordrecht, 1997. Theory and applications.
3. Mathias Beiglböck. An ultrafilter approach to Jin's theorem. *Israel Journal of Mathematics*, 185(1):369–374, October 2011.
4. Mathias Beiglböck, Vitaly Bergelson, and Alexander Fish. Sumset phenomenon in countable amenable groups. *Advances in Mathematics*, 223(2):416–432, January 2010.
5. Vieri Benci and Mauro Di Nasso. A purely algebraic characterization of the hyperreal numbers. *Proceedings of the American Mathematical Society*, 133(9):2501–2505.
6. Vitaly Bergelson and Neil Hindman. Nonmetrizable topological dynamics and Ramsey theory. *Transactions of the American Mathematical Society*, 320(1):293–320, 1990.
7. Emmanuel Breuillard, Ben Green, and Terence Tao. The structure of approximate groups. *Publications mathématiques de l'IHÉS*, 116(1):115–221, 2012.
8. Chen C. Chang and H. Jerome Keisler. *Model theory*. North-Holland Publishing Co., Amsterdam-New York-Oxford, second edition, 1977. Studies in Logic and the Foundations of Mathematics, 73.
9. Peter Cholak, Gregory Igusa, Ludovic Patey, Mariya Soskova, and Dan Turetsky. The Rado Path Decomposition Theorem. *arXiv:1610.03364 [math]*, October 2016. arXiv: 1610.03364.
10. Péter Csikvári, Katalin Gyarmati, and András Sárközy. Density and Ramsey type results on algebraic equations with restricted solution sets. *Combinatorica*, 32(4):425–449.
11. Mauro Di Nasso, Isaac Goldbring, Renling Jin, Steven Leth, Martino Lupini, and Karl Mahlburg. High density piecewise syndeticity of product sets in amenable groups. *The Journal of Symbolic Logic*, 81(4):1555–1562.
12. Mauro Di Nasso, Isaac Goldbring, Renling Jin, Steven Leth, Martino Lupini, and Karl Mahlburg. High density piecewise syndeticity of sumsets. *Advances in Mathematics*, 278:1–33, June 2015.
13. Mauro Di Nasso, Isaac Goldbring, Renling Jin, Steven Leth, Martino Lupini, and Karl Mahlburg. On a sumset conjecture of Erdös. *Canadian Journal of Mathematics. Journal Canadien de Mathmatiques*, 67(4):795–809, 2015.
14. Mauro Di Nasso and Maria Riggio. Fermat-like equations that are not partition regular. *Combinatorica*. In press.
15. Paul Erdős. A survey of problems in combinatorial number theory. *Annals of Discrete Mathematics*, 6:89–115.
16. Paul Erdős and Ronald L. Graham. *Old and new problems and results in combinatorial number theory*, volume 28 of *Monographies de L'Enseignement Mathématique*. Universit de Genève, L'Enseignement Mathématique, Geneva, 1980.
17. Paul Erdős and Paul Turán. On some sequences of integers. *The Journal of the London Mathematical Society*, S1-11(4):261.
18. Gregory A. Freiman. *Foundations of a structural theory of set addition*. American Mathematical Society, Providence, R. I.
19. Robert Goldblatt. *Lectures on the hyperreals*, volume 188 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998.
20. Isaac Goldbring. Hilbert's fifth problem for local groups. *Annals of Mathematics. Second Series*, 172(2):1269–1314.
21. W. Timothy Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Annals of Mathematics. Second Series*, 166(3):897–946.
22. Ronald L. Graham and Bruce L. Rothschild. A short proof of van der Waerden's theorem on arithmetic progressions. *Proceedings of the American Mathematical Society*, 42:385–386, 1974.
23. Ben Green and Imre Z. Ruzsa. Freiman's theorem in an arbitrary abelian group. *Journal of the London Mathematical Society. Second Series*, 75(1):163–175.
24. Heini Halberstam and Klaus Friedrich Roth. *Sequences*. Springer-Verlag, New York-Berlin, second edition, 1983.
25. Karel Hrbacek and Thomas Jech. *Introduction to set theory*, volume 220 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker, Inc., New York, third edition.
26. Ehud Hrushovski. Stable group theory and approximate subgroups. *Journal of the American Mathematical Society*, 25(1):189–243, 2012.
27. Renling Jin. Applications of nonstandard analysis in additive number theory. *The Bulletin of Symbolic Logic*, 6(3):331–341.
28. Renling Jin. Nonstandard methods for upper Banach density problems. *Journal of Number Theory*, 91(1):20–38, 2001.
29. Renling Jin. The sumset phenomenon. *Proceedings of the American Mathematical Society*, 130(3):855–861, 2002.
30. Renling Jin. Introduction of nonstandard methods for number theorists. *Integers. Electronic Journal of Combinatorial Number Theory*, 8(2):A7, 30, 2008.
31. Teturo Kamae. A simple proof of the ergodic theorem using nonstandard analysis. *Israel Journal of Mathematics*, 42(4):284–290, 1982.
32. H. Jerome Keisler. An infinitesimal approach to stochastic analysis. *Memoirs of the American Mathematical Society*, 48(297):x+184.
33. H. Jerome Keisler and Steven Leth. Meager sets on the hyperfinite time line. *The Journal of Symbolic Logic*, 56(1):71–102.
34. Steven Leth. Near arithmetic progressions in sparse sets. *Proceedings of the American Mathematical Society*, 134(6):1579–1589.
35. Steven Leth. Nonstandard methods and the erdös-turán conjecture. In *The strength of nonstandard analysis*, pages 133–142. Springer.
36. Steven Leth. Some nonstandard methods in combinatorial number theory. *Polish Academy of Sciences. Institute of Philosophy and Sociology. Studia Logica. An International Journal for Symbolic Logic*, 47(3):265–278, 1988.
37. Tom Lindstrøm. An invitation to nonstandard analysis. In *Nonstandard analysis and its applications (Hull, 1986)*, volume 10 of *London Math. Soc. Stud. Texts*, pages 1–105. Cambridge Univ. Press, Cambridge, 1988.
38. Amir Maleki. Solving equations in $\beta\mathbb{N}$. *Semigroup Forum*, 61(3):373–384, 2000.
39. Melvyn B. Nathanson. Sumsets contained in infinite sets of integers. *Journal of Combinatorial Theory, Series A*, 28(2):150–155, March 1980.
40. Hans Jürgen Prömel. *Ramsey theory for discrete structures*. Springer, Cham, 2013.

41. Cameron L. Stewart and Robert Tijdeman. On infinite-difference sets. *Canadian Journal of Mathematics. Journal Canadien de Mathmatiques*, 31(5):897–910, 1979.
42. Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, 27:199–245, 1975.
43. Terence Tao. *Hilbert's fifth problem and related topics*, volume 153 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI.
44. Terence Tao. *An introduction to measure theory*, volume 126 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2011.
45. Stevo Todorcevic. *Introduction to Ramsey spaces*, volume 174 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2010.
46. Lou van den Dries. Approximate groups. *Astérisque*, (367):Exp. No. 1077, vii, 79–113.
47. Bartel Leendert van der Waerden. Beweis einer Baudetschen Vermutung. *Nieuw Archief voor Wiskunde*, 15:212–216.
48. Hidehiko Yamabe. A generalization of a theorem of Gleason. *Annals of Mathematics. Second Series*, 58:351–365.

Index