Lecture Notes for 2023 Fudan Logic Summer School

Nonstandard Analysis and Combinatorial Number Theory

Renling Jin
Department of Mathematics
College of Charleston
Charleston, SC 29424

Abstract

This one-week course is for the students with some background in mathematical logic in a typical one-semester undergraduate level course on mathematical logic. In the first two days, we will cover basic ideas, concepts, properties, principles, etc. in nonstandard analysis with some applications in calculus and finance. In the last two days, we will focus on applications of nonstandard methods to the problems in combinatorial number theory. In the third day, we study some density problems by working in a simple nonstandard universe. In the last day, we study van der Waerden Theorem and Szemerédi Theorem related problems by working in a nonstandard universe with multiple levels of infinities.

Contents

1 Foundation of Nonstandard Analysis

1.1 Introduction

Can we incorporate new real numbers such as non-zero infinitesimals which is non-zero but closer to 0 than all old non-zero real numbers into our existing number system?

The geometric understanding of real numbers can be achieved by identifying each of them with the position of a point on a line relative to a pre-fixed location called origin on the line with a pre-determined unit length. From the origin and unit one can naturally generates the set of all integers. It is easy to imagine the addition–subtraction and multiplication–division in terms of the physical reality. Hence, the admission of the set of all rational numbers into our number system should not be controversial. Can we say the same for admitting more numbers beyond rational numbers?

There have been historical and psychological controversies when mathematicians tried to admit real numbers beyond rational numbers into existing number system. In ancient Greece, it costed, according to a legend, the life of Hippasus after he discovered the secret that $\sqrt{2}$ is not a rational number. Pre and during Hilbert's time, it was natural to assume that the real line should be complete, i.e., every bounded nonempty set of reals has a least upper bound. One of the consequences of the completeness property is, discovered by Georg Cantor, that the set of all real numbers is uncountable. But since there are only countably many ways to describe or identify individual real numbers, there should be a lot of real numbers which can never be identified or described. But why should we assume the existence of these real numbers? Can we do mathematics without them?

The completeness property adds to the structural beauty of the real number system which allows mathematicians to develop real analysis and prove theorems in an elegant and simple way. It might be fine to develop real analysis without the completeness property. But that could be much more cumbersome and tedious since the set of all describable reals has vague boundary.

The completeness property of the real line is an example of mathematical concept as the product of human's imagination to stretch a pattern of the reality to create an idealize structure, which in turn offers a better and more efficient tool for mathematicians to study real physical problems.

There are other examples of this nature. For another example, one can define an angle between two vectors in four or higher dimensional Euclidean space over the real field by following the parallel pattern of angles in the three dimensional space using inner product and Schwarz Inequality. Note that there seems no way to measure an angle physically between two vectors in a four or higher dimensional space. However, the angle between two vectors in a four or higher dimensional space is the key concept used to establish the connection between the value of a correlation coefficient of a paired data and the linear associativity of the data in statistics.

During the time when calculus was developed in seventeen century, both I. Newton and G. Leibniz used infinitesimals to create differentiation and integration theory. Due to lack of logical foundation despite its effectiveness, the admission of infinitesimals into existing number system for calculus became the target of criticism for potential inconsistency in the early eighteen century. Clearly, incorporating infinitesimals into existing real number system, if possible, demonstrates another example of human's imagination contributing to the mathematical reality and makes the enlarged real number system more useful. The problem is how the current real number system can be enlarged consistently so that the enlarged system satisfies still many useful properties such as the axioms of ordered field, and contains infinitesimal elements.

Due to the limitation of human's intuition, it is hard to imagine a positive but infinitesimal distance between two points. However, we can consider the real number system from an algebraic point of view. Imagine that the real numbers are pebbles (or calculi, which is how the branch of mathematics "Calculus" got its name). All relations and functions on the real field describe merely how these pebbles are related to each other. Now we just want to add some new pebbles to the collection and generalize the old relations and functions to the new collection so that the new collection remains to be an ordered field and some new pebbles act like infinitesimals. Can we do this consistently and if we can, why mathematicians in nineteen century didn't already do this?

There has been a huge progress in mathematical logic since the first half of twentieth century. Based on the compactness theorem in mathematical logic, A. Robinson in the early 1960s (cf. [23]) proved the consistency of admitting infinitesimals into the existing real field $\mathbb{R}$ so that the expanded system $^*\mathbb{R}$ is an ordered field and satisfies all first-order sentences which are true in $\mathbb{R}$. What does the word "first-order" mean?

1.2 First-order Logic and Ultrapower of Real Field

We will limit the breadth of our non-traditional introduction of the first-order logic. We will touch only the part enough for the purpose of this course.

The language of the first-order logic contains the logical symbols and non-logical symbols. The logical symbols include logical connectives: $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$, quantifiers: $\forall$, $\exists$, equality symbol: $=$, and variables: x, y, z, ... Logical symbols are used in the study of all branches of mathematics. A set of non-logical symbols, denoted by $\mathscr{L}$, is for some specific branch of mathematics. For example, the set of symbols

$$\mathscr{L} = \{+, \cdot, \leq, 0, 1, P\}_{P \in \mathcal{P}}$$

is the language of ordered field, where $+$ and $\cdot$ are three dimensional relation symbols, $\leq$ is a two dimension relation symbol, 0 and 1 are constant symbols which can also be considered as zero-dimensional relation symbols, and $\mathcal{P}$ is a collection of other relation symbols of finite arity. Although $\mathcal{P}$ can be assumed to contains $+$, $\cdot$, $\leq$, $0, 1$, we list these arithmetic operation symbols explicitly for clarity. By a language we mean the set of non-logical symbols. For notational simplicity all non-logical symbols considered are relational symbols (note that an n-variable function can be identified with the graph of the function which is an $(n+1)$-dimensional relation and a constant symbol is a 0-dimensional relation symbol). We do not distinguish each symbol from its intended interpretation. For example, $+$ represents a three dimensional relation symbol in the language of ordered field as well as the actual addition in an ordered field.

Definition 1.1 *An $\mathscr{L}$-formula can be formed inductively by the following two steps.*

1. *Basic step: The atomic formulas are those in the form of $P(\overline{x}, \overline{c})$ where $P \in \mathscr{L} \cup \{=\}$ is a relation symbol with arity m and $\overline{x}$ represents the k-tuple $(x_1, x_2, \ldots, x_k)$ of variables and $\overline{c}$ represents the $m-k$ tuple of constant symbols;*

2. *Inductive step: If φ and ψ are $\mathscr{L}$-formulas, so are $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \rightarrow \psi$, $\varphi \leftrightarrow \psi$, $\forall x\, \varphi$, and $\exists x\, \varphi$.*

By the complexity of a formula, we mean the number of steps in Definition 1.1 used to form the formula. The sub-formula φ is called the scope of the quantifier $\forall$ or $\exists$ in the formula $\forall x\,\varphi$ or $\exists x\,\varphi$, respectively. The variable x in $\forall x\,\varphi$ or $\exists x\,\varphi$ is called bounded. An occurrence of a variable x is called bounded in an formula φ if it is bounded in a sub-formula $\forall x\,\psi$ or $\exists x\,\psi$ of φ. An occurrence of a variable x is called free in φ if it is not bounded. We write $\overline{x}$ for a tuple of variables, and write $\varphi(\overline{x})$ to indicate implicitly that all free variables in φ are among the variables in $\overline{x}$.

Definition 1.2 *An $\mathscr{L}$-model $\mathcal{M} := (M; P^{\mathcal{M}})_{P\in\mathscr{L}}$ contains a non-empty base set M together with the interpretation $P^{\mathcal{M}} \subseteq M^m$ of each relation symbol $P \in \mathscr{L}$ with arity m.*

Note that if P is a constant symbol, sometimes denoted by c, then $P^{\mathcal{M}} = c^{\mathcal{M}}$ is an element in M. We sometimes write $\mathcal{M}$ for a model as well as its base set. Since the languages considered in the notes are either the language of ordered field or the language of set theory, we will omit $\mathscr{L}$ when we mention $\mathscr{L}$-formula or $\mathscr{L}$-model unless otherwise specific. The word "iff" is an abbreviation of "if and only if."

Definition 1.3 *For each formula $\varphi(\overline{x})$ and a tuple $\overline{a}$ of elements in a model $\mathcal{M}$, we define $\mathcal{M} \models \varphi(\overline{a})$, i.e., $\varphi(\overline{a})$ is true in $\mathcal{M}$, inductively on the complexity of the formula.*

1. *φ is an atomic formula $P(\overline{x}, \overline{c})$: $\mathcal{M} \models P(\overline{a}, \overline{c})$ iff $(\overline{a}, \overline{c^{\mathcal{M}}}) \in P^{\mathcal{M}}$;*
2. *$\mathcal{M} \models \neg\varphi$ iff $\mathcal{M} \not\models \varphi$, i.e., it's not true that $\mathcal{M} \models \varphi$ (so "$\neg$" means "not");*
3. *$\mathcal{M} \models \varphi \wedge \psi$ iff $\mathcal{M} \models \varphi$ and $\mathcal{M} \models \psi$ (so "$\wedge$" means "and");*
4. *$\mathcal{M} \models \varphi \vee \psi$ iff $\mathcal{M} \models \varphi$ or $\mathcal{M} \models \psi$ (so "$\vee$" means "or");*
5. *$\mathcal{M} \models \varphi \to \psi$ iff $\mathcal{M} \models \varphi$ implies $\mathcal{M} \models \psi$ (so "$\to$" means "imply");*
6. *$\mathcal{M} \models \varphi \leftrightarrow \psi$ iff $\mathcal{M} \models (\varphi \to \psi) \wedge (\psi \to \varphi)$ (so "$\leftrightarrow$" means "equivalent to");*
7. *$\mathcal{M} \models \forall x\,\varphi(x, \overline{a})$ iff $\mathcal{M} \models \varphi(b, \overline{a})$ for every $b \in \mathcal{M}$ (so "$\forall$" means "for every");*
8. *$\mathcal{M} \models \exists x\,\varphi(x, \overline{a})$ iff $\mathcal{M} \models \varphi(b, \overline{a})$ for some $b \in \mathcal{M}$ (so "$\exists$" means "for some").*

Note that by (1), (7), and (8) the intended value in $\mathcal{M}$ for a variable x is always an element of $\mathcal{M}$. This is the reason why we call the logic system above the first-order logic. When we say a formula, we mean a first-order formula unless otherwise specified. A formula without free variable is called a sentence. If a model $\mathcal{M}$ is given and every free variable of a formula $\varphi(\overline{x})$ is substituted by an element in $\overline{a}$ in $\mathcal{M}$, we call also $\varphi(\overline{a})$ a sentence or a sentence with parameters $\overline{a}$. So, the truth value of a sentence in a model is always determined.

It is an easy fact that each formula φ is logically equivalent to a formula ψ, i.e., φ and ψ have the same truth value in any model, where ψ does not use any of the symbols $\vee$, $\rightarrow$, $\leftrightarrow$, or $\forall$. Hence, it suffices to consider only the formulas using logic connectives $\neg$, $\wedge$, and quantifier $\exists$ in some of the proofs later on.

Example 1.4 *Let $\mathscr{L} = \{+, \cdot, \leq, 0, 1, P\}_{P \in \mathcal{P}}$ be the language of ordered field and $\mathcal{R} := (\mathbb{R}; +, \cdot, \leq, 0, 1, P^{\mathcal{R}})_{P \in \mathcal{P}}$ be the usual real ordered field with some extra relations. Then $\mathcal{R}$ is an $\mathscr{L}$-model. If φ is the sentence*

$$\forall x, y, z\, (x \leq y \rightarrow x + z \leq y + z),$$

then $\mathcal{R} \models \varphi$.

Note that the sentence above can formally be written as a logic sentence

$$\forall x \forall y \forall z \forall u \forall v\, (\leq (x, y) \wedge +(x, z, u) \wedge +(y, z, v) \rightarrow \leq (u, v)).$$

We will use conventional expressions more often than the formal ones. The reader is guaranteed that all conventional expressions can be re-written as formal ones.

The following is a familiar sentence in the language of real ordered field which is not first-order because variable X in the sentence takes a set not an element of $\mathbb{R}$ as its value.

Example 1.5 *Let φ be the sentence*

$$\forall X \subseteq [0, 1]\, \exists \beta\, (\beta \text{ is the least upper bound of } X).$$

then φ is true in $\mathcal{R}$.

We now construct an ultrapower of $\mathcal{R}$. Let $\mathbb{N} := \{0, 1, 2, \ldots\}$ and $\mathbb{Z} := \{0, \pm 1, \pm 2, \ldots\}$. If n is a positive integer, let $[n] := \{0, 1, \ldots, n - 1\}$.

Definition 1.6 *Let X be an infinite set and $\mathscr{P}$ be the power set operator. A collection $\mathcal{F} \subseteq \mathscr{P}(X)$ is called a non-principal ultrafilter on X if for any $A, B \subseteq X$*

1. *$\emptyset$ is not in $\mathcal{F}$ and every co-finite subset A of X (i.e., $X \setminus A$ is finite) is in $\mathcal{F}$;*
2. *if A, B are in $\mathcal{F}$, then $A \cap B$ is in $\mathcal{F}$;*

3. *if A is in $\mathcal{F}$ and $A \subseteq B$, then B is in $\mathcal{F}$;*

4. *if A is not in $\mathcal{F}$, then $X \setminus A$ is in $\mathcal{F}$.*

The existence of a non-principal ultrafilter on an infinite set X is guaranteed by the axiom of choice. For simplicity we use only a fixed non-principal ultrafilter $\mathcal{F}$ on $X := \mathbb{N}$. In fact, any non-principal ultrafilter on an infinite set X works as long as it is countably incomplete ($\mathcal{F}$ on a countable set such as $\mathbb{N}$ is trivially countably incomplete.)

Definition 1.7 *Let $\mathcal{M}$ be a model. Let $\mathcal{M}^{\mathbb{N}}$ be the set of all functions from $\mathbb{N}$ to $\mathcal{M}$. For any $f, g \in \mathcal{M}^{\mathbb{N}}$ define $f \sim g$ if $\{n \in \mathbb{N} \mid f(n) = g(n)\} \in \mathcal{F}$. The equivalence class of $f \in \mathcal{M}^{\mathbb{N}}$ is the set $[f] := \{g \in \mathcal{M}^{\mathbb{N}} \mid f \sim g\}$. Set $\mathcal{M}^{\mathbb{N}}/\mathcal{F} := \{[f] \mid f \in \mathcal{M}^{\mathbb{N}}\}$. The ultrapower of $\mathcal{M}$ modulo $\mathcal{F}$, denoted by $\mathcal{M}^{\mathbb{N}}/\mathcal{F}$, is a model with the base set $\mathcal{M}^{\mathbb{N}}/\mathcal{F}$ and for each relation symbol P, the interpretation of P in $\mathcal{M}^{\mathbb{N}}/\mathcal{F}$ is defined by*

$$\overline{[f]} \in P^{\mathcal{M}^{\mathbb{N}}/\mathcal{F}} \text{ iff } \{n \in \mathbb{N} \mid \overline{f(n)} \in P^{\mathcal{M}}\} \in \mathcal{F}. \tag{1}$$

For each $a \in \mathcal{M}$ let $\phi_a : \mathbb{N} \to \mathcal{M}$ be the constant function with a unique value a. If $\overline{b} = (b_1, b_2, \ldots, b_k)$, we write $\overline{[\phi_b]}$ for $([\phi_{b_1}], [\phi_{b_2}], \ldots, [\phi_{b_k}])$.

Definition 1.8 *Let $i : \mathcal{M} \to \mathcal{M}^{\mathbb{N}}/\mathcal{F}$ be the function such that $i(a) = [\phi_a]$. The function i is called an elementary embedding associated with the ultrapower construction.*

Theorem 1.9 (J. Łoś) *Let $\mathcal{M}^{\mathbb{N}}/\mathcal{F}$ be the ultrapower of a model $\mathcal{M}$ modulo $\mathcal{F}$. Let $\varphi(\overline{x}, \overline{b})$ be a formula with parameters $\overline{b}$ in $\mathcal{M}$. Then*

$$\mathcal{M}^{\mathbb{N}}/\mathcal{F} \models \varphi(\overline{[f]}, \overline{[\phi_b]}) \text{ iff } \{n \in \mathbb{N} \mid \mathcal{M} \models \varphi(\overline{f(n)}, \overline{b})\} \in \mathcal{F}.$$

Proof: If φ is an atomic formula $P(\overline{x}, \overline{c})$, then the theorem follows from the definition of $P^{\mathcal{M}^{\mathbb{N}}/\mathcal{F}}$. If φ is $\neg\psi$, then the theorem follows from Part 4 of Definition 1.6 and induction hypothesis for ψ. If φ is $\psi \wedge \chi$, then the theorem follows from Part 2 of Definition 1.6 and induction hypothesis for ψ and χ.

Assume φ is $\exists y\, \psi(x, \overline{y}, \overline{b})$. If $A := \{n \in \mathbb{N} \mid \mathcal{M} \models \exists x\, \psi(x, \overline{f(n)}, \overline{b})\} \in \mathcal{F}$, define a function $g : \mathbb{N} \to \mathcal{M}$ by letting $g(n)$ be any fixed element in $\mathcal{M}$ if $n \notin A$, and $g(n) = a_n$ for some $a_n \in \mathcal{M}$ with $\mathcal{M} \models \psi(a_n, \overline{f(n)}, \overline{b})$ if $n \in A$. Then

$$A \subseteq \{n \in \mathbb{N} \mid \mathcal{M} \models \psi(g(n), \overline{f(n)}, \overline{b})\}.$$

By Part 3 of Definition 1.6 and the induction hypothesis on ψ we have $\mathcal{M}^{\mathbb{N}}/\mathcal{F} \models \psi([g], \overline{[f]}, \overline{[\phi_b]})$, which implies $\mathcal{M}^{\mathbb{N}}/\mathcal{F} \models \exists x\, \psi(x, \overline{[f]}, \overline{[\phi_b]})$.

On the other hand, if $\mathcal{M}^{\mathbb{N}}/\mathcal{F} \models \exists x\, \psi(x, \overline{[f]}, \overline{[\phi_b]})$, then there is a $g : \mathbb{N} \to \mathcal{M}$ such that $\mathcal{M}^{\mathbb{N}}/\mathcal{F} \models \psi([g], \overline{[f]}, \overline{[\phi_b]})$. By the induction hypothesis for ψ we have

$$B := \{n \in \mathbb{N} \mid \mathcal{M} \models \psi(g(n), \overline{f(n)}, \bar{b})\} \in \mathcal{F}.$$

So, if $n \in B$, we have $\mathcal{M} \models \exists x\, \psi(x, \overline{f(n)}, \bar{b})$. Hence,

$$B \subseteq \{n \in \mathbb{N} \mid \mathcal{M} \models \exists x\, \psi(x, \overline{[f]}, \overline{[\phi_b]})\} \in \mathcal{F}$$

by Part 2 of Definition 1.6. □

Corollary 1.10 *Let $i : \mathcal{M} \to \mathcal{M}^{\mathbb{N}}/\mathcal{F}$ be the embedding defined in Definition 1.8. For any sentence $\varphi(\bar{b})$ with parameters $\bar{b}$ in $\mathcal{M}$, we have*

$$\mathcal{M} \models \varphi(\bar{b}) \text{ iff } \mathcal{M}^{\mathbb{N}}/\mathcal{F} \models \varphi(\overline{i(b)}). \tag{2}$$

Proof: The corollary follows from Theorem 1.9, Part 1 of Definition 1.6, and the fact that the set $\{n \in \mathbb{N} \mid \mathcal{M} \models \varphi(\overline{\phi_b(n)})\}$ is either $\mathbb{N}$ or $\emptyset$ depending on whether $\mathcal{M} \models \varphi(\bar{b})$ is true or not. □

The map i satisfying (2) is called an elementary embedding from a model $\mathcal{M}$ to another model $\mathcal{M}' = \mathcal{M}^{\mathbb{N}}/\mathcal{F}$. In fact, the elementary embedding from $\mathcal{M}$ to $\mathcal{M}'$ can be defined for any two models $\mathcal{M}$ and $\mathcal{M}'$ when (2) is true. Denote by $\mathcal{M} \preceq \mathcal{M}'$ for the existence of such an elementary embedding from $\mathcal{M}$ to $\mathcal{M}'$. If we want to emphasize that an elementary embedding $i : \mathcal{M} \to \mathcal{M}'$ is not surjective, we can just write $\mathcal{M} \prec \mathcal{M}'$ instead. The statement (2) is also called the transfer principle between $\mathcal{M}$ and $\mathcal{M}' = \mathcal{M}^{\mathbb{N}}/\mathcal{F}$. The embedding i is often written as * in nonstandard analysis. For example, a nonstandard analyst may write $^*A$ more often than $i(A)$.

Example 1.11 *Let $^*\mathcal{R}$ be the ultrapower of the "standard" real ordered field $\mathcal{R}$ modulo $\mathcal{F}$.*

1. *$^*\mathcal{R}$ satisfies the same first-order sentences with parameters from $\mathcal{R}$, in particular, $^*\mathcal{R}$ is an ordered field and contains a copy of $\mathcal{R}$ as its (elementary) sub-model. We call real numbers in $\mathbb{R}$ the standard real numbers.*

2. *By identifying each $\alpha \in \mathbb{R}$ with $^*\alpha = [\phi_\alpha] \in {}^*\mathbb{R}$, we can assume that $\mathcal{R} \subseteq {}^*\mathcal{R}$.*

3. *A real $r \in {}^*\mathbb{R}$ is called an infinitesimal, denoted by $r \approx 0$, if $|r| < |\alpha|$ for every non-zero $\alpha \in \mathbb{R}$. Two reals $r_1, r_2 \in {}^*\mathbb{R}$ are said to be infinitesimally close, denoted by $r_1 \approx r_2$, if $r_1 - r_2$ is an infinitesimal.*

4. *If Id* $\in \mathcal{R}^{\mathbb{N}}$ *is the identity function, i.e., Id*$(n) = n$ *for every* $n \in \mathbb{N}$*, then* $[Id] \in {}^*\mathcal{R}$ *and* $[Id] > r$ *for every* $r \in \mathbb{R}$*. So,* ${}^*\mathcal{R}$ *contains numbers larger than every* $r \in \mathbb{R}$*.*

5. $1/[Id]$ *in* ${}^*\mathcal{R}$ *is a positive infinitesimal;*

6. *A number* $N \in {}^*\mathbb{N} \setminus \mathbb{N}$ *is called a hyperfinite integer. For example,* $[Id]$ *is a hyperfinite integer. A hyperfinite integer is infinitely large from the standard point of view, but is finite from nonstandard point of view.*

Proposition 1.12 *A real number* $r \in {}^*\mathbb{R}$ *is called near standard if* $|r| \leq \alpha$ *for some* $\alpha \in \mathbb{R}$*. If* r *is near standard, then there exists a unique* $\beta \in \mathbb{R}$ *such that* $r \approx \beta$*.*

Proof: Let $S = \{\gamma \in \mathbb{R} \mid \gamma < r\}$. Then the set $S \subseteq \mathbb{R}$ is bounded above by α. By the completeness property S has a least upper bound β. It is easy to check that $r \approx \beta$. The uniqueness follows from the fact that two distinct standard reals can never be infinitesimally close. □

Example 1.13 *Let* S *be the set of all infinitesimals in* ${}^*\mathcal{R}$*. Then* S *is nonempty and bounded above by* 1*. Note that* S *does not have a least upper bound. Indeed, if* $\beta > 0$ *were the least upper bound of* S*, then* β *being infinitesimal would imply* 2β *being also an infinitesimal which violates* β *being upper bound of* S*, and* β *being non-infinitesimal would imply* $\beta/2$ *being also a non-infinitesimal which violates* β *being the least. Either way we have a contradiction.*

The example above shows that $\mathcal{R}$ and ${}^*\mathcal{R}$ may not share the same truth beyond the first-order. Besides the transfer principle, the standard part map is another way to connect ${}^*\mathcal{R}$ to $\mathcal{R}$.

Definition 1.14 *Let* $ns({}^*\mathcal{R})$ *be the set of all near standard reals in* ${}^*\mathcal{R}$*. We define the standard part map* $st : {}^*\mathcal{R} \to \mathbb{R} \cup \{\pm\}$ *by letting* $st(r) = \alpha$ *for every* $r \in ns({}^*\mathcal{R})$ *where* α *is the unique number in* $\mathbb{R}$ *such that* $r \approx \alpha$*,* $st(r) = \infty$ *if* $r > \alpha$ *for every* $\alpha \in \mathbb{R}$*, and* $st(r) = -\infty$ *if* $r < \alpha$ *for every* $\alpha \in \mathbb{R}$*.*

We would like to present very simple applications of nonstandard analysis to calculus. Note that the arguments in these applications avoid the use of limit process.

Definition 1.15 *Let* $s : \mathbb{N} \to \mathbb{R}$ *be a standard sequence. The sequence* s *is convergent if* ${}^*s(N) \approx {}^*s(N')$ *for any hyperfinite integers* N, N'*.*

Theorem 1.16 (Bolzano–Weierstrass) *Every standard bounded sequence contains a convergent subsequence.*

Proof: Suppose s is the bounded sequence in $[a,b]$. Let $N \in {}^*\mathbb{N} \setminus \mathbb{N}$. Then ${}^*s(N) \in ns({}^*\mathbb{R})$. Let $L = st({}^*s(N))$. We show that there exists a subsequence s' of s such that s' converges to L.

For each $n \in \mathbb{N}$, if there is an $m_n \in \mathbb{N}$ such that $s(m) \notin (L - 1/n, L + 1/n)$ for all $m \geq m_n$ in $\mathbb{N}$, then ${}^*s(m) \notin {}^*(L - 1/n, L + 1/n)$ for any $m \geq m_n$ in ${}^*\mathbb{N}$ by the transfer principle, which contradicts $st({}^*s(N)) = L$ because $N \geq m_n$. Hence, $(L - 1/n, L + 1/n)$ contains infinitely many terms of s for each $n \in \mathbb{N}$. So, one can choose $m_1 < m_2 < \cdots$ such that $s(m_n) \in (L - 1/n, L + 1/n)$. Now for any hyperfinite integers $N < N'$ we have ${}^*s(m_N), {}^*s(m_{N'}) \in {}^*(L - 1/N, L + 1/N)$. Hence, ${}^*s(m_N) \approx {}^*s(m_{N'})$. This shows that the subsequence $s(m_1), s(m_2), \ldots$ is a convergent subsequence of s. □

Definition 1.17 *Let $f : [a,b] \to \mathbb{R}$ be a standard function. Then,*

1. *f is continuous at $c \in [a,b]$ if for any $r \in {}^*[a,b]$ we have $r \approx c$ implies ${}^*f(r) \approx f(c)$; f is continuous on $[a,b]$ if f is continuous at every $c \in [a,b]$;*
2. *f is uniformly continuous on $[a,b]$ if $r_1 \approx r_2$ implies ${}^*f(r_1) \approx {}^*f(r_2)$ for any $r_1, r_2 \in {}^*[a,b]$.*

Theorem 1.18 *If a standard function $f : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$, then f is uniformly continuous on $[a,b]$.*

Proof: Suppose that f is continuous on $[a,b]$ but not uniformly continuous on $[a,b]$. Then, there exist $r_1, r_2 \in {}^*[a,b]$ such that $r_1 \approx r_2$ but ${}^*f(r_1) \not\approx {}^*f(r_2)$. Since $r_1 \approx r_2$ we have $st(r_1) = st(r_2) = c \in [a,b]$. Since $r_1 \approx c \approx r_2$, then ${}^*f(r_1) \approx f(c) \approx {}^*f(r_2)$, which contradicts the assumption that ${}^*f(r_1) \not\approx {}^*f(r_2)$. □

Definition 1.19 *Let $f : (a,b) \to \mathbb{R}$ be a standard function and $c \in (a,b)$. The function f is differentiable at c if there exists an $\alpha \in \mathbb{R}$ such that*

$$f'(c) := st\left(\frac{{}^*f(r) - f(c)}{r - c}\right) = \alpha$$

for any $r \in {}^*\mathbb{R}$ with $r \approx c$ and $r \neq c$.

Given a function $f : X \to Y$, $a \in X$, and $A \subseteq X$, we write $f(a)$ for some element in Y, and write $f[A]$ for the set $\{f(a) \mid a \in A\}$.

Theorem 1.20 (Chain Rule) *If the standard function $f : (a, b) \to \mathbb{R}$ is differentiable at $c \in (a, b)$, $f[(a, b)] \subseteq (\alpha, \beta)$, and a standard function $g : (\alpha, \beta) \to \mathbb{R}$ is differentiable at $f(c)$, then $g(f(x)) : (a, b) \to \mathbb{R}$ is differentiable at c and $(g(f(x)))'_c = g'(f(c))f'(c)$.*

Proof: Given any $r \approx c$ and $r \neq c$. Since f is differentiable at c, there is an infinitesimal t_1 such that that ${}^*f(r) - f(c) = (f'(c) + t_1)(r - c)$. Since g is differentiable at $f(c)$ and ${}^*f(r) \approx f(c)$ there is another infinitesimal t_2 such that ${}^*(g(f(r)) - g(f(c)) = (g'(f(c)) + t_2)({}^*f(r) - f(c))$. Hence, we have

$$\begin{aligned}
{}^*(g(f(r)) - g(f(c))) &= (g'(f(c)) + t_2)({}^*f(r) - f(c)) \\
&= (g'(f(c)) + t_2)(f'(c) + t_1)(r - c) \\
&= (g'(f(c))f'(c) + t_2 f'(c) + t_1 g'(f(c)) + t_2 t_1)(r - c),
\end{aligned}$$

which implies

$$\begin{aligned}
&st\left(\frac{g(f(r)) - g(f(c))}{r - c}\right) \\
&\quad = st\,(g'(f(c))f'(c) + t_2 f'(c) + t_1 g'(f(c)) + t_2 t_1) = g'(f(c))f'(c).
\end{aligned}$$

□

To deal with integration we encounter the integral operator which is a linear functional. Therefore, it cannot be handled in ${}^*\mathcal{R}$. We need a structure not only containing functions from $\mathbb{R}^n$ to $\mathbb{R}$ but also containing the functions of the functions, the functions of the functions of the functions, etc. This is why we introduce another model called superstructure to deal with this and many other needs in the next subsection.

1.3 Ultrapower of Superstructure

The use of superstructure and its elementary extension as the model of nonstandard analysis appeared in [24] for the first time. Fix a sufficiently large positive integer $\mathfrak{n}$, say $\mathfrak{n} = 100$. Let $\mathscr{L} := \{\in\}$ contain only one binary relation symbol. Given an infinite set X of urelements, i.e., elements without members, the *superstructure* on X, denoted by $\mathcal{V}(X)$, is an $\mathscr{L}$-model $(V(X); \in)$ where $V(X)$ is defined inductively by letting

$$V(X, 0) := X,\ V(X, n+1) := V(X, n) \cup \mathscr{P}(V(X, n))$$

for every $n < 2\mathfrak{n}$, $V(X) = V(X; 2\mathfrak{n})$, and letting $\in$ be the true set theoretic membership relation on $V(X)$ as the interpretation of the symbol $\in$ in $\mathscr{L}$. For notational convenience, we don't distinguish $\mathcal{V}(X)$ for the model from the base set of the model. We write also $\mathcal{V}(X,n)$ for both $(V(X,n);\in)$ and $V(X,n)$.

For each element $a \in \mathcal{V}(X)$ we define the *rank* of a, denoted by $\mathrm{rank}(a)$, by that if $a \in X$, then $\mathrm{rank}(a) = 0$ and if $a \in \mathcal{V}(X, n+1) \setminus \mathcal{V}(X,n)$ for some number $n < 2\mathfrak{n}$, then $\mathrm{rank}(a) = n+1$. The rank function on $\mathcal{V}(X)$ is bounded by $2\mathfrak{n}$ and is definable in $\mathcal{V}(X)$ by a first-order formula.

We assume always that $\mathbb{N} \subseteq \mathbb{R} \subseteq X$. For simplicity, set $X = \mathbb{R}$. Note that all standard mathematical objects mentioned in the lecture notes have ranks below $\mathfrak{n} = 100$. We set the highest rank to be $2\mathfrak{n}$ instead of $\mathfrak{n}$ for convenience. Note that an ordered pair (a,b) of real numbers $a,b \in \mathbb{R}$ can be viewed as the set $\{\{a\},\{a,b\}\} \in \mathcal{V}(\mathbb{R},2)$ and a function $f : D \subseteq \mathbb{R} \to \mathbb{R}$ can be viewed as a set of ordered pairs in $\mathcal{V}(\mathbb{R},2)$. Hence, $f \in \mathcal{V}(\mathbb{R},3)$. A linear functional L on functions from $\mathbb{R}$ to $\mathbb{R}$ is a set of pairs $(f,r) = \{\{f\},\{f,r\}\} \in \mathcal{V}(\mathbb{R},5)$. Hence, $L \in \mathcal{V}(\mathbb{R},6)$. Note also that the ultrafilter $\mathcal{F}$ on $\mathbb{N}$ is in $\mathcal{V}(\mathbb{R},3)$.

The superstructure $\mathcal{V}(\mathbb{R})$ is often called the *standard universe*, which means that all discussion of ordinary mathematical problems at the moment can be conducted in $\mathcal{V}(\mathbb{R})$. Since X is always $\mathbb{R}$ in the notes we omit $\mathbb{R}$ and write $\mathcal{V}$ for $\mathcal{V}(\mathbb{R})$. Recall that $\mathcal{F}$ is a non-principal ultrafilter on $\mathbb{N}$.

Definition 1.21 *The ultrapower of $\mathcal{V}$ modulo $\mathcal{F}$, denoted by ${}^*\mathcal{V}$, is the model*

$$({}^*V;\, {}^*\!\in),$$

where the base set ${}^*V = V(X)^{\mathbb{N}}/\mathcal{F}$ and the interpretation ${}^*\!\in$ of the binary relation symbol $\in$ is defined by letting $[f]\,{}^*\!\in [g]$ iff $\{n \in \mathbb{N} \mid f(n) \in g(n)\} \in \mathcal{F}$ for any $[f],[g] \in V(X)^{\mathbb{N}}/\mathcal{F}$. Let ${}^* : \mathcal{V} \to {}^*\mathcal{V}$ be the elementary embedding, i.e., ${}^*a := [\phi_a]$ for every $a \in \mathcal{V}$.

Note that the real ordered field $\mathcal{R}$ is in $\mathcal{V}$. Hence, ${}^*\mathcal{R}$ is in ${}^*\mathcal{V}$.

The model ${}^*\mathcal{V}$ is called a *nonstandard universe*, or a nonstandard elementary extension of the standard universe $\mathcal{V}$.

One of the advantages of using nonstandard methods is to replace a limit argument, which has a higher set theoretic complexity in the standard model, by an infinitesimal argument, which has a lower set theoretic complexity in a nonstandard model. The reader is encouraged to treat all $r \in {}^*\mathbb{R}$ as urelements instead of equivalence classes of functions from $\mathbb{N}$ to $\mathbb{R}$ to take this advantage and treat ${}^*\!\in$ as a real membership relation.

Formally, assume that elements in ${}^*\mathbb{R}$ are urelements with rank 0. Let $\mathscr{M}$ be the Mostowski collapsing map on ${}^*\mathcal{V}$, i.e., $\mathscr{M}(a) = a$ for every $a \in {}^*\mathbb{R}$ and

$$\mathscr{M}(b) := \{\mathscr{M}(a) \mid a\, {}^*\!\in b\}$$

for every $b \in {}^*\mathcal{V} \setminus {}^*\mathbb{R}$. Then $\mathscr{M}$ is an injection and $a \,{}^*\!\!\in b$ iff $\mathscr{M}(a) \in \mathscr{M}(b)$. If one identifies ${}^*\mathcal{V}$ with the image of ${}^*\mathcal{V}$ under $\mathscr{M}$, one can pretend that ${}^*\!\!\in$ is the true membership relation and consider ${}^*\mathcal{V}$ as a subset of the superstructure $\mathcal{V}({}^*\mathbb{R})$. Hence, we can drop the upper-left superscript * from ${}^*\!\!\in$ for notational convenience.

Similar to the elements in $\mathcal{V}$, the rank function can also be defined for elements in ${}^*\mathcal{V}$. Every element in ${}^*\mathbb{R}$ has rank 0. It is easy to check that every element in ${}^*\mathcal{V}(\mathbb{R}, n+1) \setminus {}^*\mathcal{V}(\mathbb{R}, n)$ has the rank $n+1$ for $n \in [2\mathfrak{n}]$.

Fix a hyperfinite integer K. Let $\Gamma := \{z/K \mid z \in {}^*\mathbb{Z}\}$ and $\Delta t = 1/K$.

An element $A \in {}^*\mathcal{V}$ which happens to be a set, function, relation, etc. is called an internal set, function, relation, etc., respectively.

Given an interval $[a, b]$ in $\mathcal{V}$, an internal set $T \subseteq [a, b]$ is called a set of tag points (with respect to Γ) if T contains exactly one element in each subinterval $[c, d]$ where $c, d \in \Gamma \cap [a, b]$ with $d - c = \Delta t$.

Definition 1.22 *A standard bounded function $f : [a, b] \to \mathbb{R}$ is said to be Riemann integrable on $[a, b]$ if for any two internal sets $T, T' \subseteq {}^*[a, b]$ of tag points, we have*

$$\sum_{t \in T} {}^*f(t)\Delta t \approx \sum_{t \in T'} {}^*f(t)\Delta t.$$

If f is Riemann integrable on $[a, b]$, define the integration of f on $[a, b]$ by

$$\int_a^b f(x)dx := st\left(\sum_{t \in T} {}^*f(t)\Delta t\right)$$

for some internal set $T \subseteq {}^*[a, b]$ of tag points.

Theorem 1.23 *Given a standard bounded continuous function $g : [0, 1] \times \mathbb{R} \to \mathbb{R}$ and a number $\alpha \in \mathbb{R}$, there exists a standard function $y : [0, 1] \to \mathbb{R}$ satisfying Lipschitz condition such that*

$$y(x) = \alpha + \int_0^x g(s, y(s))ds \tag{3}$$

for every $x \in [0, 1]$.

Proof: Let $B \in \mathbb{R}$ be a bound of g and $\Gamma \cap [0, 1] = \{t_0 < t_1 < \cdots < t_N\}$. Define inductively on $n \le N$ such that $Y(t_0) = \alpha$ and

$$Y(t_{n+1}) = \alpha + \sum_{i=0}^{n} {}^*g(t_i, Y(t_i))\Delta t \tag{4}$$

for $n \in [N]$. Note that $Y(t_n) \in ns({}^*\mathbb{R})$ because g being bounded by $B \in \mathbb{R}$ implies ${}^*g$ being bounded by B by the transfer principle. Let $y(0) = st(Y(t_0))$. For each $x \in (0,1]$ let $y(x) = st(Y(x^-))$ where x^- is the largest $t_n \le x$ in $\Gamma \cap [0,1]$. It is easy to see that y satisfies Lipschitz condition on $[0,1]$. Indeed, if $0 \le z_1 \le z_2 \le 1$ are standard, then $\sum_{z_1^- < t_i \le z_2^-} \Delta t \approx z_2 - z_1$ and

$$|y(z_2) - y(z_1)| = \left| st \left(\sum_{z_1^- < t_i \le z_2^-} {}^*g(t_i, Y(t_i))\Delta t \right) \right| \le B(z_2 - z_1).$$

Similarly, $|Y(t_i) - Y(t_j)| \le B|t_j - t_i|$ for $t_i \le t_j$ in $\Gamma \cap [0,1]$. We show that y satisfies (3). By Definition 1.22 the integral at the right side of (3) is infinitesimally close to $\alpha + \sum_{i=0}^{n} {}^*g(t_i, {}^*y(t_i))\Delta t$ and the left side of (3) is infinitesimally close to $\alpha + \sum_{i=0}^{n} {}^*g(t_i, Y(t_i))\Delta t$. Hence, it suffices to show that

$$\sum_{i=0}^{n} \left({}^*g(t_i, {}^*y(t_i)) - {}^*g(t_i, Y(t_i))\right) \Delta t \approx 0.$$

If $st(t_i) = \beta$, then $Y(t_i) \approx Y(\beta^-) \approx y(\beta) \approx {}^*y(t_i)$. Hence, $\eta(t_i) := {}^*g(t_i, {}^*y(t_i)) - {}^*g(t_i, Y(t_i)) \approx 0$ by the continuity of g. This verifies (3). □

1.4 Exercises

1. Prove that $1/[Id]$ is a non-zero infinitesimal in ${}^*\mathcal{R}$ as defined in Example 1.11.

2. Prove that a standard sequence s of real numbers being convergent as defined in Definition 1.15 is equivalent to that s is a Cauchy sequence in the standard sense.

3. Let $f : [a,b] \to \mathbb{R}$ be a standard function. Prove that f being continuous at some $c \in (a,b)$ or uniformly continuous on $[a,b]$ in terms of Definition 1.17 is equivalent to that f is continuous at c or uniformly continuous on $[a,b]$, respectively, in the standard sense ($\epsilon - \delta$ definition).

4. Let $f : [a,b] \to \mathbb{R}$ be a standard bounded function. Prove that f being Riemann integrable on $[a,b]$ as defined in Definition 1.22 is equivalent to that f is Riemann integrable on $[a,b]$ in the standard sense.

2 Basic Methods

In Example 1.13 we mentioned that the nonstandard real field ${}^*\mathcal{R}$ does not satisfy the completeness property and pointed out that the property is not first-order. It is also true that a superstructure $\mathcal{V}$ as the model of standard mathematics, contains all sets in $\mathscr{P}(\mathbb{R})$ as its elements. Note that $\mathcal{V}$ satisfies the first-order sentence φ:

$$\forall x \in \mathscr{P}([0,1])(x \text{ has a least upper bound in } [0,1]). \tag{5}$$

Can we conclude by the transfer principle that the sentence

$$\forall x \in \mathscr{P}(^*[0,1])(x \text{ has a least upper bound in } ^*[0,1])$$

is true in $^*\mathcal{V}$? Of course, $^*\mathcal{R}$ in $^*\mathcal{V}$ should not satisfy the completeness property because there is no least upper bound of all infinitesimals. Does this cause inconsistency? To clarify the issue we should pay attention to the difference between internal sets and external sets.

2.1 Properties and Principles

Let $A \in \mathcal{V}$ be a set with rank $\leq \mathfrak{n}$. A subset A_0 of A is finite iff there is a bijection in $\mathcal{V}$ between A_0 and $[n]$ for some $n \in \mathbb{N}$. We denote $|A_0| = n$ for saying that A_0 has a cardinality n. The cardinality function $|\cdot|$ can be extend to a function $^*|\cdot|$ from all *finite subsets of $^*A$ to $^*\mathbb{N}$. So, $^*|A_1| = n$ iff there is a bijection in $^*\mathcal{V}$ between A_1 and $[n]$. For notational convenience, we omit * from $^*|\cdot|$. A set A_1 is called a hyperfinite set if $|A_1|$ is a hyperfinite integer.

Definition 2.1 *Every element or set of the form $^*a$ for some $a \in \mathcal{V}$ is called standard and every element or set $a \in {}^*\mathcal{V}$ is called internal. If an element or a set is not in $^*\mathcal{V}$, we call it external.*

Example 2.2 *Recall that $\mathbb{R}$ is a subset of $^*\mathbb{R}$.*

1. *Every $r \in \mathbb{R}$ is standard, and $^*\mathbb{N}$, $^*\mathbb{R}$ are standard.*
2. *For each hyperfinite integer N the sets $[N]$ and $[-N,N] \cap {}^*\mathbb{R}$ are internal but not standard.*
3. *The sets $\mathbb{N}$ and $\mathbb{R}$ are external subsets of $^*\mathbb{R}$.*

For Part 2 above let $N - 1 = [g]$ where $g : \mathbb{N} \to \mathbb{N}$ and $\{n \in \mathbb{N} \mid g(n) > m\} \in \mathcal{F}$ for each $m \in \mathbb{N}$. Then $^*\mathbb{N} \cap [0,[g]] \in {}^*\mathcal{V}$. If $^*a = {}^*\mathbb{N} \cap [0,[g]] \in {}^*\mathcal{V}$ is standard, then $^*a$ being bounded above in $^*\mathbb{N}$ implies a being bounded above in $\mathbb{N}$ by the transfer property. This means that a is a finite subset of $\mathbb{N}$. Since $^*r = r$ for every $r \in \mathbb{R}$, we have $^*a = a$ which is a finite set contradicting that $^*a$ is a hyperfinite set. Hence,

${}^*\mathbb{N} \cap [N]$ is internal but not standard. By a similar reason, the set ${}^*\mathbb{R} \cap [-N, N]$ is internal but not standard.

Note that the statement $\mathcal{V} \models \varphi$ for φ being in (5) is transferred to ${}^*\mathcal{V}$ to become

$$ {}^*\mathcal{V} \models \forall x \in {}^*\mathscr{P}({}^*[0,1])(x \text{ has a least upper bound in } {}^*[0,1]). $$

The reader should notice the difference between ${}^*\mathscr{P}({}^*[0,1])$ and $\mathscr{P}({}^*[0,1])$. The former is the collection of all internal subsets of ${}^*[0,1]$ and the latter is the collection of all subsets (internal or external) of ${}^*[0,1]$. So, in ${}^*\mathcal{V}$ every internal subset of ${}^*[0,1]$ has a least upper bound. Therefore, the set of all infinitesimals in ${}^*\mathbb{R}$ is not an internal set.

For Part 3 above, since every bounded subset of $\mathbb{N}$ is finite and has a maximal element in $\mathbb{N}$, by the transfer principle, every bounded internal subset of ${}^*\mathbb{N}$ is finite or hyperfinite and has a maximal element. But $\mathbb{N}$ as a subset of ${}^*\mathbb{N}$ does not have a maximal element. Therefore, $\mathbb{N}$ is not internal in ${}^*\mathbb{N}$. By a similar reason, $\mathbb{R}$ is not an internal subset of ${}^*\mathbb{R}$.

Proposition 2.3 (Definability of Internal Sets) *Let $A \in {}^*\mathcal{V}$ be an internal set with rank$(A) \leq \mathfrak{n}$ and $\varphi(\overline{x}, \overline{b})$ be a formula with parameters $\overline{b}$ in ${}^*\mathcal{V}$ where $\overline{x}$ is an m-tuple of variables. Then*

$$ \left\{\overline{a} \in A^m \mid {}^*\mathcal{V} \models \varphi(\overline{a}, \overline{b})\right\} \tag{6} $$

is again an internal subset of A^m.

Proof: Let $A = [f]$ and $\overline{b} = \overline{[g]}$. Define a function $h : \mathbb{N} \to \mathcal{V}$ by letting

$$ h(n) := \{\overline{a} \in f(n)^m \mid \mathcal{V} \models \varphi(\overline{a}, \overline{g(n)})\} $$

for each $n \in \mathbb{N}$. Let $B = [h]$. Then B is an internal subset of A^m. The proposition follows because

$$ \begin{aligned} \overline{[p]} \in B \text{ iff } & \{n \in \mathbb{N} \mid \overline{p(n)} \in h(n)\} \in \mathcal{F} \\ \text{iff } & \{n \in \mathbb{N} \mid \mathcal{V} \models \varphi(\overline{p(n)}, \overline{g(n)})\} \in \mathcal{F} \text{ iff } {}^*\mathcal{V} \models \varphi(\overline{[p]}, \overline{b}) \end{aligned} $$

by Theorem 1.9. □

If a subset B of an internal set A is itself internal, then B can be trivially defined by the formula $x \in B$ with parameter B. So, Proposition 6 says that a subset of an internal set is internal iff the subset is first-order definable.

A nonempty set $U \subseteq {}^*\mathbb{N}$ is an initial segment of ${}^*\mathbb{N}$ if $n \in U$ and $m < n$ imply $m \in U$ for any $m, n \in {}^*\mathbb{N}$. For example, $\mathbb{N}$ is an external initial segment of ${}^*\mathbb{N}$.

Proposition 2.4 (Overspill and Underspill Principle) *Let U be an external initial segment of ${}^*\mathbb{N}$ and A be an internal subset of ${}^*\mathbb{N}$.*

1. *If $A \cap U$ is unbounded above in U, then $A \setminus U \neq \emptyset$;*

2. *If $A \setminus U$ is unbounded below in ${}^*\mathbb{N} \setminus U$, then $A \cap U \neq \emptyset$.*

Proof: Part 1: Suppose $A \setminus U = \emptyset$. Then $U = \{x \in {}^*\mathbb{N} \mid \exists a \in A\,(x \leq a)\}$ is internal by Proposition 2.3 which contradicts the assumption that U is external. The proof of Part 2 is similar. □

The overspill and underspill principles are frequently used tools in nonstandard analysis.

Proposition 2.5 (Countable Saturation) *Let A be an infinite internal set in ${}^*\mathcal{V}$ with rank $\leq \mathfrak{n}$ and $A \supseteq B_0 \supseteq B_1 \supseteq \cdots$ be a nested sequence of nonempty internal sets. Then,*

$$\bigcap_{m\in\mathbb{N}} B_m \neq \emptyset.$$

Proof: Let $B_m = [b_m]$ for some $b_m \in \mathcal{V}^{\mathbb{N}}$ and choose an $[f_m] \in [b_m]$. For each $m \in \mathbb{N}$ let $U_m := \{n \in \mathbb{N} \mid n > m,\ f_m(n) \in b_m(n),\ \text{and}\ b_0(n) \supseteq b_1(n) \supseteq \cdots \supseteq b_m(n)\}$. Then $U_m \in \mathcal{F}$. For each $n \in \mathbb{N}$, let $m_n := \max\{m \in \mathbb{N} \mid n \in U_m\}$. Note that m_n exists because $\bigcap_{m\in\mathbb{N}} U_m = \emptyset$. Note also that $n \in U_{m_n}$. Let $f \in \mathcal{V}^{\mathbb{N}}$ be a function such that $f(n) = f_{m_n}(n)$ for every $n \in \mathbb{N}$. It suffices to show that $[f] \in [b_m]$ for every $m \in \mathbb{N}$.

Given $m \in \mathbb{N}$, let $U := \{n \in \mathbb{N} \mid f(n) \in b_m(n)\}$. For each $n \in U_m$, we have $m \leq m_n$ by the maximality of m_n. Since $n \in U_{m_n}$, we have $f(n) = f_{m_n}(n) \in b_{m_n}(n) \subseteq b_m(n)$. Hence, $n \in U$ which means $U_m \subseteq U$. Since $U_m \in \mathcal{F}$, we have that $U \in \mathcal{F}$, which implies $[f] \in [b_m]$. □

Countable saturation appeared first time in [19]. It is a key property in the development of Loeb measure. In Proposition 2.3 and Proposition 2.5 the set A is assumed to have rank $\leq \mathfrak{n}$ because some collection of subsets of A are mentioned which may have rank greater than $\mathfrak{n}$. Since the elements with rank higher than $\mathfrak{n}$ are still in $\mathcal{V}$ as long as the rank is $\leq 2\mathfrak{n}$. If the set A has a rank $2\mathfrak{n}$, then some objects needed will be outside of $\mathcal{V}$. Since all mathematical objects in our applications will have a rank $\leq \mathfrak{n}$ the restriction $\text{rank}(A) \leq \mathfrak{n}$ will not cause any problem. Although the rank of some element used in the proofs may not be mentioned, the reader should understand when it is assumed to have a rank below $\mathfrak{n}$.

Proposition 2.5 is still true if the sequence B_m is assumed to satisfy the finite intersection property, i.e., the intersection of any finite collection of B_m's is nonempty, instead of the sequence being nested. Proposition 2.5 is also true if $\mathcal{F}$ is a non-principal ultrafilter on any infinite set X as long as it is countably incomplete. For any infinite cardinal κ, there exist ultrafilters $\mathcal{F}$ such that the ultrapower of $\mathcal{V}$ modulo $\mathcal{F}$ satisfies κ-saturation property, i.e., any collection of less than κ many internal subsets of an internal set satisfying finite intersection property has a nonempty intersection.

The next two corollaries are trivial.

Corollary 2.6 *Every internal set A in ${}^*\mathcal{V}$ is either finite or uncountable.*

Corollary 2.7 *Let U be an infinite initial segment of ${}^*\mathbb{N}$. Let $\{x_n \in U \mid n \in \mathbb{N}\}$ be increasing and $\{y_n \in {}^*\mathbb{N} \setminus U \mid n \in \mathbb{N}\}$ be decreasing. Then either $\{x_n \in U \mid n \in \mathbb{N}\}$ is bounded above by some $z \in U$ or $\{y_n \in {}^*\mathbb{N} \setminus U \mid n \in \mathbb{N}\}$ is bounded below by some $z \in {}^*\mathbb{N} \setminus U$.*

Corollary 2.8 *Let $A \in {}^*\mathcal{V}$ and $s : \mathbb{N} \to A$ be an external sequence. There exists an internal function $S : {}^*\mathbb{N} \to A$ such that $S \restriction \mathbb{N} = s$.*

Proof: For each $m \in \mathbb{N}$ let

$$\mathcal{S}_m := \{t \in {}^*\mathcal{V} \mid t : {}^*\mathbb{N} \to A\, (t(i) = s(i) \text{ for } i \in [m+1]).$$

Note that $\mathcal{S}_m \in {}^*\mathscr{P}(A^{{}^*\mathbb{N}} \cap {}^*\mathcal{V})$ is nonempty because it contains at least an internal function s' such that $s'(i) = s(i)$ for $i \in [m+1]$ and $s'(i) = s(0)$ for any $i \in {}^*\mathbb{N} \setminus [m]$. It is easy to see that $\mathcal{S}_m \supseteq \mathcal{S}_{m+1}$. By Proposition 2.5 we can find $S : {}^*\mathbb{N} \to A$ such that $S \restriction \mathbb{N} = s$. □

Remark 2.9 *Note that if $s : \mathbb{N} \to A$ is an injection, we cannot require that $S : {}^*\mathbb{N} \to A$ be an injection in Corollary 2.8. However, if*

$$B := \{m \in {}^*\mathbb{N} \mid S \restriction [m+1] \text{ is an injection}\},$$

then B is internal and upper unbounded in $\mathbb{N}$. By Proposition 2.4 the set B contains some hyperfinite integer N. Hence, $S \restriction [N+1]$ is an injection from $[N+1]$ to A. For example, a strictly increasing sequence $\{r_i \mid i \in \mathbb{N}\}$ in some interval $[a,b] \subseteq {}^*\mathbb{R}$ may not be extended to an internal strictly increasing sequence $\{r_i \mid i \in {}^*\mathbb{N}\}$ in $[a,b]$. Instead, it can be extended to a hyperfinite strictly increasing sequence $\{r_i \mid 0 \leq i \leq N\}$ for some hyperfinite integer N.

2.2 Loeb Space Construction

In this course we introduce only Loeb probability space generated by an internal normalized counting measure on a hyperfinite set. For general approach to Loeb measure theory the reader should consult some literature such as [18, 25].

Definition 2.10 *Let Ω be a hyperfinite set in ${}^*\mathcal{V}$ and $\Sigma_0 := {}^*\mathscr{P}(\Omega)$ be the set of all internal subsets of Ω. Clearly, each $A \in \Sigma_0$ is a finite or hyperfinite set. For $A \in \Sigma_0$ define*

$$\delta(A) := \frac{|A|}{|\Omega|} \in {}^*[0,1] \text{ and } \mu_\Omega(A) := st(\delta(A)) \in [0,1].$$

Then, $(\Omega; \Sigma_0, \delta)$ is called a normalized counting measure space, and $(\Omega; \Sigma_0, \mu_\Omega)$ is called a standardized normalized counting measure space.

Definition 2.11 *Let $(\Omega; \Sigma_0, \mu_\Omega)$ be the standardized normalized counting measure space on a hyperfinite set Ω. For each $X \subseteq \Omega$ where X could be external, the upper measure and lower measure of X are defined by*

$$\overline{\mu}_\Omega(X) := \inf\{\mu_\Omega(A) \mid X \subseteq A \text{ and } A \in \Sigma_0\} \text{ and}$$

$$\underline{\mu}_\Omega(X) := \sup\{\mu_\Omega(A) \mid X \supseteq A \text{ and } A \in \Sigma_0\}.$$

Let $\Sigma := \{X \subseteq \Omega \mid \overline{\mu}_\Omega(X) = \underline{\mu}_\Omega(X)\}$. For each $X \in \Sigma$ define $\mu_\Omega(X) = \overline{\mu}_\Omega(X)$. Then, $(\Omega; \Sigma, \mu_\Omega)$ is called a Loeb probability space, or just Loeb space, generated by the normalized counting measure on Ω.

Proposition 2.12 *Let $(\Omega; \Sigma, \mu_\Omega)$ be a Loeb space defined in Definition 2.11. Then,*

1. *$\Sigma_0 \subseteq \Sigma$;*

2. *$\mu_\Omega(\Omega) = 1$ and $\mu_\Omega(\{x\}) = 0$ for each $x \in \Omega$;*

3. *If $Y \subseteq X \subseteq \Omega$, $X \in \Sigma$, and $\mu_\Omega(X) = 0$, then $Y \in \Sigma$ and $\mu_\Omega(Y) = 0$;*

4. *If $X, Y \in \Sigma$ and $Y \subseteq X$, then $\mu_\Omega(Y) \leq \mu_\Omega(X)$;*

5. *Let $X \subseteq \Omega$. Then, $X \in \Sigma$ iff X has squeezing sandwich sequences of internal sets A_i and B_i for $i \in \mathbb{N}$, i.e., (sandwich)*

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \subseteq X \subseteq \cdots \subseteq B_3 \subseteq B_2 \subseteq B_1 \subseteq \Omega,$$

   and (squeezing) $\lim_{m\to\infty} \mu_\Omega(B_m \setminus A_m) = 0$. Furthermore, if A_m, B_m are squeezing sandwich sequences for X, then $\mu_\Omega(X) = \lim_{m\to\infty} \mu_\Omega(A_m) = \lim_{m\to\infty} \mu_\Omega(B_m)$;

6. *Let* $X, Y \in \Sigma$.

   (a) $X \cup Y \in \Sigma$ *and* $\mu_\Omega(X \cup Y) \le \mu_\Omega(X) + \mu_\Omega(Y)$*;*

   (b) *If* $Y \subseteq X$*, then* $X \setminus Y \in \Sigma$ *and* $\mu_\Omega(X \setminus Y) = \mu_\Omega(X) - \mu_\Omega(Y)$*;*

   (c) *If* $X \cap Y = \emptyset$*, then* $\mu_\Omega(X \cup Y) = \mu_\Omega(X) + \mu_\Omega(Y)$*;*

   (d) $X \setminus Y \in \Sigma$.

7. *If* $X \in \Sigma$*, then there exists* $K \in \Sigma_0$ *such that* $\mu_\Omega(X \Delta K) = 0$*, where* $X \Delta K := (X \setminus K) \cup (K \setminus X)$*;*

8. *If* $X_i \in \Sigma$ *for* $i \in \mathbb{N}$ *is a pairwise disjoint sequence, then*

$$\mu_\Omega\left(\bigcup_{i\in\mathbb{N}} X_i\right) = \sum_{i\in\mathbb{N}} \mu_\Omega(X_i);$$

9. Σ *is a* σ*-algebra and* $(\Omega; \Sigma, \mu_\Omega)$ *is an atomless, complete, countably additive probability space in the standard sense.*

Proof: Part 1 is true because of the definition of lower and upper measure.

Part 2 is true because $|\Omega|/|\Omega| = 1$ and $st(1/|\Omega|) = 0$.

Part 3 is true because $0 = \overline{\mu}_\Omega(X) \ge \overline{\mu}_\Omega(Y) \ge \underline{\mu}_\Omega(Y) \ge 0$ implies $\overline{\mu}_\Omega(Y) = \underline{\mu}_\Omega(Y) = 0$.

Part 4 follows from the fact that if $A \subseteq B$ for internal sets $A, B \in \Sigma_0$, then $\delta(A) \le \delta(B)$.

Part 5: "$\Rightarrow$": Assume $X \in \Sigma$. For each $m \in \mathbb{N}$ there are internal sets $A_m, B_m \in \Sigma_0$ with $A_m \subseteq X \subseteq B_m$ such that $\delta(A_m) > \mu_\Omega(X) - 1/m$ and $\delta(B_m) < \mu_\Omega(X) + 1/m$. By taking unions of A_m's and intersections of B_m's we can assume that A_m's and B_m's are sandwich sequences of X. Since $\delta(B_m \setminus A_m) = \delta(B_m) - \delta(A_m) < 2/m$, we have that the sequences are squeezing, i.e., $\lim_{m\to\infty} \mu_\Omega(B_m \setminus A_m) = 0$.

"$\Leftarrow$": Since $\mu_\Omega(B_m \setminus A_m) \to 0$ we have that

$$\alpha = \lim_{m\to\infty} \mu_\Omega(A_m) = \lim_{m\to\infty} \mu_\Omega(B_m) = \beta.$$

Note that $\alpha \le \underline{\mu}_\Omega(X) \le \overline{\mu}_\Omega(X) \le \beta$. So, $\underline{\mu}_\Omega(X) = \overline{\mu}_\Omega(X) = \mu_\Omega(X) = \alpha = \beta$, which clearly implies $X \in \Sigma$.

Part 6: Let A_m and B_m be squeezing sandwich sequences for X, and A'_m and B'_m be squeezing sandwich sequences for Y.

(a): Since $A_m \cup A'_m$ and $B_m \cup B'_m$ are sandwich sequences of $X \cup Y$ and

$$(B_m \cup B'_m) \setminus (A_m \cup A'_m) \subseteq (B_m \setminus A_m) \cup (B'_m \setminus A'_m),$$

we have that

$$\lim_{m\to\infty} \mu_\Omega((B_m \cup B'_m) \setminus (A_m \cup A'_m)) \le \lim_{m\to\infty} \mu_\Omega(B_m \setminus A_m) + \lim_{m\to\infty} \mu_\Omega(B'_m \setminus A'_m) = 0,$$

which implies $X \cup Y \in \Sigma$ by Part 5 and hence,

$$\mu_\Omega(X \cup Y) = \lim_{m\to\infty} \mu_\Omega(B_m \cup B'_m) \le \lim_{m\to\infty} \mu_\Omega(B_m) + \lim_{m\to\infty} \mu_\Omega(B'_m) = \mu_\Omega(X) + \mu_\Omega(Y).$$

(b): Note that $A_m \setminus B'_m \subseteq X \setminus Y \subseteq B_m \setminus A'_m$, which mean $A_m \setminus B'_m$ and $B_m \setminus A'_m$ are sandwich sequences for $X \setminus Y$. Since $(B_m \setminus A'_m) \setminus (A_m \setminus B'_m) \subseteq (B_m \setminus A_m) \cup (B'_m \setminus A'_m)$, we have that $A_m \setminus B'_m$ and $B_m \setminus A'_m$ are squeezing. So, $X \setminus Y \in \Sigma$ and $\mu_\Omega(X \setminus Y) = \lim_{m\to\infty} \mu_\Omega(B_m \setminus A'_m) = \lim_{m\to\infty} \mu_\Omega(B_m) - \lim_{m\to\infty} \mu_\Omega(A'_m) = \mu_\Omega(X) - \mu_\Omega(Y)$.

In particular, we have $X^c \in \Sigma$, where $X^c := \Omega \setminus X$, and $\mu_\Omega(X^c) = 1 - \mu_\Omega(X)$.

(c): If $X \cap Y = \emptyset$, then $Y \subseteq X^c$. Hence, $X \cup Y = (X^c \cap Y^c)^c = (X^c \setminus Y)^c$ and $\mu_\Omega(X \cup Y) = 1 - (\mu(X^c \setminus Y)) = 1 - (\mu_\Omega(X^c) - \mu_\Omega(Y)) = 1 - (1 - \mu_\Omega(X) - \mu_\Omega(Y)) = \mu_\Omega(X) + \mu_\Omega(Y)$.

(d): $X \setminus Y = X \cap Y^c = (X^c \cup Y)^c \in \Sigma$.

Part 7: Let A_m and B_m be a squeezing sandwich sequences for X. Let $\mathcal{K}_m = \{K \in \Sigma_0 \mid A_m \subseteq K \subseteq B_m\}$. Then, $\mathcal{K}_m$ is nonempty, internal, and $\mathcal{K}_{m+1} \subseteq \mathcal{K}_m$. By Proposition 2.5 there is a $K \in \bigcap_{m\in\mathbb{N}} \mathcal{K}_m$. Clearly, A_m, B_m are squeezing sandwich sequences for K. Since $X \Delta K \subseteq B_m \setminus A_m$, we have that $\mu_\Omega(X \Delta K) \le \mu_\Omega(B_m \setminus A_m) \to 0$. So, $\mu_\Omega(X \Delta K) = 0$.

Part 8: By passing to subsequences we can find squeezing sandwich sequences $A_m^{(i)}, B_m^{(i)}$ for each X_i such that

$$\max\{\mu_\Omega(B_m^{(i)} \setminus X_i), \mu_\Omega(X_i \setminus A_m^{(i)})\} \le \mu_\Omega(B_m^{(i)} \setminus A_m^{(i)}) < 1/2^i m.$$

Note that $A_m^{(i)}$ for $i = 1, 2, \dots$ are pairwise disjoint. For each $m \in \mathbb{N}$ we can find a hyperfinite integer N_m such that the sequences $\{A_m^{(i)}, B_m^{(i)} \mid i \in \mathbb{N}\}$ can be extended to internal sequences $\{A_m^{(i)}, B_m^{(i)} \mid 1 \le i \le N_m\}$ such that $\delta(B_m^{(i)} \setminus A_m^{(i)}) < 1/2^i m$ for $0 \le i \le N_m$. By Corollary 2.7 there is a hyperfinite integer $N \le N_m$ for every $m \in \mathbb{N}$. So, for any $m \in \mathbb{N}$ and $0 \le i \le N$ we have $\delta(B_m^{(i)} \setminus A_m^{(i)}) < 1/2^i m$. For each $m \in \mathbb{N}$ let

$$B_m := \bigcup_{i=1}^{N} B_m^{(i)} \text{ and } A_m := \bigcup_{i=1}^{m} A_m^{(i)}.$$

Clearly, A_m, B_m are sandwich sequences of internal sets for $X := \bigcup_{i\in\mathbb{N}} X_i$. It suffices to show that the sequences are also squeezing.

Since

$$\sum_{i=1}^{m} \mu_\Omega(X_i) = \mu_\Omega\left(\bigcup_{i=1}^{m} X_i\right) \leq 1$$

by Part 6, we have that $\lim_{m\to\infty} T_m = 0$ where $T_m := \sum_{i=m+1}^{\infty} \mu_\Omega(X_i)$. Given any $m' > m$ in $\mathbb{N}$, we have

$$\begin{aligned}\sum_{i=m+1}^{m'} \delta(B_m^{(i)}) &\leq \frac{1}{m} + \sum_{i=m+1}^{m'} \mu_\Omega(B_m^{(i)} \setminus X_i) + \sum_{i=m+1}^{m'} \mu_\Omega(X_i)\\ &\leq \frac{1}{m} + \frac{1}{m}\sum_{i=m+1}^{m'} \frac{1}{2^i} + T_m \leq \frac{1}{m} + \frac{1}{m2^m} + T_m \leq \frac{2}{m} + T_m.\end{aligned}$$

By extending m' to hyperfinite we can assume that

$$\delta\left(\bigcup_{i=m+1}^{N} B_m^{(i)}\right) \leq \sum_{i=m+1}^{N} \delta(B_m^{(i)}) \leq \frac{2}{m} + T_m.$$

So,

$$\begin{aligned}\mu_\Omega(B_m \setminus A_m) &\leq \mu_\Omega\left(\bigcup_{i=1}^{m} B_m^{(i)} \setminus \bigcup_{i=1}^{m} A_m^{(i)}\right) + \mu_\Omega\left(\bigcup_{i=m+1}^{N} B_m^{(i)}\right)\\ &\leq \mu_\Omega\left(\bigcup_{i=1}^{m}(B_m^{(i)} \setminus A_m^{(i)})\right) + \frac{3}{m} + T_m \leq \sum_{i=1}^{m} \frac{1}{2^i m} + \frac{3}{m} + T_m \leq \frac{4}{m} + T_m \to 0\end{aligned}$$

as $m \to \infty$. Therefore, A_m, B_m are squeezing for X which implies $X \in \Sigma$. Note that

$$\begin{aligned}\mu_\Omega(X) &= \sum_{i=1}^{m} \mu_\Omega(X_i) + \mu_\Omega\left(\bigcup_{i=m+1}^{\infty} X_i\right)\\ &\leq \sum_{i=1}^{m} \mu_\Omega(X_i) + \frac{1}{m} + \delta\left(\bigcup_{i=m+1}^{N} B_m^{(i)}\right)\\ &\leq \sum_{i=1}^{m} \mu_\Omega(X_i) + \frac{3}{m} + T_m \to \sum_{i=1}^{\infty} \mu_\Omega(X_i)\end{aligned}$$

as $m \to \infty$, and

$$\mu_\Omega(X) = \lim_{m\to\infty} \mu_\Omega(A_m) = \lim_{m\to\infty} \sum_{i=1}^{m} \mu_\Omega(A_m^{(i)})$$

$$= \lim_{m\to\infty} \sum_{i=1}^{m} \left(\mu_\Omega(X_i) - \mu_\Omega(X_i \setminus A_m^{(i)})\right) \geq \lim_{m\to\infty} \sum_{i=1}^{m} \left(\mu_\Omega(X_i) - \frac{1}{2^i m}\right)$$
$$= \lim_{m\to\infty} \sum_{i=1}^{m} \mu_\Omega(X_i) - \lim_{m\to\infty} \sum_{i=1}^{m} \frac{1}{2^i m} = \sum_{i=1}^{\infty} \mu_\Omega(X_i).$$

We conclude that $\mu_\Omega(X) = \sum_{i\in\mathbb{N}} \mu_\Omega(X_i)$.

Part 9: Σ is a σ-algebra by Part 6 and 8. $(\Omega; \Sigma, \mu_\Omega)$ is complete by Part 3, and countably additive by Part 8. If $X \in \Sigma$ with $\mu_\Omega(X) > 0$, we can find an internal set $A \subseteq X$ such that $\delta(A) > \mu_\Omega(X)/2 > 0$. Since A is *finite, we can find an internal set $B \subseteq A$ such that $|A| = 2|B|$ or $|A| = 2|B| + 1$. For each case $\mu_\Omega(B) = \mu_\Omega(A)/2$ and $\mu_\Omega(X \setminus B) \geq \mu_\Omega(A)/2$. So, $(\Omega; \Sigma, \mu_\Omega)$ is atomless. □

Theorem 2.13 *Let* $(\Omega; \Sigma, \mu_\Omega)$ *be a Loeb space on a hyperfinite set* Ω *and* $f : \Omega \to \mathbb{R}\cup\{\pm\infty\}$ *be a measurable function, i.e.,* $f^{-1}(O) \in \Sigma$ *for any open set* O *in* $\mathbb{R}\cup\{\pm\infty\}$*, then, there is an internal function* $F : \Omega \to {}^*\mathbb{R}$ *such that for almost all* $\omega \in \Omega$ *we have*

$$st(F(\omega)) = f(\omega).$$

Proof: Let $\mathcal{U} := \{O_n \mid n \in \mathbb{N}\}$ be a topological basis of $\mathbb{R} \cup \{\pm\infty\}$. For each $O_n \in \mathcal{U}$ let $A_{n,m} \subseteq f^{-1}(O_n)$ be increasing with respect to m such that $\lim_{m\to\infty} \mu_\Omega(A_m) = \mu_\Omega(f^{-1}(O_n))$. For each $m \in \mathbb{N}$ let

$$\mathcal{G}_m := \left\{ g : \bigcup_{n<m} A_{n,m} \to {}^*\mathbb{R} \mid g \text{ is internal and } g[A_{n,m}] \subseteq {}^*O_n \right\}.$$

It is easy to see that $\mathcal{G}_m$ is nonempty, internal, and decreasing. By Proposition 2.5 there is an $F \in \bigcap_{m\in\mathbb{N}} \mathcal{G}_m$. Note that the set

$$Z := \bigcup_{n\in\mathbb{N}} \left(f^{-1}(O_n) \setminus \bigcup_{m\in\mathbb{N}} A_{n,m} \right)$$

is a countable union of Loeb measure zero sets. Hence, $\mu_\Omega(Z) = 0$. For each $\omega \in \Omega \setminus Z$ and $O_n \in \mathcal{U}$, if $f(\omega) \in O_n$, then $\omega \in A_{n,m}$ for some $m > n$. Hence, $F(\omega) \in {}^*O_n$ which implies $st(F(\omega)) = f(\omega)$. □

2.3 Application to Finance

We present an application of nonstandard analysis to finance theory due to Dr. Yeneng Sun. This application may technically be the simplest one among all Dr. Sun's contributions to mathematical economics.

Given two hyperfinite Loeb spaces $(\Omega; \Sigma, \mu_\Omega)$ and $(\Psi; \Gamma, \nu_\Psi)$, one can form two different product measure spaces on $\Omega \times \Psi$. The first one is the standard product measure space. For any two standard probability spaces $(\Omega; \Sigma, \mu)$ and $(\Psi; \Gamma, \nu)$ a rectangle is a set of form $A \times B$ for some $A \in \Sigma$ and $B \in \Gamma$. The measure $\mu \times \nu(A \times B) := \mu(A) \cdot \nu(B)$. Let $\Sigma \times \Gamma$ be the collection of all finite union of disjoint rectangles. The measure $\mu \times \nu$ can be trivially generalized to sets in $\Sigma \times \Gamma$. Note that

$$(\Omega \times \Psi; \Sigma \times \Gamma, \mu \times \nu)$$

is a finitely additive probability space. By the same process as in Proposition 2.12 the measure $\mu \times \nu$ can uniquely be extended to the σ-algebra $\sigma(\Sigma \times \Gamma)$ generated by $\Sigma \times \Gamma$. By including in all subsets of zero-measure sets one can make the measure $\mu \times \nu$ complete. The space

$$(\Omega \times \Gamma; \sigma(\Sigma \times \Psi), \mu \times \nu)$$

is called the standard product measure space on $\Omega \times \Psi$.

The product measure space on $\Omega \times \Psi$ in the rest of this subsection is different from the standard one.

Let's consider the product space of two hyperfinite Loeb spaces $(\Omega; \Sigma, \mu_\Omega)$ and $(\Psi; \Gamma, \nu_\Psi)$. Since $\Omega \times \Psi$ is again a hyperfinite set, one can form the Loeb probability space generalized by the normalized counting measure on all internal subsets of $\Omega \times \Psi$. Denote this *Loeb product space* by

$$(\Omega \times \Psi; \Sigma \otimes \Gamma, \mu_\Omega \otimes \nu_\Psi).$$

Since a finite union of disjoint rectangles is an internal subset of $\Omega \times \Psi$, we have that $\Sigma \times \Gamma \subseteq \Sigma \otimes \Gamma$. Since $\Sigma \otimes \Gamma$ is a σ-algebra and contains all subsets of zero-measure sets with respect to $\mu_\Omega \otimes \nu_\Psi$, we have that

$$\sigma(\Sigma \times \Gamma) \subseteq \Sigma \otimes \Gamma \text{ and } \mu_\Omega \otimes \nu_\Psi \upharpoonright \sigma(\Sigma \times \Gamma) = \mu_\Omega \times \nu_\Psi.$$

Theorem 2.14 (Keisler's Fubini Theorem, Corollary 6.3.17 in [18])
Let $(\Omega; \Sigma, \mu_\Omega)$ and $(\Psi; \Gamma, \nu_\Psi)$ be two Loeb spaces. Assume that $f : \Omega \times \Psi \to \mathbb{R}$ is an integrable function on the Loeb product space $(\Omega \times \Gamma, \Sigma \otimes \Gamma, \mu_\Omega \otimes \nu_\Psi)$. Then,

1. *for ν_Ψ-almost all $y \in \Psi$, $f_y(x) := f(x, y)$ is μ_Ω-integrable,*
2. *$F(y) := \int_\Omega f(x, y) d\mu_\Omega(x)$ is ν_Ψ-integrable, and*
3. *$\int_\Psi \int_\Omega f(x, y) d\mu_\Omega(x) d\nu_\Psi(y) = \int_{\Omega \times \Psi} f(x, y) d\mu_\Omega \otimes \nu_\Psi.$*

Imagine that an insurance company has a life insurance policy for people satisfying certain conditions. Each policy could bring a gain or loss of some values for the company with certain probability distribution. It is a common sense that if the identical policy is sold to enough many policy holders and each of these policy holders lives an independent life, then the company's financial risk of selling the policy can be diminished.

How can this phenomenon be mathematically modeled?

Definition 2.15 *Fix a probability space* $(\Omega; \Sigma, \mu)$. *A random variable is a measurable function* $v(\omega) : \Omega \to \mathbb{R}$.

1. *By an individual insurance agent (for example, an insurance policy holder) we mean a random variable* $f_i(\omega) : \Omega \to \mathbb{R}$.

2. *By an insurance system we mean a function* $f : \Omega \times I \to \mathbb{R}$ *such that* $f_i(\omega) := f(\omega, i)$ *for each* $i \in I$ *is an insurance agent.*

To find an idealize the model of the phenomenon, the number of insurance agents $|I|$ should be infinite. To measure the size of the certain group of agents, there should be a measure on the set I. Since a measure should be countably additive, the size of I should be uncountable. For example, the Lebesgue measure on the unit interval of reals $[0, 1]$ is the measure space on an uncountable set $[0, 1]$.

Definition 2.16 *Let* $(\Omega; \Sigma, \mu)$ *and* $(\Psi; \Gamma, \nu)$ *be two probability spaces.*

1. *A function* $f : \Omega \times \Psi \to \mathbb{R}$ *is said to be jointly measurable if* f *is measurable with respect to the standard product space* $(\Omega \times \Psi; \sigma(\Sigma \times \Gamma), \mu \times \nu)$;

2. *Suppose a function* $f : \Omega \times \Psi \to \mathbb{R}$ *satisfies that* $f^\omega(i) := f(\omega, i)$ *is* $(\Psi; \Gamma, \nu)$ *measurable for almost every* $\omega \in \Omega$ *and* $f_i(\omega) := f(\omega, i)$ *is* $(\Omega; \Sigma, \mu)$ *measurable for almost every* $i \in \Psi$. *The function* f *is almost pairwise independent on* Ψ *if for* $\nu \times \nu$*-almost all pairs* $(i, i') \in \Psi \times \Psi$, *the random variables* $f_i(\omega)$ *and* $f_{i'}(\omega)$ *are independent.*

Theorem 2.17 (Joseph L. Doob, Proposition 8.3.3 in [18]) *Let* $(\Omega; \Sigma, \mu)$ *and* $(\Psi; \Gamma, \nu)$ *be two probability spaces and* $f : \Omega \times \Psi \to \mathbb{R}$ *be a function such that*

1. f *is jointly measurable and square-integrable;*

2. f *is almost pairwise independent on* Ψ.

Then, for ν-almost all $i \in \Psi$, the random variable $f_i(\omega)$ is μ-almost surely a constant function.

By Theorem 2.17 there is no non-trivial insurance system can be jointly measurable with respect to the standard product of the insurance policy space and the space of insurance agents which are pairwise independent.

Example 2.18 *Let N be a hyperfinite integer and $\Omega = \{\omega \mid \omega : [N] \to [2]\}$. Then, Ω is a hyperfinite set and $|\Omega| = 2^N$. Let $(\Omega; \Sigma, \mu_\Omega)$ be the Loeb space on Ω. Let $\Psi = [N]$ and $(\Psi; \Gamma, \nu_\Psi)$ be the Loeb space on Ψ. For each $i \in \Psi$ let $f_i : \Omega \to \mathbb{R}$ be defined as $f_i(\omega) := \omega(i)$. Then each f_i is a $0,1$-valued random variable on Ω and*

$$\mu_\Omega(\{\omega \mid f_i(\omega) = 0\}) = 1/2.$$

Each f_i can be viewed as a coin flip.

For any $i \neq i'$ in T, f_i and $f_{i'}$ are independent and have identical probability distribution.

Clearly, $f(\omega, i) := f_i(\omega)$ defines a measurable function on the Loeb product $(\Omega \times \Psi; \Sigma \otimes \Gamma, \mu_\Omega \otimes \nu_\Psi)$ such that all f_i are non-trivial.

Theorem 2.19 (Y. Sun, Theorem 8.5.3 in [18]) *Let $(\Omega; \Sigma, \mu_\Omega)$ and $(\Psi; \Gamma, \nu_\Psi)$ be two Loeb spaces and $f : \Omega \times \Psi \to \mathbb{R}$ be a square-integrable insurance system in $(\Omega \times \Psi, \Sigma \otimes \Gamma, \mu_\Omega \otimes \nu_\Psi)$. If the insurance agents f_i and $f_{i'}$ are independent for almost all (i, i') in $\Psi \times \Psi$, then for almost all $\omega \in \Omega$*

$$\int_\Psi f(\omega, i) d\nu_\Psi = \int_{\Psi \times \Omega} f(\omega, i) d\mu_\Omega \otimes \nu_\Psi = \int_\Psi \int_\Omega f(\omega, i) d\mu_\Omega d\nu_\Psi.$$

The theorem above is called the Exact Law of Large Numbers which indicates that the average pay-off of all insurance agents under particular realization ω for almost all $\omega \in \Omega$ is a constant which is the average pay-off of one agent.

2.4 Exercises

1. Let A be a set in $\mathcal{V}$. Prove that ${}^*A = \{{}^*a \mid a \in A\}$ iff A is a finite set.
2. Prove that an internal set $A \in {}^*\mathcal{V}$ is either finite or uncountable.
3. Let N be a hyperfinite integer, $\Omega := \{j/N \mid j = 0, 1, \ldots, N-1\}$, and $(\Omega; \Sigma, \mu_\Omega)$ be the Loeb space on Ω. Note that $st \restriction \Omega$ is a function from Ω to the standard unit interval $[0, 1]$ (cf. Definition 1.14). Let $\Gamma := \{U \subseteq [0, 1] \mid st^{-1}[U] \cap \Omega \in \Sigma\}$ and $\lambda(U) := \mu_\Omega(st^{-1}[U] \cap \Omega)$ for each $U \in \Gamma$. Prove that $([0, 1]; \Gamma, \lambda)$ is the Lebesgue measure space on $[0, 1]$.

4. Let $(\Omega; \Sigma, \mu_\Omega)$ and $(\Psi; \Gamma, \nu_\Psi)$ be two Loeb spaces defined in Example 2.18. Let

$$A := \{(\omega, i) \in \Omega \times \Psi \mid \omega(i) = 0\}.$$

Note that $A \in \Sigma \otimes \Gamma$ because A is internal. Prove that $\mu_\Omega \otimes \nu_\Psi\,(A) = 1/2$ and $A \notin \sigma(\Sigma \times \Gamma)$.

3 Easy Applications to Combinatorics

An apparent reason why nonstandard analysis should be a useful tool for other fields of mathematics is that a limit process which involves rank 3 objects in $\mathcal{V}$ such as the limit of a sequence or a function with real values can be changed to an infinitesimal argument with rank 0 objects such as infinitesimals in ${}^*\mathcal{V}$. So, good candidates for the applications of nonstandard analysis should be something involving limit processes. This may be why the density problems receive attention from nonstandard analysts. The densities introduced in this section are Shnirel'man density, lower and upper (asymptotic) density, and lower and upper Banach density.

For two sets $A, B \subseteq \mathbb{N}$, let $A + B := \{a + b \mid a \in A \text{ and } b \in B\}$. If $A = \{a\}$ we write $a + B$ instead of $\{a\} + B$ for simplicity. If $r, r' \in {}^*\mathbb{R}$, we write $r \gtrapprox r'$ for $r > r'$ or $r \approx r'$ and $r \lessapprox r'$ for $r < r'$ or $r \approx r'$.

Definition 3.1 *Let $A \subseteq \mathbb{N}$. The Shnirel'man density $\sigma(A)$, lower density $\underline{d}(A)$, upper density $\overline{d}(A)$, upper Banach density $\overline{BD}(A)$, and lower Banach density $\underline{BD}(A)$ of A are defined by*

1. $\sigma(A) := \inf_{n \geq 1} \dfrac{|A \cap (1 + [n])|}{n}$;
2. $\underline{d}(A) := \liminf_{n \to \infty} \dfrac{|A \cap [n]|}{n}$;
3. $\overline{d}(A) := \limsup_{n \to \infty} \dfrac{|A \cap [n]|}{n}$;
4. $\overline{BD}(A) := \lim_{n \to \infty} \sup_{k \in \mathbb{N}} \dfrac{|A \cap (k + [n])|}{n}$;
5. $\underline{BD}(A) := \lim_{n \to \infty} \inf_{k \in \mathbb{N}} \dfrac{|A \cap (k + [n])|}{n}$.

Remark 3.2 *1. In the definition of $\sigma(A)$, we have $1 + [n] = \{1, 2, \ldots, n\}$. Hence, 0, in or not in A, does not play any role. If $\sigma(A) > 0$, then $1 \in A$;*

2. *If $\underline{d}(A) = \overline{d}(A)$, we say that the (asymptotic) density of A exists and is denoted by $d(A)$;*

3. *If $\underline{BD}(A) = \overline{BD}(A)$, we say that the Banach density of A exists and is denoted by $BD(A)$;*

4. *In the definition of $\overline{BD}(A)$ the limit of $\sup_{k \in \mathbb{N}} \frac{|A \cap (k + [n])|}{n}$ as $n \to \infty$ always exists.*

The following Proposition is direct consequences of the definition.

Proposition 3.3 *For any $A \subseteq \mathbb{N}$ we have*

$$0 \leq \min\{\sigma(A), \underline{BD}(A)\} \leq \max\{\sigma(A), \underline{BD}(A)\} \leq \underline{d}(A) \leq \overline{d}(A) \leq \overline{BD}(A) \leq 1.$$

Lemma 3.4 *Let $A \subseteq \mathbb{N}$. Then, $\overline{BD}(A)$ is the largest real α in $[0, 1]$ such that there exist $k_m, n_m \in \mathbb{N}$ with $n_m \to \infty$ as $m \to \infty$ such that*

$$\lim_{m \to \infty} \frac{|A \cap (k_m + [n_m])|}{n_m} = \alpha.$$

The proof of the lemma is left to the reader.

3.1 Nonstandard Versions of Densities

Proposition 3.5 *Let $A \subseteq \mathbb{N}$ and $\alpha \in \mathbb{R}$. Then*

1. *$\underline{d}(A) \geq \alpha$ iff $\frac{|{}^*A \cap [N]|}{N} \gtrapprox \alpha$ for any hyperfinite integer N;*

2. *$\overline{d}(A) \geq \alpha$ iff $\frac{|{}^*A \cap [N]|}{N} \gtrapprox \alpha$ for some hyperfinite integer N.*

Proof: Part 1. "⇒": Let N be an arbitrary hyperfinite integer. Since for each $\epsilon > 0$, there exists an $n_0 \in \mathbb{N}$ such that

$$\forall n \in \mathbb{N} \left(n \geq n_0 \to \frac{|A \cap [n]|}{n} > \alpha - \epsilon \right).$$

By the transfer principle, it is true that

$$\forall n \in {}^*\mathbb{N} \left(n \geq n_0 \to \frac{|{}^*A \cap [n]|}{n} > \alpha - \epsilon \right).$$

Since $N \in {}^*\mathbb{N}$ and $N \geq n_0$, we have $\dfrac{|{}^*A \cap [N]|}{N} > \alpha - \epsilon$. Since $\epsilon > 0$ can be arbitrarily small, we have that $\dfrac{|{}^*A \cap [N]|}{N} \gtrapprox \alpha$.

Part 1. "$\Leftarrow$": Suppose $\underline{d}(A) < \alpha$. Let $\alpha' = (\alpha + \underline{d}(A))/2$, then there is an increasing sequence $n_1 < n_2 < \cdots$ such that $\forall i \in \mathbb{N} \left(\dfrac{|A \cap [n_i]|}{n_i} < \alpha'\right)$. By the transfer principle the sentence $\forall i \in {}^*\mathbb{N} \left(\dfrac{|{}^*A \cap [n_i]|}{n_i} < \alpha'\right)$ is true in ${}^*\mathcal{V}$. Let N' be a hyperfinite integer and $N := n_{N'}$. Then, N is hyperfinite and $\dfrac{|{}^*A \cap [N]|}{N} \lessapprox \alpha' < \alpha$. Hence, the right side of Part 1 is false.

The proof of Part 2 is left to the reader. □

Proposition 3.6 *Let $A \subseteq \mathbb{N}$ and $\alpha \in \mathbb{R}$. Then*

1. $\underline{BD}(A) \geq \alpha$ *iff* $\dfrac{|{}^*A \cap (k + [N])|}{N} \gtrapprox \alpha$ *for any $k \in {}^*\mathbb{N}$ and any hyperfinite integer N;*

2. $\overline{BD}(A) \geq \alpha$ *iff* $\dfrac{|{}^*A \cap (k + [N])|}{N} \gtrapprox \alpha$ *for some $k \in {}^*\mathbb{N}$ and some hyperfinite integer N.*

Proof: We prove Part 2. The proof of Part 1 is left to the reader.

Part 2. "$\Rightarrow$": Given $m \in \mathbb{N}$, there exist $k_m \in \mathbb{N}$ and $n_m > m$ such that

$$\frac{|A \cap (k_m + [n_m])|}{n_m} > \alpha - \frac{1}{m}.$$

By the transfer principle, we have that for any $m \in {}^*\mathbb{N}$ there exist $k_m \in {}^*\mathbb{N}$ and $n_m > m$ such that

$$\frac{|{}^*A \cap (k_m + [n_m])|}{n_m} > \alpha - \frac{1}{m}.$$

Now let m be a hyperfinite integer, $k := k_m$, and $N := n_m > m$. Then,

$$\frac{|{}^*A \cap (k + [N])|}{N} \gtrapprox \alpha.$$

Part 2. "$\Leftarrow$": Assume that $\overline{BD}(A) < \alpha$. Let $\alpha' = (\alpha + \overline{BD}(A))/2$. Then, there exists an $n_0 \in \mathbb{N}$ such that the following sentence is true in $\mathcal{V}$:

$$\forall k, n \in \mathbb{N} \left(n \geq n_0 \rightarrow \frac{|A \cap (k + [n])|}{n} \leq \alpha'\right).$$

By the transfer principle, the following is true in ${}^*\mathcal{V}$:

$$\forall k, n \in {}^*\mathbb{N} \left(n \geq n_0 \rightarrow \frac{|A \cap (k + [n])|}{n} \leq \alpha'\right).$$

Since hyperfinite integers are greater than n_0, the right side of Part 2 is false. □

3.2 By-one-get-one-free Thesis

Shnirel'man density and lower density are most used densities by number theorists. For example, Shnirel'man proved that if a set A has positive Shnirel'man density, then there is a fixed k such that every positive integer is the sum of at most k numbers in A. If P is the set of all prime numbers, then $A := (\{0, 1\} \cup P) + (\{0, 1\} \cup P)$ has positive Shnirel'man density, therefore, every positive integer is the sum of at most $2k$ prime numbers. This is the first nontrivial result towards the solution of Goldbach conjecture.

The buy-one-get-one-free thesis is the following statement:

> There is a parallel result involving upper Banach density for every existing result involving Shnirel'man density or lower density.

The thesis makes sense because of the following two theorems.

Theorem 3.7 *If $A \subseteq \mathbb{N}$ and $\overline{BD}(A) = \alpha$, then there is an $k \in {}^*\mathbb{N}$ and a hyperfinite integer N such that for μ_Ω-almost all $n \in k + [N]$ where μ_Ω is the Loeb measure on $\Omega := k + [N]$, we have $\underline{d}((^*A - n) \cap \mathbb{N}) = \alpha$. On the other hand, if $A \subseteq \mathbb{N}$ and there is a positive integer $n \in {}^*\mathbb{N}$ such that $\underline{d}((^*A - n) \cap \mathbb{N}) \geq \alpha$, then $\overline{BD}(A) \geq \alpha$.*

Theorem 3.8 *If $A \subseteq \mathbb{N}$ and $\overline{BD}(A) = \alpha$, then there is an $n \in {}^*\mathbb{N}$ such that*

$$\sigma((^*A - n) \cap \mathbb{N}) = \alpha.$$

To present short proofs of Theorem 3.7 and Theorem 3.8 we borrow the following Birkhoff's Ergodic Theorem.

Theorem 3.9 (Birkhoff's Ergodic Theorem, Theorem 2.3 in [22])
Let (Ω, Σ, μ) be a probability space and T be a measure-preserving transformation from Ω to Ω. For every $f \in L_1(\Omega)$, there exists a $\bar{f} \in L_1(\Omega)$ such that for μ-almost all $x \in \Omega$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \bar{f}(x),$$

where T^0 is the identity map and $T^{k+1}(x) = T(T^k(x))$ for every $k \in \mathbb{N}$.

Prove of Theorem 3.7: We prove the second part first. Assume that $\underline{d}((^*A - k) \cap \mathbb{N}) \geq \alpha$ for some $k \in {}^*\mathbb{N}$. For each $m \in \mathbb{N}$ there exists $n_m \in \mathbb{N}$ such that

$$\frac{|^*A \cap (k + [n])|}{n} \geq \alpha - \frac{1}{m}$$

for every $n \geq n_m$. By Proposition 2.8 there is a hyperfinite integer N' such that

$$\frac{|{}^*A \cap (k + [n])|}{n} \geq \alpha - \frac{1}{N'} \approx \alpha$$

for every $n \geq n_{N'}$. Choose $N \geq n_{N'}$ to be hyperfinite. Then,

$$\frac{|{}^*A \cap (k + [N])|}{N} \gtrapprox \alpha,$$

which implies $\overline{BD}(A) \geq \alpha$ by Part 2 of Proposition 3.6.

Now we prove the first part. Assume $\overline{BD}(A) = \alpha$. By Part 2 of Proposition 3.6 there is a $k \in {}^*\mathbb{N}$ and hyperfinite integer N such that $|{}^*A \cap (k + [N])|/N \approx \alpha$. Let $\Omega := k + [N]$, $(\Omega; \Sigma, \mu_\Omega)$ be the Loeb space, $B := {}^*A \cap \Omega$, and $f : \Omega \to \mathbb{R}$ be the characteristic function of B. Then, $f \in L_1(\Omega)$, i.e., f is integrable. Let $T(n) = n + 1$ for all $n \in \Omega, n \neq k + N - 1$ and $T(k + N - 1) = k$. Then T is a measure-preserving transformation on Ω. By Theorem 3.9 there is a $\bar{f} \in L_1(\Omega)$ such that there is a $X \subseteq \Omega$ with $\mu_\Omega(X) = 1$ such that for all $n \in X$ we have

$$\begin{aligned} \bar{f}(n) &= \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} f(T^i(n)) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} f(n + i) \\ &= \lim_{m\to\infty} \frac{|B \cap (n + [m])|}{m} = d(({}^*A - n) \cap \mathbb{N}). \end{aligned}$$

Since $\bar{f}(n) > \alpha$ implies $\underline{d}(({}^*A - n) \cap \mathbb{N}) > \alpha$ which implies $\overline{BD}(A) > \alpha$ by the first part, we have that $\bar{f}(n) \leq \alpha$ for all $n \in \Omega$. Since

$$\int_\Omega \bar{f} d\mu_\Omega = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} \int_\Omega f(T^i(n)) d\mu_\Omega = \int_\Omega f d\mu_\Omega = \mu_\Omega(B) = \alpha,$$

we conclude that $\bar{f}(n) = \alpha$ for μ_Ω-almost all $n \in \Omega$. Hence, $\underline{d}(({}^*A - n) \cap \mathbb{N}) = d(({}^*A - n) \cap \mathbb{N}) = \alpha$ for μ_Ω-almost all $n \in \Omega$. □

Proof of Theorem 3.8: By Theorem 3.7 we can find $k \in {}^*\mathbb{N}$ such that $\underline{d}(({}^*A - k) \cap \mathbb{N}) = \alpha$. For each $m \in \mathbb{N}$ let $n_m := \max\{n \in \mathbb{N} \mid |({}^*A - k) \cap [n_m] \leq \alpha - 1/m\}$. Note that n_m exists because otherwise we would have $\underline{d}(({}^*A - k) \cap \mathbb{N}) \leq \alpha - 1/m$. Note that $|({}^*A - k - n_m) \cap [n]|/n > \alpha - 1/m$ for any $n \in 1 + [m]$. By Proposition 2.8 we can find a hyperfinite integer N such that $|({}^*A - k - n_N) \cap [n]|/n > \alpha - 1/N$ for any $n \in 1 + [N]$. This implies that $\sigma(({}^*A - k - n_N) \cap \mathbb{N}) \geq \alpha$. Since $\sigma(({}^*A - k - n_N) \cap \mathbb{N}) > \alpha$ implies $\underline{d}(({}^*A - k - n_N) \cap \mathbb{N}) > \alpha$ which is impossible by Theorem 3.7 we conclude that $\sigma(({}^*A - k - n_N) \cap \mathbb{N}) = \alpha$. □.

Theorem 3.10 (Mann's Theorem, Theorem 3 in [9]) *Let $A, B \subseteq \mathbb{N}$ and $0 \in A \cap B$. Then*

$$\sigma(A+B) \geq \min\{\sigma(A)+\sigma(B), 1\}.$$

Theorem 3.11 (Upper Banach Density Version, Theorem 2 in [12])
Let $A, B \subseteq \mathbb{N}$. Then

$$\overline{BD}(A+B+\{0,1\}) \geq \min\{\overline{BD}(A)+\overline{BD}(B), 1\}.$$

Definition 3.12 *Let $B \subseteq \mathbb{N}$. For a positive integer $h \in \mathbb{N}$, let*

$$hB := \{b_1 + b_2 + \cdots + b_h \mid b_i \in B \ \textit{for} \ i = 1, 2, \ldots, h\}.$$

1. *The set B is a basis if $hB = \mathbb{N}$ for some $h \in \mathbb{N}$. The least such h is called the order of B. Clearly, a basis must contain 0;*

2. *Suppose B is a basis of order h. For each $m \geq 1$ let $h(m) := \min\{h' \in \mathbb{N} \mid m \in h'B\}$. Then, the number*

$$h^* := \sup_{n \geq 1} \frac{1}{n} \sum_{m=1}^{n} h(m)$$

is called the average order of B. Note that $h^* \leq h$;

3. *The set B is an asymptotic basis if $\mathbb{N} \setminus h_a B$ is finite for some $h_a \in \mathbb{N}$. The least such h_a is called the asymptotic order of B;*

4. *Suppose B is an asymptotic basis of order $h_a \in \mathbb{N}$ and $\mathbb{N} \setminus [n_0] \subseteq h_a B$ for some minimal $n_0 \in \mathbb{N}$. For each $m \geq n_0$ let $h(m) := \min\{h' \in \mathbb{N} \mid m \in h'B\}$. Then, the number*

$$h_a^* := \limsup_{n \to \infty} \frac{1}{n} \sum_{m=n_0}^{n_0+n-1} h(m)$$

is called the asymptotic average order of B. Note that $h_a^* \leq h_a$;

5. *The set B is a piecewise basis if there exists some $h_p \in \mathbb{N}$ such that one can find a sequence $k_n + [m_n]$ with $m_n \to \infty$ as $n \to \infty$ satisfying*

$$k_n + ([m_n]) \subseteq h_p((B - k_n) \cap \mathbb{N}) + k_n$$

for every $n \in \mathbb{N}$. The least such h_p is called the piecewise order of B;

6. *The set B is a piecewise asymptotic basis if there is an $h_{pa} \in \mathbb{N}$ such that one can find a sequence $k_n + [m_n]$ with $m_n \to \infty$ as $n \to \infty$ and a number $n_0 \in \mathbb{N}$ satisfying*

$$k_n + ([m_n] \setminus [n_0]) \subseteq h_{pa}((B - k_n) \cap \mathbb{N}) + k_n$$

for every $n \in \mathbb{N}$. The least such h_{pa} is called the piecewise asymptotic order of B;

7. *Suppose that B is a piecewise asymptotic basis of piecewise asymptotic order h_{pa}. Let $\mathcal{I}$ be the sequence $k_n + [m_n]$ and $n_0 \in \mathbb{N}$ such that $k_n + ([m_n] \setminus [n_0]) \subseteq h_{pa}((B - k_n) \cap \mathbb{N}) + k_n$ for every $n \in \mathbb{N}$. For each $m \in k_n + ([m_n] \setminus [n_0])$ let $h(m) := \min\{h' \in \mathbb{N} \mid m \in h'((B - k_n) \cap \mathbb{N}) + k_n$. Let*

$$h_n^* := \frac{1}{m_n - n_0} \sum_{i=k_n+n_0}^{k_n+m_n-1} h(m) \text{ and}$$

$$h_{\mathcal{I}}^* := \limsup_{n \to \infty} h_n^*.$$

Then, the number

$$h_{pa}^* := \inf\{h_{\mathcal{I}}^* \mid \text{ for all suitable } \mathcal{I}\}$$

is called a piecewise asymptotic average order of B.

Theorem 3.13 (Rohrback's Theorem, Theorem 13 in [9]) *If B is an asymptotic basis of asymptotic average order h_a^*, then for any $A \subseteq \mathbb{N}$ we have*

$$\underline{d}(A + B) \geq \underline{d}(A) + \frac{1}{2h_a^*}\underline{d}(A)(1 - \underline{d}(A)).$$

Theorem 3.14 (Upper Banach Density Version, Theorem 4 in [12]) *If B is a piecewise asymptotic basis of piecewise asymptotic average order h_{pa}^*, then for any $A \subseteq \mathbb{N}$ we have*

$$\overline{BD}(A + B) \geq \overline{BD}(A) + \frac{1}{2h_{pa}^*}\overline{BD}(A)(1 - \overline{BD}(A)).$$

3.3 Plünnecke's Inequalities

Theorem 3.13 is a generalization of Erdős' theorem (cf. [9, Theorem 5]) that if B is a basis of order h, then for any $A \subseteq \mathbb{N}$ it is true that

$$\sigma(A + B) \geq \sigma(A) + \frac{1}{2h}\sigma(A)(1 - \sigma(A)).$$

Erdős' theorem is for the study of so-called essential component problems. A set B is called essential component if $\sigma(A+B) > \sigma(A)$ for any $A \subseteq \mathbb{N}$ with $0 < \sigma(A) < 1$. Hence, a basis must be an essential component.

There is another generalization of Erdős' theorem, which is much more significant than Theorem 3.13 does. The following generalization of Erdős' theorem used a completely different idea from Erdős'.

Theorem 3.15 (Plünnecke's Theorem, Theorem 7.10 in [21]) *Let B be a basis of order h. Then, for any $A \subseteq \mathbb{N}$ we have*

$$\sigma(A+B) \geq \sigma(A)^{1-\frac{1}{h}}.$$

It is not too hard to show that $\sigma(A)^{1-\frac{1}{h}} \geq \sigma(A) + \dfrac{1}{h}\sigma(A)(1-\sigma(A))$ (cf. [21, Corollary 7.2]).

The key component used in the proof of Theorem 3.15 is a version of Plünnecke's inequality based on graph theoretic argument. The following lemma is a translation of [21, Theorem 7.4] from the language of graph theory to the language of additive number theory.

Lemma 3.16 (Plünnecke's Inequality, Theorem 7.4 in [21])
Let $A, B \subseteq \mathbb{N}$ and $h, n \geq 1$ be such that $A \cap [n] \neq \emptyset$. For each $1 \leq i \leq h$ define

$$D_{A,B,n,i} = \min\left\{\frac{|(A'+iB)\cap[n]|}{|A'\cap[n]|} : \emptyset \neq A' \subseteq A \cap [n]\right\}.$$

Then

$$D_{A,B,n,1} \geq (D_{A,B,n,2})^{1/2} \geq \cdots \geq (D_{A,B,n,h})^{1/h}.$$

Many interesting subsets of $\mathbb{N}$ are not bases but asymptotic bases. For example, $P := \{p \in \mathbb{N} \mid p \text{ is a prime number}\}$, $C_k := \{n^k \mid n \in \mathbb{N}\}$ for $k \geq 1$, $P^2 := \{a^2b^3 \mid a, b \in \mathbb{N} \text{ and } a, b \geq 1\}$, etc. are asymptotic bases. Therefore, it is interesting to see whether Plünnecke's Theorem can be generalized to some versions involving other densities.

Definition 3.17 *Let $B \subseteq \mathbb{N}$.*

1. *The set B is a lower asymptotic basis of order $h \in \mathbb{N}$ if $\underline{d}(hB) = 1$;*

2. *The set B is an upper asymptotic basis of order $h \in \mathbb{N}$ if $\overline{d}(hB) = 1$;*

3. *The set B is an upper Banach basis of order $h \in \mathbb{N}$ if $\overline{BD}(hB) = 1$;*

4. *The set B is a lower Banach basis of order $h \in \mathbb{N}$ if $\underline{BD}(hB) = 1$.*

Note that P is an asymptotic basis of order 4 by Vinogradov's Theorem, or 3 if Goldbach conjecture is true. It is also known that P is a lower asymptotic basis of order 3. P^2 is an asymptotic basis of order 3 by a result of Heath-Brown (cf. [10]). C_2 is a basis of order 4 and C_3 is an asymptotic basis of order at most 7 (cf. [20]). Note also that P, C_k, P^2 are all have upper density 0.

Theorem 3.18 (Theorem 1.5 in [14]) *Let $A, B \subseteq \mathbb{N}$ and B be a lower asymptotic basis of order h. Then*

$$\underline{d}(A + B) \geqslant \underline{d}(A)^{1-\frac{1}{h}}.$$

Corollary 3.19 *For any $A \subseteq \mathbb{N}$ we have*

1. $\underline{d}(A + P) \geq \underline{d}(A)^{2/3}$;
2. $\underline{d}(A + C_2) \geq \underline{d}(A)^{3/4}$;
3. $\underline{d}(A + C_3) \geq \underline{d}(A)^{6/7}$;
4. $\underline{d}(A + P^2) \geq \underline{d}(A)^{2/3}$.

Theorem 3.20 (Theorem 1.6 in [14]) *There are $A, B \subseteq \mathbb{N}$ with $\overline{d}(A) = \frac{1}{2}$, $\overline{d}(2B) = 1$, and*

$$\overline{d}(A + B) = \overline{d}(A).$$

Theorem 3.21 (Theorem 1.7 in [14]) *Let $A, B \subseteq \mathbb{N}$ and B be a upper Banach basis of order h. Then*

$$\overline{BD}(A + B) \geq \overline{BD}(A)^{1-\frac{1}{h}}.$$

Theorem 3.22 (Theorem 7 in [15]) *Let $A, B \subseteq \mathbb{N}$ and B be an upper Banach basis of order h. Then,*

$$\underline{BD}(A + B) \geq \underline{BD}(A)^{1-\frac{1}{h}}.$$

Note that Theorem 3.18 and Theorem 3.20 show that lower density and upper density are asymmetrical on generalizing Plünnecke's Theorem. Theorem 3.21 and Theorem 3.22 look like following the same pattern but they show also that upper Banach density and lower Banach density are mildly asymmetrical. Both of the theorems require B be upper Banach basis.

We will prove Theorem 3.18 and Theorem 3.21. The reader can find the proofs of the other two theorems in [14, 15]. The arguments used in the proof of Theorem 3.15

deal with finite intervals of integers and are purely combinatorial. It becomes messy when the limit processes for $\underline{d}$ or $\overline{BD}$ are involved. Using nonstandard analysis, we can transfer the limit processes to combinatorial arguments on intervals of hyperfinite length, which simplify the proofs.

Proof of Theorem 3.18: Let A and B be in Theorem 3.18 such that $\underline{d}(A) = \alpha$ and $\underline{d}(hB) = 1$. Without loss of generality, we can assume $0 < \alpha < 1$. Let N be any hyperfinite integer. We want to show that

$$\frac{|{}^*(A+B)\cap[N]|}{N} = \frac{|({}^*A + {}^*B)\cap[N]}{N} \gtrapprox \alpha^{1-\frac{1}{h}},$$

which implies Theorem 3.18 by Proposition 3.5. Choose hyperfinite integers $N' < K < N$ such that $(N-K)/N \approx 0$ and $(K-N')/(N-N') \approx 0$ (for example $K = N - \left\lfloor \sqrt{N} \right\rfloor$ and $N' = K - \left\lfloor \sqrt[4]{N} \right\rfloor$ satisfy the requirements). Let $C_0 = {}^*A \cap [K]$. Then $(|C_0 \cap [N]|)/N \gtrapprox \alpha$. Next we want to trim C_0 so that the density of the trimmed set in each interval $\{x, x+1, \ldots, N-1\}$ for every $x \leq K$ would not be too large. We define C_k inductively for $k = 0, 1, \ldots, N'-1$ so that

$$C_0 \supseteq C_1 \supseteq \cdots \supseteq C_{N'-1},\ \frac{|C_{N'-1}\cap[N]|}{N} \approx \alpha, \text{ and}$$
$$\frac{|C_{N'-1}\cap\{x, x+1, \ldots, N-1\}|}{N-x} \lessapprox \alpha$$

for any $x \leq K$. Start with C_0. For each $k < N'-1$ let

$$C_{k+1} = \begin{cases} C_k, & \text{if } \frac{|C_k\cap\{N'-k, N'-k+1, \ldots, N-1\}|}{N-N'+k} \leq \alpha \\ C_k \setminus \{N'-k\}, & \text{otherwise.} \end{cases}$$

It is easy to see that $C_0, C_1, \ldots, C_{N'-1}$ has the desired properties. Let $A_0 = C_{N'-1}$ and nonempty $A' \subseteq A_0$ be such that

$$D_{A_0, {}^*B, N, h} = \frac{|(A' + h\,{}^*B)\cap[N]|)}{|A'\cap[N]|}.$$

Let $z = \min A'$. Then $z < K$ because $A_0 \subseteq [K]$. Hence $N - z$ is hyperfinite, which implies $\dfrac{|(h\,{}^*B)\cap[N-z]|}{N-z} \approx 1$. By Lemma 3.16 we have

$$\frac{|(A_0 + {}^*B)\cap[N]|}{|A_0\cap[N]|}$$
$$\geq D_{A_0, {}^*B, N, 1} \geq (D_{A_0, {}^*B, N, h})^{1/h} = \left(\frac{|(A' + h\,{}^*B)\cap[N]}{|A'\cap[N]|}\right)^{1/h}$$

$$\gtrapprox \left(\frac{|(z+h\,{}^*B)\cap[N]|}{|A'\cap[N]|)}\right)^{1/h} \gtrapprox \left(\frac{|(h\,{}^*B)\cap[N-z]|/(N-z)}{|A'\cap\{z,z+1,\dots,N-1\}|/(N-z)}\right)^{1/h}$$
$$\gtrapprox \left(\frac{1}{|A_0\cap\{z,z+1,\dots,N-1\}|/(N-z)}\right)^{1/h} \gtrapprox \frac{1}{\alpha^{1/h}},$$

which implies

$$\frac{|{}^*(A+B)\cap[N]|}{N} \geq \frac{|(A_0+{}^*B)\cap[N]|}{N} \gtrapprox \frac{|A_0\cap[N]|}{N}\cdot\frac{1}{\alpha^{1/h}} \gtrapprox \alpha^{1-\frac{1}{h}}.$$

Since N is an arbitrary hyperfinite integer, Theorem 3.18 is proven with the help of Proposition 3.5. □

Proof of Theorem 3.21: Let A and B be in Theorem 3.21 with $\overline{BD}(A)=\alpha$ and $\overline{BD}(hB)=1$ for some $h\in\mathbb{N}$. Theorem 3.21 is trivially true if $\overline{BD}(A)=0$ or $\overline{BD}(A)=1$. So, we can assume that $0<\alpha=\overline{BD}(A)<1$. Let $n\in{}^*\mathbb{N}$ and K be a hyperfinite integer such that $n+[K]\subseteq(h\,{}^*B)$. Choose N large enough so that $(n+K)/N\approx 0$ and $|{}^*A\cap(m+[N])|/N\approx\alpha$ for some $m\in{}^*\mathbb{N}$. It suffices to show that

$$\frac{|({}^*A\cap(m+[N])+{}^*B)\cap(m+[N])|}{N} \gtrapprox \alpha^{1-\frac{1}{h}}$$

by Proposition 3.6. Let $A_0=({}^*A\cap(m+[N-n-K])-m$. By the choice of N and A_0 we have

$$\frac{|A_0\cap[N]|}{N}\approx\alpha \text{ and } \frac{|(A_0+{}^*B)\cap[N]|}{N} \lessapprox \frac{|({}^*A+{}^*B)\cap(m+[N])|}{N}.$$

It now suffices to show that

$$\frac{|(A_0+{}^*B)\cap[N]|}{N} \gtrapprox \alpha^{1-\frac{1}{h}}.$$

Let $A'\subseteq A_0$ be nonempty such that $D_{A_0,{}^*B,N,h}=|(A'+h\,{}^*B)\cap[N]|/|A'\cap[N]|$.

Claim:

$$\frac{|(A'+h\,{}^*B)\cap[N]|}{|A'\cap[N]|} = D_{A_0,{}^*B,H,h} \gtrapprox \frac{1}{\alpha}.$$

Proof of Claim: Let $H=\lfloor K/2\rfloor$ and let $I_i=iH+[H]$ for $i=0,1,\dots\lfloor N/H\rfloor-1$, and let $I_{\lfloor N/H\rfloor}=\lfloor N/H\rfloor\cdot H+[N-\lfloor N/H\rfloor\cdot H]$. Denote

$$\mathcal{I}:=\{I_i \mid i\in[\lfloor N/H\rfloor+1] \text{ and } I_i\cap A'\neq\emptyset\}.$$

Then

$$|(A'+h\,{}^*B)\cap[N]| \geqslant |\mathcal{I}|\cdot H$$

because $H \leq K/2$, every element in A' is less than or equal to $N - n - K$, and $H + n + I_i \subseteq (A' + h\,^*B) \cap [N]$ if $A' \cap I_i \neq \emptyset$ for every $i = 0, 1, \ldots, \lfloor H/N \rfloor$. Given a positive standard real ϵ, we have

$$|A' \cap [N]| \leqslant |\mathcal{I}| \cdot (\alpha + \epsilon)H$$

because $|A' \cap I_i|/|I_i| \lessapprox \alpha$ when $|I_i|$ is hyperfinite by Proposition 3.6. Because ϵ is an arbitrary standard positive real number, we have that

$$\frac{|(A' + h\,^*B) \cap [N]|}{|A' \cap [N]|} \gtrapprox \frac{|\mathcal{I}| \cdot H}{|\mathcal{I}| \cdot \alpha H} = \frac{1}{\alpha}.$$

This completes the proof of the claim.

We continue to prove Theorem 3.21. Combine the arguments above and Theorem 3.16 we now have

$$\begin{aligned} \frac{|(A_0 + {}^*B) \cap [N]|}{|A_0 \cap [N]|} &\gtrapprox D_{A_0, {}^*B, N, 1} \geq (D_{A_0, {}^*B, N, h})^{1/h} \\ &= \left(\frac{|(A' + h\,^*B) \cap [N]|}{|A' \cap [N]|} \right)^{1/h} \gtrapprox \frac{1}{\alpha^{1/h}}. \end{aligned}$$

Hence

$$\frac{|{}^*(A + B) \cap [N]|}{N} \gtrapprox \frac{|(A_0 + {}^*B) \cap [N]|}{N} \gtrapprox \frac{|A_0 \cap [N]|}{N} \cdot \frac{1}{\alpha^{1/h}} \approx \alpha^{1 - \frac{1}{h}},$$

which implies Theorem 3.21 by Proposition 3.6. □

3.4 Exercises

1. Let $A \subseteq \mathbb{N}$. Prove that the limit of the sequence

$$s_n := \sup_{k \in \mathbb{N}} \frac{|A \cap (k + [n])|}{n}$$

as $n \to \infty$ exists.

2. Prove Lemma 3.4.

3. Prove Part 2 of Proposition 3.5.

4. Prove that Theorem 3.14 using Theorem 3.13 and By-one-get-one-free Thesis.

4 Hard Applications to Combinatorics

There have been many recent applications of nonstandard analysis to Ramsey type problems in combinatorial number theory (cf. [4, 5, 6, 16].) One of the characteristics of these new applications is the use of multiple levels of infinities. We will first construct nonstandard universes with multiple levels of infinities and then solve some combinatorial problems in these nonstandard universes.

4.1 Multiple Levels of Infinities and Ramsey's Theorem

Our first goal in this subsection is to construct a sequence of nonstandard universes and two types of correspondent elementary embeddings satisfying some nice properties.

Proposition 4.1 *There exists a sequence of nonstandard universes*

$$\mathcal{V}_0 = \mathcal{V} \prec \mathcal{V}_1 \prec \mathcal{V}_2 \prec \cdots \mathcal{V}_n \prec \cdots$$

and elementary embeddings

$$i_{m,n} : \mathcal{V}_n \to \mathcal{V}_{n+1}$$

for all $0 \le m \le n$ *in* $\mathbb{N}$ *such that*

1. $\mathbb{N}_0 := \mathbb{N}$ *and* $\mathbb{N}_{n+1} := i_{n,n}(\mathbb{N}_n) \supseteq i_{n,n}[\mathbb{N}_n] = \mathbb{N}_n$ *is an end-extension of* $\mathbb{N}_n$*, i.e., every number in* $\mathbb{N}_{n+1} \setminus \mathbb{N}_n$ *is greater than any number in* $\mathbb{N}_n$*, for* $n = 0, 1, \ldots$*;*
2. $i_{m,n}[\mathbb{N}_k \setminus \mathbb{N}_{k-1}] \subseteq \mathbb{N}_{k+1} \setminus \mathbb{N}_k$ *for* $k = m+1, m+2, \ldots, n$*;*
3. $i_{m,n}(x) = x$ *for every* $x \in \mathbb{N}_m$ *and* $i_{m,n} \restriction \mathcal{V}_k = i_{m,k}$ *for* $m \le k \le n$*;*
4. $i_{m,n} \restriction \mathcal{V}_k : (\mathcal{V}_k; \mathbb{R}_{k-l+1}, \mathbb{R}_{k-l}) \to (\mathcal{V}_{k+1}; \mathbb{R}_{k-l+2}, \mathbb{R}_{k+1-l})$ *is an elementary embedding where* $(\mathcal{V}_k; \mathbb{R}_{k-l+1}, \mathbb{R}_{k-l})$ *and* $(\mathcal{V}_{k+1}; \mathbb{R}_{k-l+2}, \mathbb{R}_{k+1-1})$ *represent the models* $\mathcal{V}_k$ *and* $\mathcal{V}_{k+1}$ *augmented by unary relations* $\mathbb{R}_{k+1-l}, \mathbb{R}_{k-l} \notin \mathcal{V}_k$ *and* $\mathbb{R}_{k-l+2}, \mathbb{R}_{k+1-l} \notin \mathcal{V}_k$*, respectively, for* $m \le k \le n$ *and* $2 \le l \le k - m$*;*

Recall that the ultrafilter $\mathcal{F}$ is fixed after Definition 1.6. Let $\mathcal{V}_0 := \mathcal{V}$, $\mathcal{F}_0 := \mathcal{F}$, $\mathcal{V}_1 := {}^*\mathcal{V}$ be the ultrapower of $\mathcal{V}_0$ modulo $\mathcal{F}_0$, and $i_{0,0} := {}^*$ be the elementary embedding from $\mathcal{V}_0$ to $\mathcal{V}_1$ constructed in Definition 1.21. Note that $\mathcal{F}_0 \in \mathcal{V}_0$.

Let $\mathcal{F}_1 := i_{0,0}(\mathcal{F}_0) \in \mathcal{V}_1$. By the transfer principle we have that $\mathcal{F}_1$ satisfies Parts 1 – 4 of Definition 1.6 for any $A, B \in \mathcal{V}_1$ with $X = \mathbb{N}_1 := i_{0,0}(\mathbb{N}_0)$ and co-finite is replaced by co-hyperfinite in $\mathcal{V}_1$. We call $\mathcal{F}_1$ a $\mathcal{V}_1$–internal non-principal ultrafilter on $\mathbb{N}_1$. Notice that $i_{0,0}(\mathscr{P}(\mathbb{N}_0)) = \mathcal{V}_1 \cap \mathscr{P}(\mathbb{N}_1)$ and

$$i_{0,0}(\mathscr{P}_{<\mathbb{N}_0}(\mathbb{N}_0)) = \mathcal{V}_1 \cap \mathscr{P}_{<\mathbb{N}_1}(\mathbb{N}_1) := \{A \subseteq \mathbb{N}_1 \mid A \in \mathcal{V}_1 \wedge \exists N \in \mathbb{N}_1 (A \subseteq [N])\}.$$

If an $\in$–formula φ is coded by a finite sequence of numbers in $\mathbb{N}_0$, then $i_{0,0}(\varphi) = \varphi$.

Without loss of generality we can identify $i_{0,0}[\mathcal{V}_0]$ with $\mathcal{V}_0$ so that $\mathcal{V}_0$ is an elementary submodel of $\mathcal{V}_1$.

Let $\mathcal{F}'_0 := \mathcal{F}_0$ and $\mathbb{N}'_0 := \mathbb{N}_0$. We use $'$ to indicate the different location where $\mathcal{F}_0$ and $\mathbb{N}_0$ are used. To form an ultrapower of $\mathcal{V}_1$ modulo $\mathcal{F}'_0$, we obtain an elementary extension

$$\mathcal{V}_2 := (V_1^{\mathbb{N}'_0}/\mathcal{F}'_0, {}^*\!\in) = \mathcal{V}_1^{\mathbb{N}'_0}/\mathcal{F}'_0 = \left(\mathcal{V}_0^{\mathbb{N}_0}/\mathcal{F}_0\right)^{\mathbb{N}'_0}/\mathcal{F}'_0 \tag{7}$$

and associated elementary embedding $i_{0,1} : \mathcal{V}_1 \to \mathcal{V}_2$ as we did in Definition 1.8 and Corollary 1.10. By applying Mostowski collapsing map again we can assume that ${}^*\!\in$ is the real membership relation $\in$ and $\mathbb{N}_1 \subseteq \mathbb{N}_2 := i_{0,1}(\mathbb{N}_1)$. Note that $\mathbb{N}_1$ and $i_{0,1}[\mathbb{N}_1]$ are not the same even after Mostowski collapsing. Let's call $\mathcal{V}_1^{\mathbb{N}'_0}/\mathcal{F}'_0$ the external ultrapower of $\mathcal{V}_1$ modulo $\mathcal{F}'_0$.

If $\mathbb{N}_1$ had been identified with $i_{0,1}[\mathbb{N}_1]$, then $\mathbb{N}_2$ won't be an end-extension of $\mathbb{N}_1$. Therefore, we should look at $\mathcal{V}_2$ from a different angle.

Definition 4.2 *The $\mathcal{V}_1$-internal ultrapower of $\mathcal{V}_1$ modulo $\mathcal{F}_1$ is the model with the base set $\mathcal{V}_1^{\mathbb{N}_1} \cap \mathcal{V}_1 := \{[f]_{\mathcal{F}_1} \mid f \in \mathcal{V}_1^{\mathbb{N}_1}$ and $f \in \mathcal{V}_1\}$, where*

$$f \sim_{\mathcal{F}_1} g \text{ iff } \{n \in \mathbb{N}_1 \mid f(n) = g(n)\} \in \mathcal{F}_1 \text{ and}$$

$$[f]_{\mathcal{F}_1} := \{g \in \mathcal{V}_1^{\mathbb{N}_1} \cap \mathcal{V}_1 \mid f \sim_{\mathcal{F}_1} g\},$$

and the membership relation $\in_2$ defined by

$$[f]_{\mathcal{F}_1} \in_2 [g]_{\mathcal{F}_1} \text{ iff } \{n \in \mathbb{N}_1 \mid f(n) \in g(n)\} \in \mathcal{F}_1.$$

The map $i_{1,1} : \mathcal{V}_1 \to (\mathcal{V}_1^{\mathbb{N}_1} \cap \mathcal{V}_1)/\mathcal{F}_1$ with $i_{1,1}(c) = [\phi_c]_{\mathcal{F}_1}$ is the elementary embedding from $\mathcal{V}_1$ to $(\mathcal{V}_1^{\mathbb{N}_1} \cap \mathcal{V}_1)/\mathcal{F}_1$ associated with the $\mathcal{V}_1$-internal ultrapower of $\mathcal{V}_1$ modulo the $\mathcal{V}_1$-internal ultrafilter $\mathcal{F}_1$.

By applying Mostowski collapsing map again we can assume that $\in_2$ is $\in$. An element $a \in \mathcal{V}_2$ is called $\mathcal{V}_2$-internal. An element $a \in \mathcal{V}_2$ is called $\mathcal{V}_1$-internal if $a \in i_{1,1}[\mathcal{V}_1]$. Recall that $((\mathcal{V}_1^{\mathbb{N}_1} \cap \mathcal{V}_1)/\mathcal{F}_1; \in)$ is the $\mathcal{V}_1$-internal ultrapower of $\mathcal{V}_1$ modulo $\mathcal{F}_1$.

Note that the $\mathcal{V}_1$-internal ultrapower of $\mathcal{V}_1$ modulo $\mathcal{F}_1$ is really the same as the external ultrapower of $\mathcal{V}_1$ modulo $\mathcal{F}'_0$. Indeed, we can make two-step ultrapower process in two different order. In the external ultrapower of $\mathcal{V}_1$ modulo $\mathcal{F}'_0$ we view the ultrapower modulo $\mathcal{F}_0$ to get $\mathcal{V}_1$ first and the ultrapower of $\mathcal{V}_1$ modulo $\mathcal{F}'_0$ the second. If we view the two-step ultrapower process by taking the ultrapower modulo $\mathcal{F}'_0$ first, $\mathbb{N}_0$ and $\mathcal{F}_0$ in $\mathcal{V}_0$ become $\mathbb{N}_1$ and $\mathcal{F}_1$, respectively, and $\mathcal{V}_0^{\mathbb{N}_0}$ because the collection $\mathcal{V}_1^{\mathbb{N}_1} \cap \mathcal{V}_1$ of all $\mathcal{V}_1$-internal functions from $\mathbb{N}_1$ to $\mathcal{V}_1$. Hence, the process of taking ultrapower of $\mathcal{V}_0$ modulo $\mathcal{F}_0$ is lifted into $\mathcal{V}_1$ to become the $\mathcal{V}_1$-internal ultrapower of $\mathcal{V}_1$ modulo $\mathcal{F}_1$ to complete the second step. Symbolically, we have

$$\begin{aligned}\mathcal{V}_2 &= \left(\mathcal{V}_0^{\mathbb{N}_0}/\mathcal{F}_0\right)^{\mathbb{N}_0'}/\mathcal{F}_0' = (\mathcal{V}_1^{\mathbb{N}_1} \cap \mathcal{V}_1)/\mathcal{F}_1 \\ &= ((\mathcal{V}_0^{\mathbb{N}_0'}/\mathcal{F}_0')^{\mathbb{N}_0^{\mathbb{N}_0'}/\mathcal{F}_0} \cap (\mathcal{V}_0^{\mathbb{N}_0'}/\mathcal{F}_0'))/(\mathcal{F}_0^{\mathbb{N}_0'}/\mathcal{F}_0').\end{aligned} \tag{8}$$

Roughly speaking, (8) shows that one can change the order of ultrapower of $\mathcal{V}_0$ construction steps first modulo $\mathcal{F}_0$ and then modulo $\mathcal{F}_0'$ to the order that first modulo $\mathcal{F}_0'$ and then modulo $\mathcal{F}_1 = i_{0,0}(\mathcal{F}_0)$.

By applying the transfer principle to the statement that every bounded function from $\mathbb{N}_0$ to $\mathbb{N}_0$ is equivalent, modulo $\mathcal{F}_0$, to a constant function, we have that every bounded $\mathcal{V}_1$-internal function from $\mathbb{N}_1$ to $\mathbb{N}_1$ is equivalent, modulo $\mathcal{F}_1$, to a constant function. So, if $[f]_{\mathcal{F}_1} \in \mathbb{N}_2$ and $f(n) \leq m \in \mathbb{N}_1$ for every $n \in \mathbb{N}_1$, then f is equivalent, modulo $\mathcal{F}_1$, to $[\phi_c]_{\mathcal{F}_1}$ for some $c \in \mathbb{N}_1$, which implies $[f]_{\mathcal{F}_1} \in \mathbb{N}_1$. So, $\mathbb{N}_2 := i_{1,1}(\mathbb{N}_1) \supseteq i_{1,1}[\mathbb{N}_1] = \mathbb{N}_1$ is an end-extension of $\mathbb{N}_1$. Note that $i_{0,1} \restriction \mathbb{N}_0 = i_{1,1} \restriction \mathbb{N}_0 = i_{0,0}$. If $\mathcal{V}_2$ is considered as the external ultrapower of $\mathcal{V}_1$, then $\mathbb{N}_1$ can be identified as $\mathbb{N}_0^{\mathbb{N}_0'}/\mathcal{F}_0'$ in (7).

It is easy to check that the elementary embeddings $i_{0,0}, i_{0,1}, i_{1,1}$ satisfy Proposition 4.1 except Part 4, which is irrelevant.

In fact, $\mathcal{V}_2$ can be viewed as one-step ultrapower of $\mathcal{V}_0$ modulo the tensor product of $\mathcal{F}_0$ and $\mathcal{F}_0'$ (cf. [1, Proposition 6.5.2]) where

$$\mathcal{F}_0 \otimes \mathcal{F}_0' := \{A \subseteq \mathbb{N}_0 \times \mathbb{N}_0' \mid \{n' \in \mathbb{N}_0' \mid \{n \in \mathbb{N}_0 \mid (n, n') \in A\} \in \mathcal{F}_0\} \in \mathcal{F}_0'\}$$

is a non-principle ultrafilter on $\mathbb{N}_0 \times \mathbb{N}_0'$. This indicates that $\mathcal{V}_2$ is countably saturated and elements in $\mathcal{V}_2$ can be represented by the equivalence class, modulo $\mathcal{F}_0 \otimes \mathcal{F}_0'$, of functions $f : \mathbb{N}_0 \times \mathbb{N}_0' \to \mathcal{V}_0$.

Now consider a three-step ultrapower construction. Let $\mathcal{F}_0'' := \mathcal{F}_0$, $\mathbb{N}_0'' := \mathbb{N}_0$, and $\mathcal{F}_2 := i_{1,1}(\mathcal{F}_1) \in \mathcal{V}_2$. Then

$$\mathcal{V}_3 = ((\mathcal{V}_0^{\mathbb{N}_0}/\mathcal{F}_0)^{\mathbb{N}_0'}/\mathcal{F}_0')^{\mathbb{N}_0''}/\mathcal{F}_0'' = \mathcal{V}_2^{\mathbb{N}_0''}/\mathcal{F}_0'' \tag{9}$$

$$= ((\mathcal{V}_1^{\mathbb{N}_1} \cap \mathcal{V}_1)/\mathcal{F}_1)^{\mathbb{N}_1'} \cap \mathcal{V}_1)/\mathcal{F}_1' = (\mathcal{V}_2^{\mathbb{N}_1'} \cap \mathcal{V}_1)/\mathcal{F}_1' \tag{10}$$

$$= ((\mathcal{V}_1^{\mathbb{N}_1} \cap \mathcal{V}_1)/\mathcal{F}_1)^{\mathbb{N}_1'} \cap \mathcal{V}_1)/\mathcal{F}_1' = (\mathcal{V}_2^{\mathbb{N}_2} \cap \mathcal{V}_2)/\mathcal{F}_2. \tag{11}$$

The ultrapower in (9) results in the associated elementary embedding $i_{0,2} : \mathcal{V}_2 \to \mathcal{V}_3$. The ultrapower in (10) results in the associated elementary embedding $i_{1,2} : \mathcal{V}_2 \to \mathcal{V}_3$. And the ultrapower in (11) results in the associated elementary embedding $i_{2,2} : \mathcal{V}_2 \to \mathcal{V}_3$. After applying Mostowski collapsing map we can again assume that

$\mathbb{N}_3 := i_{1,2}(\mathbb{N}_2) \supseteq \mathbb{N}_2 = i_{1,2}[\mathbb{N}_2]$ and $\mathbb{N}_3$ is an end-extension of $\mathbb{N}_2$. We can also assume that $\mathcal{V}_2 \subseteq \mathcal{V}_3$ via $i_{2,2}$. It is also easy to check that $i_{0,2} \upharpoonright \mathcal{V}_1 = i_{0,1}$ and $i_{0,2} \upharpoonright \mathcal{V}_0 = i_{0,0}$. Similarly, we have $i_{1,2} \upharpoonright \mathcal{V}_1 = i_{1,1}$. Note that Part 4 in Proposition 4.1 follows from the fact that $(\mathcal{V}_3; \mathbb{R}_2, \mathbb{R}_1)$ is the ultrapower of $(\mathcal{V}_2; \mathbb{R}_1, \mathbb{R}_0)$ modulo $\mathcal{F}'_0$. Hence, $i_{0,2}$ is an elementary embedding from $(\mathcal{V}_2; \mathbb{R}_1, \mathbb{R}_0)$ to $(\mathcal{V}_3; \mathbb{R}_2, \mathbb{R}_1)$.

The validity of the remaining properties in Proposition 4.1 for $i_{m,2}$ with $m = 0, 1, 2$ is left for the reader to check.

In general, we can use the same idea to iterate the ultrapower construction. Given $0 \leq m \leq n$, if we iterate the ultrapower construction m times internally (cf. (11)) followed by iterating ultrapower construction $n - m$ times within $\mathcal{V}_m$ "externally" (by viewing $\mathcal{V}_m$ as the "standard universe") we obtain the elementary embedding $i_{m,n} : \mathcal{V}_n \to \mathcal{V}_{n+1}$. These $i_{m,n}$'s satisfy the four parts in Proposition 4.1. For more detailed discussion of iterating ultrapowers the reader may consult [16, §2].

The second gaol of this subsection is to present a probably the simplest proof of Ramsey's Theorem as a testing case for working within a nonstandard universe such as $\mathcal{V}_n$. In the remaining part of this subsection let $[X]^k_* := \{S \subseteq X \mid |S| = k\}$ for any set X and $k \in \mathbb{N}_0$. A coloring of a set Y with r colors is a function $c : Y \to [r]$. A set $Z \subseteq Y$ is monochromatic (with respect to c) if $c \upharpoonright Z$ is a constant function.

Theorem 4.3 (Ramsey's Theorem) *Let $k, r \in \mathbb{N}_0$. If $c : [\mathbb{N}_0]^k_* \to [r]$ is a coloring of $[\mathbb{N}_0]^k_*$ with at most r colors, then there exists an infinite set $\mathbb{H} \subseteq \mathbb{N}_0$ such that $[\mathbb{H}]^k_*$ is monochromatic.*

Proof: Work within $\mathcal{V}_k$. Let $x_1 = [Id_{\mathbb{N}_0}]_{\mathcal{F}_0} \in \mathbb{N}_1 \setminus \mathbb{N}_0$ and $x_{j+1} := i_{0,k-1}(x_j)$ for $j = 1, 2, \ldots, k-1$. Then $\overline{x} = \{x_1, x_2, \ldots, x_r\} \in [\mathbb{N}_k]^k_*$. Note that x_j is the equivalence class represented by the identity map $Id_{\mathbb{N}_{j-1}} : \mathbb{N}_{j-1} \to \mathbb{N}_{j-1}$.

For convenience we denote still c for the extension of c from $[\mathbb{N}_j]^k_*$ to $[r]$ in $\mathcal{V}_j$. Let $c(\overline{x}) = c_0$. We construct a sequence $A = \{a_0 < a_1 < \cdots\} \subseteq \mathbb{N}_0$ inductively such that $c \upharpoonright [A \cup \overline{x}]^k_* \equiv c_0$.

Suppose that $A_m := \{a_0, \ldots, a_{m-1}\}$ has been found that $c \upharpoonright [A_m \cup \overline{x}]^k_* \equiv c_0$. Note that the sentence

$$\exists y \in \mathbb{N}_1 \, (y > a_{m-1} \text{ and } c \upharpoonright [A_m \cup \{y\} \cup \{i_{0,k-1}(x_1), \ldots, i_{0,k-1}(x_{k-1})\}]^k_* \equiv c_0)$$

is true in $(\mathcal{V}_k; \mathbb{R}_1)$ where y is witnessed by x_1. Hence,

$$\exists y \in \mathbb{N}_0 \, (y > a_{m-1} \text{ and } c \upharpoonright [A_m \cup \{y\} \cup \{x_1, \ldots, x_{k-1}\}]^k_* \equiv c_0) \tag{12}$$

is true in $(\mathcal{V}_{k-1}; \mathbb{R}_0)$ by Part 4 of Proposition 4.1. Let $y = a_m \in \mathbb{N}_0$ be the witness of the truth of (12) in $\mathcal{V}_{k-1}$ and $A_{m+1} = A_m \cup \{a_m\}$. It suffices to show the following claim.

Claim: $c \restriction [A_{m+1} \cup \overline{x}]^k_* \equiv c_0$.

Proof of Claim: Let $\overline{b} = \{b_1 < b_2 < \cdots < b_k\} \in [A_{m+1} \cup \overline{x}]^k_*$. We show that $c(\overline{b}) = c_0$.

If $b_k < x_k$, then $c(\overline{b}) = c_0$ by (12). If $b_1 = x_1$, then $c(\overline{b}) = c(\overline{x}) = c_0$. So, we can assume that $b_1 \in \mathbb{N}_0$ and $b_k = x_k$. Let $p = \max\{j \in 1 + [k] \mid x_j \notin \overline{b}\}$. Then $p < k$, $b_p = x_{j'}$ for some $1 \leq j' < p$ or $b_p \in \mathbb{N}_0$, and $b_j = x_j$ for $j = p+1, \ldots, k$. Let $p' := 0$ if $b_p \in \mathbb{N}_0$ or $p' = j'$ if $b_p = x_{j'}$ for some $1 \leq j' \leq p - 1$. Note that $i_{p',k-1}(b_j) = b_j$ for $j \leq p$. Note also that $i_{p',k-1}(x_{j-1}) = i_{0,k-1}(x_{j-1}) = b_j$ for $j = p+1, \ldots, k$ because $i_{p',k-1}(x_{j-1})$ is an equivalence class represented by $Id_{\mathbb{N}_{j-1}}$. So, $i^{-1}_{p',k-1}(\overline{b}) \in [A_{m+1} \cup \{x_1, \ldots, x_{k-1}\}]^k_*$ and hence, $c(i^{-1}_{p',k-1}(\overline{b})) = c_0$. By the transfer principle for $i_{p',k-1}$ we have $c(\overline{b}) = c_0$. This completes the proof of the claim as well as the theorem. $\square$

4.2 Multidimensional van der Waerden's Theorem

The multidimensional van der Waerden's Theorem is also called Gallai's Theorem. Fix a dimension s and let $[n]^s = \{(x_1, x_2, \ldots, x_s) \mid x_j \in [n] \text{ for } j = 1, 2, \ldots, s\}$. A homothetic copy of $[n]^s$ is a set of the form

$$HC_{\vec{a},d,n} := \vec{a} + d[n]^s = \{\vec{a} + d\vec{x} \mid \vec{x} \in [n]^s\}$$

for some $\vec{a} \in \mathbb{N}^s$ and $d \in \mathbb{N}$, $d > 0$. The subscript n in $HC_{\vec{a},d,n}$ will be omitted after it is fixed.

Theorem 4.4 (T. Gallai) *Given any positive $r, n \in \mathbb{N}_0$, one can find an $N \in \mathbb{N}_0$ such that for every coloring $c : [N]^s \to [r]$ there exists $\vec{a}, d$ such that $HC_{\vec{a},d,n} \subseteq [N]^s$ and $c \restriction HC_{\vec{a},d,n} \equiv c_0$ for some $c_0 \in [r]$.*

The proof of Theorem 4.4 in this subsection is inspired by the proof of the one-dimensional version in [17].

Proof: Fix $n \in \mathbb{N}_0$. Let $\lhd$ be the lexicographical order of $HC_{\vec{a},d}$. For each $0 \leq l < n^s$ let $HC_{\vec{a},d}(l)$ denote the l-th element of $HC_{\vec{a},d}$ under $\lhd$. Note that $HC_{\vec{a},d}(0) = \vec{a}$.

Let $\varphi_m(r, N)$ be the following first-order sentence:

$$\begin{aligned}&\forall c : [N]^s \to [r] \ \exists HC_{\vec{a},d} \subseteq [N]^s \ \exists c_0 \in [r] \\ &\qquad (c(HC_{\vec{a},d}(l)) = c_0 \ \text{ for } \ l = 0, 1, \ldots, m)\,. \end{aligned} \tag{13}$$

It suffices to prove the following claim.

Claim 1: Let $0 \le m < n^s$. For every $r \in \mathbb{N}_0$ there exists an $N \in \mathbb{N}_0$ such that $\varphi_m(r, N)$ is true in $\mathcal{V}_0$.

Note that the claim when $m = n^s - 1$ is Theorem 4.4. It suffices to prove the claim by induction on $m \le n^s - 1$. Call $HC_{\vec{a},d}$ in (13) monochromatic up to m with respect to c.

Proof of Claim 1: The case for $m = 0$ is trivial.

Assume that the claim is true for $m - 1$. We prove that the claim is true for $m < n^s$.

Given $r \in \mathbb{N}_0$, the task now is to find $N \in \mathbb{N}_0$ such that $\varphi_m(r, N)$ is true in $\mathcal{V}_0$.

Work within $\mathcal{V}_{r+1}$. Choose any $N_r \in \mathbb{N}_{r+1} \setminus \mathbb{N}_r$. It suffices to prove that $\varphi_m(r, 2N_r)$ is true in $\mathcal{V}_r$ by the transfer principle.

Fix $c : [2N_r]^s \to [r]$. It suffices to find a $HC_{\vec{a},d} \subseteq [2N_r]^s$ which is monochromatic up to m with respect to c.

Choose any $N_j \in \mathbb{N}_{j+1} \setminus \mathbb{N}_j$ for $j = 0, 1, \ldots, r-1$. Since $\mathbb{N}_{j+1}$ is an end-extension of $\mathbb{N}_j$, the number $r^{(2N_{j-1})^s}$ is infinitely smaller than N_j. Note also that $N_j + N_{j-1} + \cdots + N_0 < N_{j+1}$.

For any $\vec{x}, \vec{y} \in [N_r]^s$ we say that $\vec{x}$ and $\vec{y}$ have the same $2N_j$-type if for any $\vec{z} \in [2N_j]^s$ we have $c(\vec{x} + \vec{z}) = c(\vec{y} + \vec{z})$, i.e., the color patterns of $\vec{x} + [2N_j]^s$ and $\vec{y} + [2N_j]^s$ with respect to c are the same.

Since the first-order sentence

$$(\forall r' \in \mathbb{N}_0)\,(\forall N \in \mathbb{N}_1 \setminus \mathbb{N}_0)\,\varphi_{m-1}(r', N)$$

is true in $(\mathcal{V}_1; \mathbb{N}_0)$, the sentence

$$(\forall r' \in \mathbb{N}_j)\,(\forall N \in \mathbb{N}_{j+1} \setminus \mathbb{N}_j)\,\varphi_{m-1}(r', N)$$

is true in $(\mathcal{V}_{j+1}; \mathbb{N}_j)$ for $j = 1, 2, \ldots, r$ by Part 4 of Definition 4.1. In particular, $\varphi_{m-1}(r^{(2N_{j-1})^s}, N_j)$ is true in $\mathcal{V}_{j+1}$ for $j = 1, 2, \ldots, r$.

Since the number of different $2N_{j-1}$-types is at most $r^{(2N_{j-1})^s}$, for any $\vec{b} + [N_j]^s$ we can find $HC_{\vec{a}_j,d_j} \subseteq \vec{b} + [N_j]^s$ such that $HC_{\vec{a}_j,d_j}$ is monochromatic up to $m - 1$ with respect to $2N_{j-1}$-types, i.e., $HC_{\vec{a}_j,d_j}(l)$ for $l = 0, 1, \ldots, m-1$ have the same $2N_{j-1}$-type. So, we can now find a sequence of homothetic copies of $[n]^s$

$$HC_{\vec{a}_r,d_r}, HC_{\vec{a}_{r-1},d_{r-1}}, \ldots, HC_{\vec{a}_0,d_0}$$

such that

- $HC_{\vec{a}_r,d_r} \subseteq [N_r]^s$ is monochromatic up to $m - 1$ with respect to $2N_{r-1}$-types;

- $HC_{\vec{a}_{r-1},d_{r-1}} \subseteq [N_{r-1}]^s$ such that $HC_{\vec{a}_r,d_r}(0) + HC_{\vec{a}_{r-1},d_{r-1}}$ is monochromatic up to $m-1$ with respect to $2N_{r-2}$-types. Note that $HC_{\vec{a}_r,d_r}(l) + HC_{\vec{a}_{r-1},d_{r-1}}(l')$ for $0 \le l, l' \le m-1$ have the same $2N_{r-2}$-type;

- $HC_{\vec{a}_{r-2},d_{r-2}} \subseteq [N_{r-2}]^s$ such that $HC_{\vec{a}_r,d_r}(0) + HC_{\vec{a}_{r-1},d_{r-1}}(0) + HC_{\vec{a}_{r-2},d_{r-2}}$ is monochromatic up to $m-1$ with respect to $2N_{r-3}$-types. Note that $HC_{\vec{a}_r,d_r}(l) + HC_{\vec{a}_{r-1},d_{r-1}}(l') + HC_{\vec{a}_{r-2},d_{r-2}}(l'')$ for $0 \le l, l', l'' \le m-1$ have the same $2N_{r-3}$-type;

- ......;

- $HC_{\vec{a}_1,d_1} \subseteq [N_1]^s$ such that $\sum_{j=2}^{r} HC_{\vec{a}_j,d_j}(0) + HC_{\vec{a}_1,d_1}$ is monochromatic up to $m-1$ with respect to $2N_0$-types. Note that $\sum_{j=2}^{r} HC_{\vec{a}_j,d_j}(l_j) + HC_{\vec{a}_1,d_1}(l_1)$ for $0 \le l_1, l_2, \ldots, l_r \le m-1$ have the same $2N_0$-type;

- $HC_{\vec{a}_0,d_0} \subseteq [N_0]^s$ such that $\sum_{j=1}^{r} HC_{\vec{a}_j,d_j}(0) + HC_{\vec{a}_0,d_0}$ is monochromatic up to $m-1$ with respect to coloring c. Note that $\sum_{j=1}^{r} HC_{\vec{a}_j,d_j}(l_j) + HC_{\vec{a}_0,d_0}(l_0)$ for $0 \le l_0, l_1, \ldots, l_r \le m-1$ have the same c-value.

Define $HC_{\vec{a},d} \oplus HC_{\vec{a}',d'} := HC_{\vec{a}+\vec{a}',d+d'}$. Clearly, for any $l < n^s$ we have

$$(HC_{\vec{a},d} \oplus HC_{\vec{a}',d'})(l) = HC_{\vec{a},d}(l) + HC_{\vec{a}',d'}(l).$$

For each $j = 0, 1, \ldots, r$ let

$$\vec{y}_j := HC_{\vec{a}_r,d_r}(0) + \cdots + HC_{\vec{a}_j,d_j}(0) + HC_{\vec{a}_{j-1},d_{j-1}}(m) + \cdots + HC_{\vec{a}_0,d_0}(m).$$

Since there are $r+1$ many y_j's and r colors, there must exist $0 \le j_1 < j_2 \le r$ such that $c(\vec{y}_{j_1}) = c(\vec{y}_{j_2})$. Let

$$\begin{aligned} D := {} & HC_{\vec{a}_r,d_r}(0) + \cdots + HC_{\vec{a}_{j_2},d_{j_2}}(0) \\ & + HC_{\vec{a}_{j_2-1},d_{j_2-1}} \oplus \cdots \oplus HC_{\vec{a}_{j_1},d_{j_1}} \\ & + HC_{\vec{a}_{j_1-1},d_{j_1-1}}(m) + \cdots + HC_{\vec{a}_0,d_0}(m). \end{aligned} \tag{14}$$

Then D is a homothetic copy of $[n]^s$.

Claim 2: The homothetic copy D of $[n]^s$ in (14) is monochromatic up to $m-1$ with respect to c.

Claim 1 follows from Claim 2 because $D(0) = \vec{y}_{j_1}$ and $D(m) = \vec{y}_{j_2}$ have the same c–value and hence, the homothetic copy D of $[n]^s$ is monochromatic up to m with respect to c.

Proof of Claim 2: By the construction of $HC_{\vec{a}_j,d_j}$ we have that

$$\sum_{j=j_2}^{r} HC_{\vec{a}_j,d_j}(0) + \sum_{j=j_1}^{j_2-1} HC_{\vec{a}_j,d_j}(l)$$

for $0 \leq l \leq m-1$ have the same $2N_{j_1-1}$-type. Note that

$$\vec{b} := \sum_{j=0}^{j_1-1} HC_{\vec{a}_j,d_j}(m) \in [2N_{j_1-1}]^s.$$

Hence,

$$\begin{aligned} D(l) := {} & HC_{\vec{a}_r,d_r}(0) + \cdots + HC_{\vec{a}_{j_2},d_{j_2}}(0) \\ & + HC_{\vec{a}_{j_2-1},d_{j_2-1}}(l) + \cdots + HC_{\vec{a}_{j_1},d_{j_1}}(l) + \vec{b} \end{aligned}$$

for $l = 0, 1, \ldots, m-1$ have the same c–value. This completes the proof of Claim 2, Claim 1, and the theorem. □

4.3 Szemerédi's Theorem

Szemerédi's Theorem is the center of attention in additive combinatorics for many years which has attracted many prominent mathematicians.

Theorem 4.5 (E. Szemerédi, 1975 [26]) *If $D \subseteq \mathbb{N}$ has a positive upper density, then D contains a k–term arithmetic progression for every $k \in \mathbb{N}$.*

Szemerédi's Theorem confirms a conjecture of P. Erdős and P. Turán made in 1936, which implies van der Waerden's Theorem.

Nonstandard versions of Furstenberg's ergodic proof and Gowers's harmonic proof of Szemerédi's Theorem have been tried by T. Tao (see Tao's blog post [28]). In August 2017, Tao gave a series of lectures to explain Szemerédi's original combinatorial proof and hope to simplify it so that the proof can be better understood. He believed that Szemerédi's combinatorial method should have a greater impact on combinatorics.

During these lectures Tao challenged the audience to produce a nonstandard proof of Szemerédi's Theorem which is noticeably simpler and more transparent than Szemerédi's original proof. However, in his later blog post [29], Tao commented that "in fact there are now signs that perhaps nonstandard analysis is not the optimal framework in which to place this argument." We disagree. The paper [16] was written to meet Tao's challenge and showed that with the help of a nonstandard universe with three levels of infinities, Szemerédi's original argument can be made simpler and more transparent.

The main simplification in the proof in [16] comparing to the standard proof in [26, 27] is that a Tower of Hanoi type induction in [27, Theorem 6.6] or in [26, Lemma 5, Lemma 6, and Fact 12] is replaced by a straightforward induction, which makes Szemerédi's idea more transparent. To achieve this, $\mathcal{V}_3$ (see Proposition 4.1) is used which supply three levels of infinities, plus various elementary embeddings from $\mathcal{V}_j$ to $\mathcal{V}_{j'}$ for some $0 \leq j < j' \leq 3$.

In this subsection we will do the following: (1) assume a weak regularity lemma and derive a nonstandard form of mixing lemma; (2) prove Theorem 4.5 for $k = 3$; (3) prove Theorem 4.5 for $k = 4$, (4) prove Theorem 4.5 for any k. The reason to present the proof for $k = 3$ and $k = 4$ is to show how the level of difficulties arises.

Let's fix some notation. The Greek letters $\alpha, \beta, \gamma, \epsilon$, etc. will represent standard real numbers unless otherwise specified. All unspecified sets mentioned are either standard or $\mathcal{V}_j$-internal for $j = 1, 2$, or 3. If $m, n \in \mathbb{N}_3$, we write $m \ll n$ if $m \in \mathbb{N}_j$ and $n \in \mathbb{N}_{j'} \setminus \mathbb{N}_{j'-1}$ for some $0 \leq j < j' \leq 3$. For example, $1 \ll n$ means that n is hyperfinite. The words "arithmetic progression" will be abbreviated to "a.p." The length of an a.p. p, denoted by $|p|$, is the number of the terms in p. A finite a.p., often with length k, will be denoted by p, q, r, etc. and an a.p. of hyperfinite length will be denoted by P, Q, R, etc. If P (or p) is an a.p., the l-th term of P is denoted by $P(l)$ for any $1 \leq l \leq |P|$. By k-term a.p. or just k-a.p. we mean an a.p. with length k. If both p and q are k-a.p., let $r := p \oplus q$ be the k-a.p. such that $r(l) = p(l) + q(l)$ for $1 \leq l \leq k$.

The following standard lemma is a consequence of Szemerédi's Regularity Lemma in [26]. The proof of the lemma can be found in the appendix section of [27].

Lemma 4.6 *Let U, W be finite sets, let $\epsilon > 0$, and for each $w \in W$, let E_w be a subset of U. Then there exists a partition $U = U_1 \cup U_2 \cup \cdots \cup U_{n_\epsilon}$ for some $n_\epsilon \in \mathbb{N}_0$, and real numbers $0 \leq c_{u,w} \leq 1$ in $\mathbb{R}_0$ for $u \in [n_\epsilon]$ and $w \in W$ such that for any set $F \subseteq U$, one has*

$$\left| |F \cap E_w| - \sum_{u=1}^{n_\epsilon} c_{u,w} |F \cap U_u| \right| \leq \epsilon |U|$$

for all but $\epsilon|W|$ values of $w \in W$.

For the mixing lemma we introduce some notion for slightly broader sense of Loeb measure, as well as strong upper Banach density in $\mathcal{V}_j$.

Definition 4.7 *Let* $0 \leq j < j' \leq 3$*. For any two numbers* $r, r' \in \mathbb{R}_{j'}$ *we write* $r \approx_j r'$ *if* $|r - r'| < 1/n$ *for every* $n \in \mathbb{N}_j$*. If* $r \in ns_j(\mathbb{R}_{j'})$ *where* $ns_j(\mathbb{R}_{j'})$ *is the set of all near standard elements when considering* $\mathcal{V}_j$ *as the "standard" universe (cf. Definition 1.12), denote* $st_j(r)$ *for the unique number* $\alpha \in \mathbb{R}_j$ *such that* $r \approx_j \alpha$*. For any bounded set* $A \subseteq \mathbb{N}_{j'}$ *and* $n \in \mathbb{N}_{j'}$ *denote*

$$\delta_n(A) := \frac{|A|}{n} \in \mathbb{R}_{j'} \text{ and } \mu_n^j(A) := st_j(\delta_n(A)).$$

Notice that δ_n is a $\mathcal{V}_{j'}$-internal function while μ_n^j are often external functions but definable in $(\mathcal{V}_{j'}; \mathbb{R}_j)$, i.e.,

$$\mu_n^j(A) = \alpha \text{ iff } \forall n \in \mathbb{N}_{j'} \cap \mathbb{R}_j \left(|\delta_H(A) - \alpha| < \frac{1}{n} \right).$$

If $A \subseteq \Omega$ and $|\Omega| = H$, then $\mu_H(A)$ coincides with the Loeb measure of A in Ω (Definition 2.11). The term δ_H is often used for an internal argument.

Definition 4.8 *Let* $0 \leq j < j' \leq 3$ *and* $A \subseteq \mathbb{N}_{j'}$ *with* $|A| \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ *the strong upper Banach density* $SD^j(A)$ *of* A *in* $\mathcal{V}_j$ *is defined by*

$$SD^j(A) := \sup_j \left\{ \mu_{|P|}^j(A \cap P) \mid |P| \in \mathbb{N}_{j'} \setminus \mathbb{N}_j \right\}. \tag{15}$$

The letter P *above always represents an a.p. and* $\sup_j$ *represents the least upper bound in* $\mathbb{R}_j \cup \{\pm\infty\}$ *of a subset of* $\mathbb{R}_j$ *in* $\mathcal{V}_j$*. If* $S \subseteq \mathbb{N}_{j'}$ *has* $SD^j(S) = \eta \in \mathbb{R}_j$ *and* $A \subseteq \mathbb{N}_{j'}$*, the strong upper Banach density* SD_S^j *of* A *relative to* S *is defined by*

$$SD_S^j(A) := \sup_j \left\{ \mu_{|P|}^j(A \cap P) \mid |P| \in \mathbb{N}_{j'} \setminus \mathbb{N}_j, \text{ and } \mu_{|P|}^j(S \cap P) = \eta \right\}. \tag{16}$$

Note that if Szemeredédi's Theorem is true, then $SD(A) > 0$ implies $SD(A) = 1$. However, we haven't proven the theorem yet. Similar to Proposition 3.6, we can derive some nonstandard version of strong upper Banach density.

Proposition 4.9 *Let* $0 \leq j < j' \leq 3$*. Given* $A \subseteq S \subseteq \mathbb{N}_{j'}$ *with* $|A| \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ *and* $\alpha, \eta \in \mathbb{R}_j$ *with* $0 \leq \alpha \leq \eta \leq 1$*. Then the following are true:*

1. $SD^j(S) \geq \eta$ *iff there exists a* P *with* $|P| \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ *and* $\mu_{|P|}^j(S \cap P) \geq \eta$*;*

2. *If* $SD^j(S) = \eta$*, then there exists a* P *with* $|P| \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ *such that* $\mu_{|P|}^j(S \cap P) = SD^j(S \cap P) = \eta$*;*

3. *Suppose* $SD^j(S) = \eta$*. Then* $SD_S^j(A) \geq \alpha$ *iff there exists a* P *with* $|P| \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$*,* $\mu_{|P|}^j(S \cap P) = \eta$*, and* $\mu_{|P|}^j(A \cap P) \geq \alpha$*;*

4. *Suppose* $SD^j(S) = \eta$. *If* $SD^j_S(A) = \alpha$, *then there exists a* P *with* $|P| \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ *such that* $\mu^j_{|P|}(S \cap P) = \eta$ *and* $\mu^j_{|P|}(A \cap P) = SD^j_{S\cap P}(A \cap P) = \alpha$.

Proof: Part 1: If $SD^j(S) \geq \eta$, then there is a P_n with $|P_n| \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ such that $\delta_{|P_n|}(S \cap P_n) > \eta - 1/n$ for every $n \in \mathbb{N}_j$. Let

$$A := \left\{ n \in \mathbb{N}_{j'} \mid \exists P \subseteq \mathbb{N}_{j'} \left(|P| \geq n \wedge \delta_{|P|}(S \cap P) > \eta - \frac{1}{n} \right) \right\}.$$

Then A is $\mathcal{V}_{j'}$–internal and $A \cap \mathbb{N}_j$ is unbounded above in $\mathbb{N}_j$. By Proposition 2.4, there is a $J \in A \setminus \mathbb{N}_j$. Hence there is an a.p. $P_J \subseteq \mathbb{N}_{j'}$ such that $|P_J| \geq J \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ and $\delta_{|P_J|}(S \cap P_J) > \eta - 1/J \approx_j \eta$. Therefore, $\mu^j_{|P_J|}(S \cap P_J) \geq \eta$. On the other hand, if $\mu^j_{|P|}(S \cap P) \geq \eta$, then $SD^j(S) \geq \eta$ by the definition of SD^j in (15).

Part 2: If $SD^j(S) = \eta$, we can find P with $|P| \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ such that $\mu^j_{|P|}(S \cap P) = \eta' \geq \eta$ by Part 1. Clearly, $\eta = SD^j(S) \geq SD^j(S \cap P) \geq \mu^j_{|P|}(S \cap P) = \eta'$ by the definition of SD^j. Hence $\eta = \eta'$.

Part 3: If $SD^j_S(A) \geq \alpha$, then there is a P with $|P| > n$ such that $|\delta_{|P|}(S \cap P) - \eta| < 1/n$ and $\delta_{|P|}(A \cap P) > \alpha - 1/n$ for every $n \in \mathbb{N}_j$. By Proposition 2.4 as in the proof of Part 1 there is a P_J for some $J \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ with $|P_J| \geq J$ such that $|\delta_{|P_J|}(S \cap P_J) - \eta| < 1/J$ and $\delta_{|P_J|}(A \cap P_J) > \alpha - 1/J$, which implies $\mu^j_{|P_J|}(S \cap P) = \eta$ and $\mu^j_{|P_J|}(A \cap P_J) \geq \alpha$. On the other hand, if $\mu_{|P|}(S \cap P) = \eta$ and $\mu^j_{|P|}(A \cap P) \geq \alpha$, then $SD^j_S(A) \geq \alpha$ by the definition of SD^j_S in (16).

Part 4: If $SD^j_S(A) = \alpha$, then $\mu^j_{|P|}(S \cap P) = \eta$ and $\mu^j_{|P|}(A \cap P) = \alpha' \geq \alpha$ for some P with $|P| \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ by Part 3. Clearly, $\alpha = SD^j_S(A) \geq SD^j_{S\cap P}(A \cap P) \geq \mu^j_{|P|}(A \cap P) = \alpha'$ by the definition of SD^j_S. Hence $\alpha = \alpha'$. ☐

The uniformity of $A \in [N]$ when $\mu_N(A) = SD(A)$ will be useful.

Lemma 4.10 *Let* $0 \leq j < j' \leq 3$. *Given* $N, H \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$, $H \leq N/2$, *and* $C \subseteq [N]$ *with* $\mu^j_N(C) = SD^j(C) = \alpha \in \mathbb{R}_j$, *for each* $n \in \mathbb{N}_{j'}$ *let*

$$D_{n,H,C} := \left\{ x \in [N - H] \mid |\delta_H(C \cap (x + [H])) - \alpha| < \frac{1}{n} \right\}. \tag{17}$$

Then there exists a $J \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ *such that* $\mu^j_{N-H}(D_{J,H,C}) = 1$.

Notice that $D_{n,H,C} \subseteq D_{n',H,C}$ if $n \geq n'$.

Proof: Fix N, H, and C. The subscripts H and C in $D_{n,H,C}$ will be omitted in the proof. If $st_j(H/N) > 0$, then for every $x \in [N - H]$ we have $\mu^j_H(x + [H]) = \alpha$ by the supremality of α. Hence the maximal J with $J \leq H$ such that $|\delta_H(A\cap(x+[H]))-\alpha| < 1/J$ for every $x \in [N-H]$ is in $\mathbb{N}_{j'} \setminus \mathbb{N}_j$. Now $D_J = [N-H]$ works.

Assume that $st_j(H/N) = 0$. So, μ^j_{N-H} and μ^j_N coincide. If $\delta_N(D_n) \approx_j 1$ for every $n \in \mathbb{N}_j$, then the maximal J satisfying $|\delta_N(D_J) - 1| < 1/J$ must be in $\mathbb{N}_{j'} \setminus \mathbb{N}_j$ by Proposition 2.4. Hence $\mu^j_N(D_J) = 1$. So we can assume that $\mu^j_N(D_n) < 1$ for some $n \in \mathbb{N}_j$ and derive a contradiction.

Notice that for each $x \in [N-H]$, it is impossible to have $\mu^j_H(C \cap (x+[H])) > \alpha = SD^j(C)$ by the definition of SD^j. Let $\overline{D}_n := [N-H] \setminus D_n$. Then $\mu^j_N(\overline{D}_n) = 1 - \mu^j_N(D_n) > 0$. Notice that $x \in \overline{D}_n$ implies $\delta_H(C \cap (x+[H])) \leq \alpha - 1/n$. By the following double counting argument, by ignoring some $\mathcal{V}_j$–infinitesimal amount inside st_j, we have

$$\begin{aligned}
\alpha &= st_j\left(\frac{1}{H}\sum_{y=1}^{H}\delta_N(C-y)\right) = st_j\left(\frac{1}{HN}\sum_{y=1}^{H}\sum_{x=1}^{N}\chi_C(x+y)\right)\\
&= st_j\left(\frac{1}{NH}\sum_{x=1}^{N}\sum_{y=1}^{H}\chi_C(x+y)\right) = st_j\left(\frac{1}{N}\sum_{x=1}^{N}\delta_H(C\cap(x+[H]))\right)\\
&= st_j\left(\frac{1}{N}\sum_{x\in D_n}\delta_H(C\cap(x+[H])) + \frac{1}{N}\sum_{x\in \overline{D}_n}\delta_H(C\cap(x+[H]))\right)\\
&\leq \alpha\mu^j_N(D_n) + \left(\alpha - \frac{1}{n}\right)\mu^j_N(\overline{D}_n) < \alpha
\end{aligned}$$

which is absurd. This completes the proof. □

Suppose $0 \leq j < j' \leq 3$, $N \geq H \gg 1$ in $\mathbb{N}_{j'}$, $U \subseteq [N]$, $A \subseteq S \subseteq [N]$, $0 \leq \alpha \leq \eta \leq 1$, and $x \in [N]$. For each $n \in \mathbb{N}_j$ let $\xi(x, \alpha, \eta, A, S, U, H, n)$ be the following internal statement:

$$\begin{aligned}
|\delta_H(x+[H]) \cap U) - 1| &< 1/n,\\
|\delta_H((x+[H]) \cap S) - \eta| &< 1/n, \text{ and}\\
|\delta_H((x+[H]) \cap A) - \alpha| &< 1/n.
\end{aligned} \tag{18}$$

The statement $\xi(x, \alpha, \eta, A, S, U, H, n)$ infers that the densities of A, S, U in the interval $x + [H]$ go to $\alpha, \eta, 1$, respectively, as $n \to \infty$ in $\mathbb{N}_j$. The statement ξ will be referred a few times in Lemma 4.18 and its proof.

The following lemma is the application of Lemma 4.10 to the sets U, S, A simultaneously.

Lemma 4.11 *Let $0 \leq j < j' \leq 3$. Let $N \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$, $U \subseteq [N]$, and $A \subseteq S \subseteq [N]$ be such that $\mu^j_N(U) = 1$, $\mu^j_N(S) = SD(S) = \eta$, and $\mu^j_N(A) = SD^j_S(A) = \alpha$ for some*

$\eta, \alpha \in \mathbb{R}_j$. For any $n, h \in \mathbb{N}_{j'}$ let

$$G_{n,h} := \{x \in [N-h] \mid \mathcal{V}_{j'} \models \xi(x, \alpha, \eta, A, S, U, h, n)\}. \tag{19}$$

(a) For each $H \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ with $H \leq N/2$ there exists a $J \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ such that $\mu^j_{N-H}(G_{J,H}) = 1$;

(b) For each $n \in \mathbb{N}_j$, there is an $h_n \in \mathbb{N}_j$ with $h_n > n$ such that $\delta_N(G_{n,h_n}) > 1 - 1/n$.

Proof: Part (a): Applying Lemma 4.10 for U and S we can find $J_1, J_2 \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ such that $\mu^j_{N-H}(D_{J_1,H,U}) = 1$ and $\mu^j_{N-H}(D_{J_2,H,S}) = 1$ where $D_{n,h,C}$ is defined in (17) and α is replaced by 1 for U and η for S. Let $G' := D_{J_1,H,U} \cap D_{J_2,H,S}$. For each $n \leq \min\{J_1, J_2\}$ let

$$\overline{G}''_n := \left\{x \in [N-H] \mid \delta_H(A \cap (x + [H])) > \alpha + \frac{1}{n}\right\}, \text{ and}$$

$$\underline{G}''_n := \left\{x \in [N-H] \mid \delta_H(A \cap (x + [H])) < \alpha - \frac{1}{n}\right\}.$$

Notice that both $\overline{G}''_n$ and $\underline{G}''_n$ are $\mathcal{V}_{j'}$–internal. If $\mu^j_{N-H}(\overline{G}''_n) > 0$ for some $n \in \mathbb{N}_j$, then $\overline{G}''_n \cap G' \neq \emptyset$. Let $x_0 \in \overline{G}''_n \cap G'$. Then we have $\mu^j_H(S \cap (x_0 + [H])) = \eta$ and $\mu^j_H(A \cap (x_0 + [H])) > \alpha + 1/n$, which contradicts $SD^j_S(A) = \alpha$. Hence $\delta_{N-H}(\overline{G}''_n) \approx_j 0$ for every $n \in \mathbb{N}_j$. By Proposition 2.4 we can find $J_+ \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ such that $\mu^j_{N-H}(\overline{G}''_n) = 0$ for any $n \leq J_+$. If $\mu^j_{N-H}(\underline{G}''_n) > 0$ for some $n \in \mathbb{N}_j$, then $\mu^j_{N-H}(\overline{G}''_m) > 0$ for some $m \in \mathbb{N}_j$ by the fact that $\mu^j_{N-H}(A) = \alpha$. Hence $\delta_{N-H}(\underline{G}''_n) \approx_j 0$ for every $n \in \mathbb{N}_j$. By Proposition 2.4 again we can find $J_- \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ such that $\mu^j_{N-H}(\underline{G}''_n) = 0$ for any $n \leq J_-$. The proof is complete by setting $J := \min\{J_1, J_2, J_+, J_-\}$ and

$$G_{J,H} := (D_{J,H,U} \cap D_{J.H,S}) \setminus (\overline{G}''_J \cup \underline{G}''_J).$$

Part (b): Suppose Part (b) is not true. Then there exists an $n \in \mathbb{N}_j$ such that $\delta_{N-h}(G_{n,h}) \leq 1 - 1/n$ for any $h > n$ in $\mathbb{N}_j$. By Proposition 2.4 there is an $H \in \mathbb{N}_{j'} \setminus \mathbb{N}_j$ such that $\delta_{N-H}(G_{n,H}) \leq 1 - 1/n$. By Part (a) there is a $J \gg n$ such that $\mu^j_{N-H}(G_{J,H}) = 1$. We have a contradiction because $n < J$ and hence $G_{J,H} \subseteq G_{n,H}$. □

Notice that for a given n one can choose h_n to be the least such that $\delta_N(G_{n,h_n}) > 1 - 1/n$ in Lemma 4.11 (b). So we can assume that h_n is an internal function of n. Hence we can assume that G_{n,h_n} is also an internal function of n.

We often write *st* for st_0, μ_n for μ^0_n, and *SD* for SD^0. One can derive a so-called mixing lemma from Lemma 4.6.

Lemma 4.12 (Mixing Lemma) *Let $N \in \mathbb{N}_{j'} \setminus \mathbb{N}_0$, $A \subseteq S \subseteq [N]$, $1 \ll H \leq N/2$, and $R \subseteq [N-H]$ be an a.p. with $|R| \gg 1$ such that*

$$\mu_N(S) = SD(S) = \eta > 0,\ \mu_N(A) = SD_S(A) = \alpha > 0, \tag{20}$$

$$\mu_H((x+[H]) \cap S) = \eta, \text{ and } \mu_H((x+[H]) \cap A) = \alpha \tag{21}$$

for every $x \in R$. Then the following are true.

(i) For any set $E \subseteq [H]$ with $\mu_H(E) > 0$, there is an $x \in R$ such that

$$\mu_H(A \cap (x+E)) \geq \alpha \mu_H(E);$$

(ii) Let $m \gg 1$ be such that the van der Waerden number $\Gamma(3^m, m) \leq |R|$. For any internal partition $\{U_n \mid n \in [m]\}$ of $[H]$ there exists an m–a.p. $P \subseteq R$, a set $I \subseteq [m]$ with $\mu_H(U_I) = 1$ where $U_I = \bigcup\{U_n \mid n \in I\}$, and an infinitesimal $\epsilon > 0$ such that

$$|\delta_H(A \cap (x+U_n)) - \alpha\delta_H(U_n)| \leq \epsilon\delta_H(U_n)$$

for all $n \in I$ and all $x \in P$;

(iii) Given an internal collection of sets $\{E_w \subseteq [H] \mid w \in W\}$ with $|W| \gg 1$ and $\mu_H(E_w) > 0$ for every $w \in W$, there exists an $x \in R$ and $T \subseteq W$ such that $\mu_{|W|}(T) = 1$ and

$$\mu_H(A \cap (x+E_w)) = \alpha\mu_H(E_w)$$

for every $w \in T$.

Proof: Part (i): Assume that (i) is not true. For each $x \in R$ let r_x be such that $\delta_H(A \cap (E+x)) = (\alpha - r_x)\delta_H(E)$. Then r_x must be positive non-infinitesimal. We can set $r := \min\{r_x \mid x \in R\}$ since the function $x \mapsto r_x$ is internal. Clearly, the number r is positive non-infinitesimal. Hence $\delta_H(A \cap (E+x)) \leq (\alpha - r)\delta_H(E)$ for all $x \in R$. Notice that by (20) and (21), for μ_H–almost all $y \in [H]$ we have $\mu_{|R|}(S \cap (y+R)) = \eta$ which implies that for μ_H–almost all $y \in [H]$ we have $\mu_{|R|}(A \cap (y+R)) = \alpha$. So

$$\alpha\mu_H(E) \approx \frac{1}{H}\sum_{y\in E}\frac{1}{|R|}\sum_{x\in R}\chi_A(x+y) = \frac{1}{|R|}\sum_{x\in R}\frac{1}{H}\sum_{y=1}^{H}\chi_{A\cap(E+x)}(x+y)$$
$$\leq \frac{1}{|R|}\sum_{x\in R}(\alpha-r)\delta_H(E) = (\alpha-r)\delta_H(E) \approx (\alpha - st(r))\mu_H(E) < \alpha\mu_H(E),$$

which is absurd.

Part (ii): To make the argument explicitly internal we use δ_H instead of μ_H. For each $t \in \mathbb{N}_{j'}$, $x \in R$, and $n \in [m]$ let

$$c_n^t(x) = \begin{cases} 1 & \text{if } \delta_H((x+U_n)\cap A) \geq \left(\alpha + \frac{1}{t}\right)\delta_H(U_n), \\ 0 & \text{if } \left(\alpha - \frac{1}{t}\right)\delta_H(U_n) < \delta_H((x+U_n)\cap A) < \left(\alpha + \frac{1}{t}\right)\delta_H(U_n), \\ -1 & \text{if } \delta_H((x+U_n)\cap A) \leq \left(\alpha - \frac{1}{t}\right)\delta_H(U_n). \end{cases}$$

and let $c^t : P \to \{-1, 0, 1\}^{[m]}$ be such that $c^t(x)(n) = c_n^t(x)$. For each $t \in \mathbb{N}_0$, since the van der Waerden number $\Gamma(3^m, m) \leq |R|$, there exists an m–a.p. $P_t \subseteq R$ such that $c^t(x) = c^t(x')$ for any $x, x' \in P_t$. For each $x \in P_t$ let

$$\begin{aligned} I_t^+ &= \{n \in [m] \mid c^t(x)(n) = 1\}, \\ I_t^- &= \{n \in [m] \mid c^t(x)(n) = -1\}, \text{ and} \\ I_t &= [m] \setminus (I_t^+ \cup I_t^-), \text{ and} \end{aligned}$$

$$\begin{aligned} U_t^+ &= \bigcup\{U_n \mid n \in I_t^+\}, \\ U_t^- &= \bigcup\{U_n \mid n \in I_t^-\}, \text{ and} \\ U_t &= [H] \setminus (U_t^+ \cup U_t^-). \end{aligned}$$

Clearly, $\delta_H((x+U_t^-)\cap A) \leq (\alpha - 1/t)\delta_H(U_t^-)$ because U_t^- is a disjoint union of the U_n's for $n \in I_t^-$. Since $t \in \mathbb{N}_0$ we have that $\mu_H(U_t^-) = 0$ by (i) with P_t in the place of R and U_t^- in the place of E. Notice that $\delta_H(A \cap (x + U_t^+)) \geq (\alpha + 1/t)\delta_H(U_t^+)$. Since $\alpha \geq \mu_H(A \cap (x + U_t^+)) \geq (\alpha + 1/t)\mu_H(U_t^+)$, we have that $\mu_H(U_t^+) < 1$, which implies $\mu_H(U_t) > 0$. If $\mu_H(U_t^+) > 0$, then $\delta_H(A \cap (x + U_t^+)) \geq (\alpha + 1/t)\delta_H(U_t^+)$ implies $\mu_H(A \cap (x+U_t)) < \alpha\mu_H(U_t)$ for all $x \in P_t$, which again contradicts (i). Hence $\mu_H(U_t^+) = 0$ and therefore, $\delta_H(U_t) > 1 - 1/t$ is true for every $t \in \mathbb{N}_0$.

Since the set of all $t \in \mathbb{N}_{j'}$ with $\delta_H(U_t) > 1 - 1/t$ is $\mathcal{V}_{j'}$–internal, by Proposition 2.4 there is a $J \gg 1$ such that $\delta_H(U_J) > 1 - 1/J \approx 1$. The proof of (ii) is completed by letting $P := P_J$, $I := I_J$, and $U_I := U_J$.

Part (iii): Choose a sufficiently large positive infinitesimal ϵ satisfying that there is an internal partition of $[H] = U_0 \cup U_1 \cup \cdots \cup U_m$ and real numbers $0 \leq c_{n,w} \leq 1$ for each $n \in [m]$ and $w \in W$ such that the van der Waerden number $\Gamma(3^m, m) \leq |R|$, and for any internal set $F \subseteq [H]$ there is a $T_F \subseteq W$ with $|W \setminus T_F| \leq \epsilon|W|$ such that

$$\left| |F \cap E_w| - \sum_{n=1}^{m} c_{n,w}|F \cap U_n| \right| \leq \epsilon H \tag{22}$$

for all $w \in T_F$. Notice that such ϵ exists because if ϵ is a standard positive real, then $m = n_\epsilon$ is in $\mathbb{N}_0$. From (22) with F being replaced by $[H]$ we have

$$\left| |E_w| - \sum_{n=1}^{m} c_{n,w} |U_n| \right| \leq \epsilon H \tag{23}$$

for all $w \in T_{[H]}$. By (ii) we can find a $P \subseteq R$ of length m, a positive infinitesimal ϵ_1, and $I \subseteq [m]$ where, for some $x \in P$,

$$I := \{n \in [m] \mid |\delta_H((x + U_n) \cap A) - \alpha\delta_H(U_n)| < \epsilon_1 \delta_H(U_n)\}$$

(I is independent of the choice of x), and $V := \bigcup\{U_n \mid n \in I\}$ with $\mu_H(V) = 1$. Let $I' = [m] \setminus I$ and $V' = [H] \setminus V$. Then for each $w \in T := T_{[H]} \cap T_{(A-x)\cap[H]}$ we have

$$\begin{aligned}
&|\delta_H(A \cap (x + E_w)) - \alpha\delta_H(E_w)| \\
&\leq \frac{1}{H}\left(\left| |A \cap (x + E_w)| - \sum_{n\in[m]} c_{n,w} |A \cap (x + U_n)| \right| \right. \\
&\quad + \left| \sum_{n\in[m]} c_{n,w} |A \cap (x + U_n)| - \sum_{n\in[m]} c_{n,w} \alpha |U_n| \right| \\
&\quad \left. + \left| \alpha \sum_{n\in[m]} c_{n,w} |U_n| - \alpha \sum_{i\in[m]} |E_w| \right| \right) \\
&\leq \epsilon + \frac{1}{H} \sum_{n\in I} c_{n,w} \epsilon_1 |U_n| + 2\delta_H(V') + \alpha\epsilon \\
&\leq \epsilon + \epsilon_1 \delta_H(V) + 2\delta_H(V') + \alpha\epsilon \approx 0.
\end{aligned}$$

Hence $\mu_H(A \cap (x + E_w)) = \alpha\mu_H(E_w)$ for all $w \in T$. Notice that $\mu_{|W|}(T) = 1$ because $\epsilon \approx 0$ and $\mu_{|W|}(T_{[H]}) = \mu_{|W|}(T_{[H]\cap(A-x)}) = 1$. □

The set S in Lemma 4.12, although seems unnecessary, is needed in the proof of Lemma 4.18.

Szemeredi's Theorem for $k = 3$:

Theorem 4.13 (K. F. Roth, 1953) *If $U \subseteq \mathbb{N}$ and $SD(U) > 0$, then U contains nontrivial 3-term arithmetic progressions.*

Proof: We work within $\mathcal{V}_1$. The elementary embedding $i_{0,0}$ is represented by * for notational convenience.

Let $\alpha = SD(U)$. Then $\alpha > 0$. Let $P \subseteq \mathbb{N}_1$ be an a.p. with $|P| \gg 1$ and $\mu_{|P|}({}^*U \cap P) = \alpha$. Without loss of generality we can assume that $P = [N] \cup \{0\}$. Let $A := {}^*U \cap [N]$. It suffices to find a 3-a.p. in A.

Let $H = \lfloor N/6 \rfloor$ and $S = [N - H]$. Notice that $\{0\} \cup (H + [H]) \cup (2H + 2[H]) \subseteq S$. For each $t \in [H]$ let

$$\mathcal{Q}_t = \{q \subseteq [H] \mid q \text{ is a 3–a.p., } q(1) \in A \cap [H], \text{ and } q(3) = t\}$$

$$\text{and } E_t = \{q(2) \mid q \in \mathcal{Q}_t\}.$$

Notice that $\mu_H(E_t) = \alpha/2 > 0$ because $p(1) - t$ must be even and the density of A in an a.p. of difference 2 and length $\geq \lfloor N/16 \rfloor$ is also α. By (iii) of Lemma 4.12, there is an $l \in [H]$ and $T \subseteq [H]$ with $\mu_H(T) = 1$ such that

$$\mu_H(A \cap (H + l + E_t)) = \alpha^2/2$$

for all $t \in T$. Since $2H + 2l \in S$ and $\mu_H(T) = 1$, we have

$$\mu_H(A \cap (2H + 2l + T)) = \alpha > 0.$$

Let $t_0 \in T$ be such that $2H + 2l + t_0 \in A \cap (2H + 2l + T)$. Let $p_0 = \{0, H + l, 2H + 2l\}$ and $q_0 \in \mathcal{Q}_{t_0}$ with $H + l + q_0(2) \in A \cap (H + l + E_{t_0})$. Then $p_0 \oplus q_0$ is an 3–a.p. Clearly, $p_0(3) + q_0(3) = 2H + 2l + t_0 \in (2H + 2l + T) \cap A \subseteq A$, $p_0(2) + q_0(2) \in (H + l + E_{t_0}) \cap A \subseteq A$, and $p_0(1) + q_0(1) = q(1) \in A$ by the definition of E_{t_0}. Note that there are at least $\alpha^2 H/2$ many 3-a.p. q's in $\mathcal{Q}_{t_0}$ with $p_0 \oplus q \subseteq A$. □

Szemeredi's Theorem for $k = 4$:

We again work in $\mathcal{V}_1$. If one wants to count the number of 4-a.p.'s such that all but the third (or second) term of the a.p. are in a set A, then the same idea of the proof of Roth's Theorem can be used to prove the following lemma.

Lemma 4.14 *Let $N \gg 1$, $A \subseteq [N]$ be such that $\mu_N(A) = SD(A) = \alpha > 0$, and $H = \lfloor N/8 \rfloor$. There exists an interval $x_0 + [H] \subseteq [N]$, a set $T \subseteq x_0 + [H]$ with $\mu_H(T) = 1$, and*

$$\mathcal{P}_t := \{p \subseteq [N] \mid p \text{ is a 4–a.p.}, p(1), p(2) \in A, \text{ and } p(4) = t\}$$

such that $\mu_H(\mathcal{P}_t) = \alpha^2/3$ for each $t \in T$.

The reason why the number of 4-a.p.'s in A is $\geq \alpha^2 H/3$ instead of $\alpha^2 H/2$ as in Theorem 4.13 is that for a 4-a.p. p with $p(4) = t$ fixed, $p(4) - p(1)$ should be a multiple of 3 in order to guarantee that $p(2)$ and $p(3)$ are integers.

Lemma 4.15 *Let $N \gg 1$, $B, S_\gamma \subseteq [N]$ be such that $B \subseteq S_\gamma$, $\mu_N(S_\gamma) = SD(S_\gamma) = \gamma > 11/12$, $\mu_N(B) = SD_{S_\gamma}(B) = \beta > 0$. There exists an interval $x_0 + [\lfloor N/24 \rfloor] \subseteq [N]$ and a set $T \subseteq x_0 + [\lfloor N/24 \rfloor]$ with $\mu_{N/24}(T) \geq 1 - 12(1-\gamma)$, and a collection of 4-a.p.'s $\{p_t \mid t \in T\}$ such that $p_t(1), p_t(2) \in B$, $p_t(3), p_t(4) \in S_\gamma$, and $p_t(3) = t$ for each $t \in T$.*

Proof Let $H := [N/8]$. Notice that $\mu_H(S_\gamma \cap (x + [H])) = \gamma$ and $\mu_H(B \cap (x + [H])) = \beta$ for every $x \in [N - H]$. Let $\mathcal{Q}$ be the collection of all 4–a.p.'s in $[H]$. For each $w \in [\lfloor H/3 \rfloor, \lfloor 2H/3 \rfloor]$ let

$$\mathcal{Q}_w^3 := \{q \in \mathcal{Q} \mid q(1) \in B \text{ and } q(3) = w\}$$
$$\text{and } E_w^3 := \{q(2) \mid q \in \mathcal{Q}_w^3\}.$$

We have that $\mu_H(E_w^3) = \beta/2$. For each $w' \in [H]$ let

$$\mathcal{R}_{w'}^i := \{q \in \mathcal{Q} \mid q(1) \in B \text{ and } q(i) = w'\}$$
$$\text{and } F_{w'}^i := \{q(2) \mid q \in \mathcal{R}_{w'}^i\}$$

for $i = 3, 4$. Clearly, $\mu_H(F_{w'}^i) \leq \beta$.

By (iii) of Lemma 4.12, there is an $l \in [H]$, $W_3 \subseteq [\lfloor H/3 \rfloor, \lfloor 2H/3 \rfloor]$ with $\mu_H(W_3) = 1/3$, and $W^i \subseteq [H]$ with $\mu_H(W^i) = 1$ such that

$$\mu_H(B \cap (H + l + E_w^3)) = \frac{\beta^2}{2} \text{ and } \mu_H(B \cap (H + l + F_{w'}^i)) \leq \beta^2$$

for all $w \in W_3$ and $w' \in W^i$ for $i = 3$ or 4. Clearly, $\mu_H(((i-1)H + (i-1)l + W^i) \cap S_\gamma) = \gamma$ for $i = 3$ or 4.

Let $T^3 := 2H + 2l + W_3$. For each $t = 2H + 2l + w \in T^3$ let

$$\mathcal{P}_t := \{p \text{ is a 4–a.p. in } [N] \mid p(1) \in B \cap [H], p(2) \in B \cap (H + l + E_w^3), p(3) = t\}$$
$$\text{and } \mathcal{P} := \bigcup_{i \in T^3} \mathcal{P}_t.$$

Notice that $\mu_H(\mathcal{P}_t) = \mu_H(B \cap (2H + 2l + E_w^3)) = \beta^2/2$ for each $t = 2H + 2l + w \in T^3$.

A 4–a.p. $p \in \mathcal{P}$ is called *good* if $p(i) \in S_\gamma \cap ((i-1)H + (i-1)l + [H])$ for $i = 3, 4$. Let $\mathcal{P}_g$ be the collection of all good 4–a.p.'s in $\mathcal{P}$. A 4–a.p. $p \in \mathcal{P}$ is *bad* if it is not good. Let $\mathcal{P}_b := \mathcal{P} \smallsetminus \mathcal{P}_g$. Let $T_g^3 := \{p(3) \mid p \in \mathcal{P}_g\}$. Then $T_g^3 \subseteq S_\gamma$. We show that $\mu_H(T_g^3) \geq \frac{1}{3} - 4(1-\gamma)$.

Notice that $\mathcal{P}_b \subseteq \bigcup_{i=3,4} \{p \in \mathcal{P} \mid p(1) \in B \cap [H], p(2) \in B \cap (h + l + [H]), p(i) \notin S_\gamma\}$. Hence

$$|\mathcal{P}_b| \leq \sum_{i=3}^{4} \sum_{w' \in [H] \smallsetminus (S_\gamma - (i-1)H - (i-1)l)} |F_{w'}^i|$$

$$\leq \sum_{i=3}^{4} \left(\sum_{w' \in [H] \smallsetminus W^i} |F^i_{w'}| + \sum_{w' \in W^i \smallsetminus (S_\gamma - (i-1)H - (i-1)l)} |F^i_{w'}| \right).$$

So $|\mathcal{P}_g| = |\mathcal{P}| - |\mathcal{P}_b|$

$$\geq \sum_{t \in T^3} |\mathcal{P}_t| - \sum_{i=3}^{4} \left(\sum_{w' \in [H] \smallsetminus W^i} |F^i_{w'}| + \sum_{w' \in W^i \smallsetminus (S_\gamma - (i-1)H - (i-1)l)} |F^i_{w'}| \right).$$

Hence we have

$$\begin{aligned}
\mu_H(T^3_g) \cdot \frac{\beta^2}{2} &= st \left(\frac{1}{H} \sum_{t \in T^3_g} \frac{1}{H} |\mathcal{P}_t| \right) \\
&\geq st \left(\frac{1}{H^2} |\mathcal{P}_g| \right) = st \left(\frac{1}{H^2} (|\mathcal{P}| - |\mathcal{P}_b|) \right) \\
&\geq st \left(\frac{1}{H^2} \sum_{t \in T^3} |\mathcal{P}_t| \right) \\
&\quad - st \left(\frac{1}{H^2} \sum_{i=3}^{4} \left(\sum_{w' \in [H] \smallsetminus W^i} |F^i_{w'}| + \sum_{w' \in W^i \smallsetminus (S_\gamma - (i-1)H - (i-1)l)} |F^i_{w'}| \right) \right) \\
&\geq \mu_H(T^3) \cdot \frac{\beta^2}{2} - 2(1-\gamma) \cdot \beta^2 = \left(\frac{1}{3} - 4(1-\gamma) \right) \cdot \frac{\beta^2}{2},
\end{aligned}$$

which implies $\mu_H(T^3_g) \geq \frac{1}{3} - 4(1-\gamma)$. Hence $\mu_{N/24}(T^3_g) \geq 1 - 12(1-\gamma)$ because $H = \lfloor N/8 \rfloor$. Now the lemma is proven if we set $x_0 := 2H + 2l + \lfloor H/3 \rfloor$, $T := T^3_g$, and choose one $p_t \in \mathcal{P}_g$ such that $P_t(3) = t$ for each $t \in T$. $\square$

Remark 4.16 *The argument for showing $\mu_{N/24}(T^3_g) > 1 - 12(1-\gamma)$ is from [27, Page 34].*

Theorem 4.17 (E. Szemerédi, 1969) *If $U \subseteq \mathbb{N}_0$ and $SD(U) > 0$, then U contains nontrivial 4-term arithmetic progressions.*

Proof Let $N \gg 1$ and $A \subseteq [N]$ be such that $\mu_N(A) = SD(A) = \alpha > 0$. Same as in the beginning of the proof of Theorem 4.13, it suffices to find a 4–a.p. in A. For each $n, j \in \mathbb{N}_0$ let

$$S_{j,n} := \{x \in [N-n] \mid \mu_n((x + [n]) \cap A) \geq \alpha - 1/j\}.$$

Then $\lim_{n\to\infty} \mu_{N-n}(S_{j,n}) = 1$ by Lemma 4.11. So, for all sufficiently large $n \in \mathbb{N}_0$ we have that $\gamma_{j,n} := SD(S_{j,n}) > 11/12$. Let $R_{j,n}$ be an a.p. in $[N]$ with difference d and $|R_{j,n}| \gg 1$ such that $\mu_{|R_{j,n}|}(R_{j,n} \cap S_{j,n}) = \gamma_{j,n}$. For each $\tau \subseteq [n]$ let

$$B_{\tau,n} := \{x \in [R_{j,n}] \mid A \cap (x + [n]) = x + \tau\}.$$

Then there is a τ_j such that $\mu_{|R_{j,n}|}(B_{j,n}) = \beta_{j,n} > 0$ because n is finite where $B_{j,n} := B_{\tau_j,n}$. Let $P_{j,n} \subseteq R_{j,n}$ be an a.p. of difference $d' = dm$ for some positive integer m with $|P_{j,n}| = N' \gg 1$, $\mu_{N'}(P_{j,n} \cap S_{j,n}) = \gamma_{j,n}$, and $\mu_{N'}(P_{j,n} \cap B_{j,n}) = \beta_{j,n}$. Let $\varphi : P_{j,n} \to [N']$ be the affine map $\varphi(x) = (x - \min P_{j,n})/d' + 1$. Applying Lemma 4.15 to $[N']$ for $S' = \varphi((S_{j,n}) \cap P_{j,n})$, and $B' = \varphi(B_{j,n} \cap P_{j,n})$, and then pulling back through φ^{-1}, we obtain $x_0 + d'\,[\lfloor |P_{j,n}|/24 \rfloor] \subseteq P_{j,n}$ and $T_{j,n} \subseteq x_0 + d'\,[\lfloor |P_{j,n}|/24 \rfloor]$ with $\mu_{N'/24}(T_{j,n}) \geq 1 - 12(1 - \gamma_{j,n})$, and there exists a collection of 4–a.p.'s $\mathcal{P}_{j,n} = \{p_t \mid t \in T_{j,n}\}$ such that $p_t(1), p_t(2) \in B_{j,n} \cap P_{j,n}$, $p_t(3), p_t(4) \in S_{j,n} \cap P_{j,n}$, and $p_t(3) = t$ for each $t \in T_{j,n}$.

By countable saturation we can find fixed hyperfinite integer H and then J such that $\gamma := \gamma_{J,H} \approx 1$, $P := P_{J,H}$ with $|P| \gg 1$, $S := S_\gamma$, $B := B_{J,H} \subseteq S$, $T := T_{J,H}$, and $\mathcal{P}_{J,H} = \{p_t \mid t \in T\}$ such that $p_t(1), P_t(2) \in B$, $p_t(3), p_t(4) \in S$, and $p_t(3) = t$ for each $t \in T$.

Notice that $\mu_{N-H}(S) = 1$, $T \subseteq x_0 + d'\,[\lfloor |P|/24 \rfloor]$, $\mu_{|P|/24}(T) = 1$, $\gamma \approx 1$, $x, y \in B$ implies $((x+[H]) \cap A) - x = ((y+[H]) \cap A) - y$, and $x \in S$ implies $\mu_H((x+[H]) \cap A) = \alpha$. It may be the case that $\mu_{|P|}(B) = 0$. But the existence of the collection $\mathcal{P}_{J,H} = \{p_x \mid x \in T\}$ is guaranteed by countable saturation.

Since $\mu_{N/24}(T) = 1$, we can find an a.p. of $P' \subseteq T$ of difference d' with $|P'| \gg 1$. Let $\mathcal{P}' := \{p_t \in \mathcal{P}_{J,H} \mid t \in P'\}$. Notice that for each $p_t \in \mathcal{P}'$ we have that $p_t(1), p_t(2) \in B$, $p_t(3) = t \in S$, and $p_t(4) \in S$.

Let $\tau_0 := ((x + [H]) \cap A) - x$ for some $x \in B$. Then $\mu_H(\tau_0) = \alpha$ because $B \subseteq S$. By Lemma 4.14 with N being replaced by H, A being replaced by τ, we can find $x_0 + [\lfloor H/8 \rfloor] \subseteq [H]$, $T_Q \subseteq x_0 + [\lfloor H/8 \rfloor]$ with $\mu_H(T_Q) = 1/8$,

$$\mathcal{Q}_w := \{q \subseteq [H] \mid q(1), q(2) \in \tau_0, \text{ and } p(4) = w\},$$

and $E_w = \{q(3) \mid q \in \mathcal{Q}_w\}$ such that $\mu_H(E_w) = \alpha^2/24$ for each $w \in T_Q$.

By (iii) of Lemma 4.12 there is an $x' \in P'$ and $T'_Q \subseteq T_Q$ with $\mu_H(T'_Q) = 1/8$ such that $\mu_H((x' + E_w) \cap A) = \alpha\mu_H(E_w) = \alpha^3/24$ for each $w \in T'_Q$.

Fix $p_{x'} \in \mathcal{P}'$. Since $p_{x'}(4) \in S$, we have that $\mu_H((p_{x'}(4) + T'_Q) \cap A) = \alpha/8$. Hence there is a $w \in T'_Q$ such that $p_{x'}(4) + w \in A$. Let $q_w \in \mathcal{Q}_w$. Then $p_{x'}(4) + q_w(4) = p_{x'}(4) + w \in A$. Notice that $p_{x'}(3) + q_w(3) \in (x + E_w) \cap A \subseteq A$. Notice also that $p_{x'}(1), p_{x'}(2) \in B$ imply $A \cap (p_{x'}(i) + [\lfloor H/8 \rfloor]) = p_{x'}(i) + \tau_0$ for $i = 1, 2$. Hence

$p_{x'}(i) + q_w(i) \in p_{x'}(i) + \tau_0 \subseteq A$ for $i = 1, 2$. Therefore, $p_{x'} \oplus q_w$ is a nontrivial 4–a.p. in A. $\square$

Szemerédi's Theorem for all $k \geq 5$:

We work within $\mathcal{V}_2$ in this section except in the proof of Claim 1 in Lemma 4.18 where $\mathcal{V}_3$ is needed.

Szemerédi's Theorem is an easy consequence of Lemma 4.18, denoted by $\mathbf{L}(m)$ for all $m \in [k]$. For an integer $n \geq 2k + 1$ define an interval $C_n \subseteq [n]$ by

$$C_n := \left[\left\lceil \frac{kn}{2k+1} \right\rceil, \left\lfloor \frac{(k+1)n}{2k+1} \right\rfloor\right]. \tag{24}$$

The set C_n is the subinterval of $[n]$ in the middle of $[n]$ with the length $\lfloor n/(2k+1) \rfloor \pm \iota$ for $\iota = 0$ or 1. If $n \gg 1$, then $\mu_n(C_n) = 1/(2k + 1)$. For notational convenience we denote

$$D := 3k^3 \text{ and } \eta_0 := 1 - \frac{1}{D}. \tag{25}$$

♦: Fix a $K \in \mathbb{N}_1 \setminus \mathbb{N}_0$. The number K is the length of an interval which will play an important role in Lemma 4.18. Keeping K unchanged is one of the advantages from nonstandard analysis, which is unavailable in the standard setting.

If p is a k-a.p. and A is a set, we denote $p \oplus A$ for the sequence $\{p(l)+A \mid 1 \leq l \leq k\}$. If p, q are k-a.p.'s and A be a set, we denote $p \sqsubseteq q \oplus A$ for the statement that $p(l) \in q(l) + A$ for $1 \leq l \leq k$.

Lemma 4.18 ($\mathbf{L}(m)$) *Given any $\alpha > 0$, $\eta > \eta_0$, any $N \in \mathbb{N}_2 \setminus \mathbb{N}_1$, and any $A \subseteq S \subseteq [N]$ and $U \subseteq [N]$ with*

$$\mu_N(U) = 1, \mu_N(S) = SD(S) = \eta, \text{ and } \mu_N(A) = SD_S(A) = \alpha, \tag{26}$$

the following are true:

$\mathbf{L}_1(m)(\alpha, \eta, N, A, S, U, K)$: *There exists a k–a.p. $\vec{x} \subseteq U$ with $\vec{x} \oplus [K] \subseteq [N]$ satisfying the statement $(\forall n \in \mathbb{N}_0)\, \xi(\vec{x}(l), \alpha, \eta, A, S, U, K, n)$ for $l \in 1 + [k]$, where ξ is defined in (18), and there exist $T_l \subseteq C_K$ with $\mu_{|C_K|}(T_l) = 1$ where C_K is defined in (24) and $V_l \subseteq [K]$ with $\mu_K(V_l) = 1$ for every $l \geq m$, and collections of k–a.p.'s*

$$\begin{aligned} \mathcal{P} &:= \bigcup\{\mathcal{P}_{l,t} \mid t \in T_l \text{ and } l \geq m\} \text{ and} \\ \mathcal{Q} &:= \bigcup\{\mathcal{Q}_{l,v} \mid v \in V_l \text{ and } l \geq m\} \text{ such that} \end{aligned}$$

$$\mathcal{P}_{l,t} \subseteq \{p \sqsubseteq (\vec{x} \oplus [K]) \cap U \mid \forall l' < m\, (p(l') \in A) \text{ and } p(l) = \vec{x}(l) + t\} \tag{27}$$

satisfying $\mu_K(\mathcal{P}_{l,t}) = \alpha^{m-1}/k$ *for all* $l \geq m$ *and* $t \in T_l$*, and*

$$\mathcal{Q}_{l,v} = \{q \sqsubseteq \vec{x} \oplus [K] \mid \forall l' < m\,(q(l') \in A) \text{ and } q(l) = \vec{x}(l) + v\} \qquad (28)$$

satisfying $\mu_K(\mathcal{Q}_{l,v}) \leq \alpha^{m-1}$ *for all* $l \geq m$ *and* $v \in V_l$*.*

$\mathbf{L}_2(m)(\alpha, \eta, N, A, S, K)$*: There exist a set* $W_0 \subseteq S$ *of* $\min\{K, \lfloor 1/D(1-\eta)\rfloor\}$*–consecutive integers where* D *is defined in (25) and a collection of* k*–a.p.'s* $\mathcal{R} = \{r_w \mid w \in W_0\}$ *such that for each* $w \in W_0$ *we have* $r_w(l) \in A$ *for* $l < m$*,* $r_w(l) \in S$ *for* $l > m$*, and* $r_w(m) = w$*.*

Remark 4.19 *(a)* $\mathbf{L}_2(m)$ *is an internal statement in* $\mathcal{V}_2$*. Both* $\mathbf{L}_1(m)$ *and* $\mathbf{L}_2(m)$ *depend on* K*. Since* K *is fixed throughout whole proof, it, as a parameter, may be omitted in some expressions.*

(b) If $H \gg 1$ *and* $T \subseteq [H]$ *with* $\mu_H(T) > 1 - \epsilon$*, then* T *contains* $\lfloor 1/\epsilon \rfloor$ *consecutive integers because otherwise we have* $\mu_H(T) \leq (\lfloor 1/\epsilon\rfloor - 1)/\lfloor 1/\epsilon \rfloor = 1 - 1/\lfloor 1/\epsilon\rfloor \leq 1 - 1/(1/\epsilon) = 1 - \epsilon$*.*

(c) The purpose of defining C_K *is that if* $t \in C_K$*, then the number of* k*–a.p.'s* $p \sqsubseteq \vec{x} \oplus [K]$ *with* $p(l) = \vec{x}(l) + t$ *is guaranteed to be at least* $K/(k-1)$*.*

(d) It is not essential to require specific constant $c = 1/k$ *for* $\mu_K(\mathcal{P}_{l,t}) = c\alpha^{m-1}$ *in* $\mathbf{L}_1(m)$*. Just requiring that* $\mu_K(\mathcal{P}_{l,t}) \geq c\alpha^{m-1}$ *for some positive standard real* c *is sufficient. We use more specific expression "*$\mu_K(\mathcal{P}_{l,t}) = \alpha^{m-1}/k$*" for notational simplicity.*

(e) Some "bad" k*–a.p.'s in* $\mathcal{P}$ *in* $\mathbf{L}_1(m)$ *will be thinned out so that* $\mathcal{R}$ *in* $\mathbf{L}_2(m)$ *can be constructed from* $\mathcal{P}$*. The collection* $\mathcal{Q}$ *is only used to prevent* $\mathcal{P}$ *from being thinned out too much. See the proof of Lemma 4.20.*

(f) It is important to notice that in $\mathbf{L}_1(m)$ *the collection* $\mathcal{P}_{l,t}$ *is a part of the collection at the right side of (27) while the collection* $\mathcal{Q}_{l,v}$ *is equal to the collection at the right side of (28).*

The following lemma is a generalization of Lemma 4.15.

Lemma 4.20 $\mathbf{L}_1(m)(\alpha, \eta, N, A, S, U)$ *implies* $\mathbf{L}_2(m)(\alpha, \eta, N, A, S)$ *for any* α, η, N, A, S, U *satisfying the conditions of Lemma 4.18.*

Proof Assume we have obtained the k–a.p. $\vec{x} \subseteq U$ with $\vec{x} \oplus [K] \subseteq [N]$, sets $T_l \subseteq C_K$ and $V_l \subseteq [K]$ with $\mu_{|C_K|}(T_l) = 1$ and $\mu_K(V_l) = 1$, and collections of k–a.p.'s $\mathcal{P}$ and $\mathcal{Q}$ as in $\mathbf{L}_1(m)$.

Call a k–a.p. $p \in \mathcal{P}_m := \bigcup\{\mathcal{P}_{m,t} \mid t \in T_m\}$ *good* if $p(l) \in S \cap (\vec{x}(l) + [K])$ for $l \geq m$ and *bad* otherwise. Let $\mathcal{P}^g_m \subseteq \mathcal{P}_m$ be the collection of all good k–a.p.'s and $\mathcal{P}^b_m := \mathcal{P}_m \setminus \mathcal{P}^g_m$ be the collection of all bad k–a.p.'s. Let $T^g_m := \{p(m) - \vec{x}(m) \mid p \in \mathcal{P}^g_m\}$. Then $T^g_m \subseteq T_m \cap (S - \vec{x}(m)) \cap C_K$. We show that $\mu_{|C_K|}(T^g_m) > 1 - D(1-\eta)$.

Let $Q := \{q \sqsubseteq \vec{x} \oplus [K] \mid q(l') \in A \text{ for } l' < m\}$. Notice that

$$\mathcal{P}^b_m \subseteq \bigcup_{l \geq m} \{q \in Q \mid q(l) \notin S\}$$

and for each $v \in V_l$, $q \in \mathcal{Q}_{l,v}$ iff $q \in Q$ and $q(l) = \vec{x}(l) + v$.

$$\begin{aligned}
\text{Hence, } |\mathcal{P}^b_m| &\leq \sum_{l=m}^{k} \sum_{w \in [K]\setminus(S - \vec{x}(l))} |\{q \in Q \mid q(l) = \vec{x}(l) + w\}| \\
&\leq \sum_{l=m}^{k} \left(\sum_{w \in [K]\setminus V_l} |\{q \in Q \mid q(l) = \vec{x}(l) + w\}| + \sum_{v \in V_l \setminus (S - \vec{x}(l))} |\mathcal{Q}_{l,v}| \right) \\
&\leq K \sum_{l=m}^{k} (|[K] \setminus V_l| + |V_l \setminus (S - \vec{x}(l))| \alpha^{m-1}).
\end{aligned}$$

$$\text{So } |\mathcal{P}^g_m| = |\mathcal{P}_m| - |\mathcal{P}^b_m| \geq \sum_{t \in T_m} |\mathcal{P}_{m,t}| - K \sum_{l=m}^{k} (|[K] \setminus V_l| + |V_l \setminus (S - \vec{x}(l))| \alpha^{m-1}).$$

Notice that $\mu_K([K] \setminus V_l) = 0$. Hence we have $|[K] \setminus V_l|/|C_K| \approx 0$ and

$$\begin{aligned}
\mu_{|C_K|}(T^g_m) \cdot \frac{\alpha^{m-1}}{k} &= st\left(\frac{1}{|C_K|} \sum_{t \in T^g_m} \frac{1}{K} |\mathcal{P}_{m,t}| \right) \geq st\left(\frac{1}{|C_K| K} |\mathcal{P}^g_m| \right) \\
&\geq st\left(\frac{1}{|C_K|} \sum_{t \in T_m} \frac{1}{K} |\mathcal{P}_{m,t}| - \frac{1}{|C_K|} \sum_{l=m}^{k} (|V_l \setminus (S - \vec{x}(l))| \alpha^{m-1}) \right) \\
&\geq \mu_{|C_K|}(T_m) \cdot \frac{\alpha^{m-1}}{k} - (2k+1)k(1-\eta) \cdot \alpha^{m-1} \\
&= \left(\frac{1}{k} - (2k+1)k(1-\eta) \right) \cdot \alpha^{m-1},
\end{aligned}$$

which implies $\mu_{|C_K|}(T^g_m) \geq 1 - (2k+1)k^2(1-\eta) > 1 - D(1-\eta)$. Recall that $T^g_m \subseteq C_K$. Hence $\vec{x}(m) + T^g_m$ contains a set W_0 of $\lfloor 1/D(1-\eta) \rfloor$ consecutive integers. So, $\mathbf{L}_2(m)$

is proven if we let $\mathcal{R} := \{r_w \mid w \in W_0\}$ where r_w is one of the k–a.p.'s in $\mathcal{P}^g_m$ such that $r_w(m) = \vec{x}(m) + w$. $\square$

Proof of Lemma 4.18 We prove $\mathbf{L}(m)$ by induction on m. By Lemmas 4.20 it suffices to prove $\mathbf{L}_1(m)$.

For $\mathbf{L}(1)$, given any $\alpha > 0$, $\eta > \eta_0$, $N \in \mathbb{N}_2 \setminus \mathbb{N}_1$, A, S, and U satisfying (26), by Lemma 4.11 (b) we can find a k–a.p. $\vec{x} \subseteq [N]$ such that

$$(\forall n \in \mathbb{N}_0)\, \xi(\vec{x}(l), \alpha, \eta, A, S, U, K, n)$$

is true for $l \in 1 + [k]$, where ξ is defined in (18). For each $l \in 1 + [k]$ let $T_l = C_K \cap U$ and $V_l = [K]$. For each $l \in 1 + [k]$, $t \in T_l$, and $v \in V_l$ let

$$\begin{aligned} \mathcal{P}_{l,t} &:= \{p \sqsubseteq (\vec{x} \oplus [K]) \cap U \mid p(l) = \vec{x}(l) + t\} \\ \mathcal{Q}_{l,v} &:= \{q \sqsubseteq (\vec{x} \oplus [K]) \mid q(l) = \vec{x}(l) + v\}. \end{aligned}$$

Clearly, we have $\mu_K(\mathcal{P}_{l,t}) \geq 1/(k-1) > 1/k$. By some pruning we can assume that $\mu_K(\mathcal{P}_{l,t}) = 1/k$. It is trivial that $\mu_K(\mathcal{Q}_{l,v}) \leq 1$ and $q \in \mathcal{Q}_{l,v}$ iff $q(l) = \vec{x}(l) + v$ for each $q \sqsubseteq \vec{x} \oplus [K]$. This completes the proof of $\mathbf{L}_1(1)(\alpha, \eta, N, A, S, U)$. $\mathbf{L}_2(1)(\alpha, \eta, N, A, S)$ follows from Lemma 4.20.

Assume $\mathbf{L}(m-1)$ is true for some $2 \leq m \leq k$.

We now prove $\mathbf{L}(m)$. Given any $\alpha > 0$ and $\eta > \eta_0$, fix $N \in \mathbb{N}_2 \setminus \mathbb{N}_1$, $U \subseteq [N]$, and $A \subseteq S \subseteq [N]$ satisfying (26). For each $n \in \mathbb{N}_1 \setminus \mathbb{N}_0$, by Lemma 4.11 (b), there is an $h_n > n$ in $\mathbb{N}_1$ and $G_{n,h_n} \subseteq [N]$ defined in (19) such that $d_n := \delta_{N-h_n}(G_{n,h_n}) > 1 - 1/n$. Notice that $d_n \approx_1 \mu^1_{N-h_n}(G_{n,h_n}) > \eta_0$ because $n \gg 1$ and $\mu_{N-h_n}(G_{n,h_n})) = 1$. Let $\eta^1_n := \mu^1_{N-h_n}(G_{n,h_n})$ and fix an $n \in \mathbb{N}_1 \setminus \mathbb{N}_0$.

Claim 1 The following internal statement $\theta(n, A, N)$ is true:
$\exists W \subseteq [N]\, \exists \mathcal{R}\, (W$ is an a.p. $\wedge\, |W| \geq \min\{K, \lfloor 1/2D(1-d_n) \rfloor\} \wedge \mathcal{R} = \{r_w \mid w \in W\}$ is a collection of k–a.p.'s such that

$$\begin{aligned} &\forall w \in W\, ((\forall l \geq m)\, (r_w(l) \in G_{n,h_n}) \wedge (r_w(m-1) = w) \\ &\quad \wedge (\forall l, l' \leq m-2)((A \cap (r_w(l) + [h_n])) - r_w(l) = (A \cap (r_w(l') + [h_n])) - r_w(l'))). \end{aligned}$$

Proof of Claim 1 Working in $\mathcal{V}_2$ by considering $\mathcal{V}_1$ as the standard universe, we can find $P \subseteq [N]$ with $|P| \in \mathbb{N}_2 \setminus \mathbb{N}_1$ by Lemma 4.9 and Proposition 4.1 such that

$$SD^1(G_{n,h_n}) = \mu^1_{|P|}(P \cap G_{n,h_n}) = SD^1(G_{n,h_n} \cap P) = \eta^1_n.$$

For each $x \in P \cap G_{n,h_n}$ let $\tau_x = ((x + [h_n]) \cap A) - x$. Since there are at most $2^{h_n} \in \mathbb{N}_1$ different τ_x's and $|P| \gg 2^{h_n}$, we can find one, say, $\tau_n \subseteq [h_n]$ such that the set

$$B_n := \{x \in P \cap G_{n,h_n} \mid \tau_x = \tau_n\}$$

satisfies $\mu^1_{|P|}(B_n) \geq \eta^1_n/2^{h_n} > 0$. Notice that $\mu^0_{|P|}(B_n)$ could be 0.

Let $P' \subseteq P$ with $|P'| = N' \in \mathbb{N}_2 \setminus \mathbb{N}_1$ be such that $\mu^1_{N'}(G_{n,h_n} \cap P') = \eta^1_n$ and

$$\begin{aligned} \beta^1_n &:= \mu^1_{N'}(B_n \cap P') = SD^1_{G_{n,h_n} \cap P}(B_n \cap P) \\ &= SD^1_{G_{n,h_n} \cap P'}(B_n \cap P') \geq \mu^1_{|P|}(B_n) > 0 \end{aligned}$$

by Part 4 of Lemma 4.9. Let d be the common difference of the a.p. P' and $\varphi : P' \to [N']$ be the order-preserving bijection, i.e.,

$$\varphi(x) := 1 + (x - \min P')/d.$$

Let $B' := \varphi[B_n \cap P']$ and $S' := \varphi[G_{n,h_n} \cap P']$. We have that B', S', N' and β^1_n, η^1_n in the place of A, S, N and α, η satisfy the $\mathcal{V}_1$–version of (26) with μ, SD and SD_S being replaced by μ^1, SD^1, and $SD^1_{S'}$.

Let $N'' = i_{1,2}(N')$, $B'' = i_{1,2}(B')$, and $S'' = i_{1,2}(S')$ where $i_{1,2}$ is described in Proposition 4.1. Recall that $i_{1,2} \restriction \mathbb{R}_1$ is an identity map. Since $N' \in \mathbb{N}_2 \setminus \mathbb{N}_1$, we have $N'' \in \mathbb{N}_3 \setminus \mathbb{N}_2$. Notice also that $\mu^1_{N''}(S'') = SD^1(S'') = \eta^1_n$ and $\mu^1_{N''}(B'') = SD^1_{S''}(B'') = \beta^1_n$. By the induction hypothesis that $\mathbf{L}(m-1)$ is true we have

$$\begin{aligned} &(\mathcal{V}_2; \mathbb{R}_0, \mathbb{R}_1) \models \forall \alpha, \eta \in \mathbb{R}_0 \, \forall N \in \mathbb{N}_2 \setminus \mathbb{N}_1 \, \forall A, S \subseteq [N] \\ &\quad (\alpha > 0 \wedge \eta > \eta_0 \wedge A \subseteq S \wedge \mu_N(S) = SD(S) = \eta \wedge \mu_N(A) = SD_S(A) \\ &\quad \to \mathbf{L}_2(m-1)(\alpha, \eta, N, A, S)). \end{aligned} \tag{29}$$

Since $(\mathcal{V}_2; \mathbb{R}_0, \mathbb{R}_1)$ and $(\mathcal{V}_3; \mathbb{R}_1, \mathbb{R}_2)$ are elementarily equivalent by Part 4 of Proposition 4.1 via $i_{0,2}$, we have, by universal instantiation, that

$$(\mathcal{V}_3; \mathbb{R}_1, \mathbb{R}_2) \models \mathbf{L}_2(m-1)(\beta^1_n, \eta^1_n, N'', B'', S''). \tag{30}$$

Notice that the right side above no longer depends on $\mathbb{R}_1$ or $\mathbb{R}_2$. So, we have

$$\mathcal{V}_3 \models \mathbf{L}_2(m-1)(i_{1,2}(\beta^1_n), i_{1,2}(\eta^1_n), i_{1,2}(N'), i_{1,2}(B'), i_{1,2}(S')) \tag{31}$$

because $i_{1,2}(\beta^1_n) = \beta^1_n$ and $i_{1,2}(\eta^1_n) = \eta^1_n$. Since $i_{1,2}$ is an elementary embedding, we have

$$\mathcal{V}_2 \models \mathbf{L}_2(m-1)(\beta^1_n, \eta^1_n, N', B', S'),$$

which means that there is a set $W' \subseteq [N']$ of $\min\{K, \lfloor 1/D(1-\eta^1_n) \rfloor\}$–consecutive integers and a collection of k–a.p.'s $\mathcal{R}' = \{r'_w \mid w \in W'\}$ such that for every $w \in W'$ we have $r'_w(l) \in B'$ for $l < m-1$, $r'_w(m-1) = w$, and $r'_w(l) \in S'$ for $l \geq m$. Notice that $\varphi^{-1}[[N']] \subseteq [N]$. Let $W = \varphi^{-1}[W']$ and $\mathcal{R} = \{r_w \mid w \in W\}$, where $r_w = \varphi^{-1}[r'_{\varphi(w)}]$, such that for each $w \in W$ we have $r_w(l) \in \varphi^{-1}[B'] \subseteq B_n$ for $l < m-1$, $r_w(m-1) = w$, and $r_w(l) \in \varphi^{-1}[S'] \subseteq G_{n,h_n}$ for $l \geq m$. If $\eta_n^1 = 1$, then $|W| \geq K$. If $\eta_n^1 < 1$, then $2(1-d_n) > 1 - \eta_n^1$. Hence $|W| \geq \min\{K, \lfloor 1/2D(1-d_n)\rfloor\}$. □ (Claim 1)

The following claim follows from Claim 1 by Proposition 2.4.

Claim 2 There exists a $J \in \mathbb{N}_2 \setminus \mathbb{N}_1$ such that the $\theta(J, A, N)$ is true, i.e., $\exists W \subseteq [N]\, \exists \mathcal{R}\, (W$ is an a.p. $\wedge\, |W| \geq \min\{K, \lfloor 1/2D(1-d_J)\rfloor\} \wedge \mathcal{R} = \{r_w \mid w \in W\}$ is a collection of k–a.p.'s such that $\forall w \in W\, ((\forall l \geq m)\, (r_w(l) \in G_{J,h_J}),\ r_w(m-1) = w$, and $(\forall l, l' \leq m-2)\, ((A \cap (r_w(l) + [h_J])) - r_w(l) = (A \cap (r_w(l') + [h_J])) - r_w(l'))))$.

For notational convenience let $W_H := W$ and $\mathcal{R}_H := \mathcal{R}$ be obtained in Claim 2 and rename $H := h_J$, $S_H := G_{J,h_J}$, $\tau_H := (A \cap (r_w(l) + [h_J])) - r_w(l)$ for some (or any) $w \in W_H$ and $l < m-1$. Let $\{w_s \mid 1 \leq s \leq |W_H|\}$ be the increasing enumeration of W_H. Notice that $H \in \mathbb{N}_2 \setminus \mathbb{N}_1$. We now go back to consider $\mathcal{V}_0$ as our standard universe. Notice that $\mu_{N-H}(S_H) = 1$, $|W_H| \gg 1$, and $(\forall n \in \mathbb{N}_0)\, \xi(x, \alpha, \eta, A, S, U, H, n)$ is true for every $x \in S_H$ where ξ is defined in (18).

Claim 3 For each $s \in \mathbb{N}_0$ we can find an internal $U_s \subseteq [H]$ with $\mu_H(U_s) = 1$ such that for each $y \in U_s$ and each $l \in 1 + [k]$, $r_{w_s}(l) + y \in U$ and $(\forall n \in \mathbb{N}_0)\, \xi(r_{w_s}(l) + y, \alpha, \eta, A, S, U, K, n)$ is true.

Proof of Claim 3 For each $l \in 1 + [k]$ we have $\xi(r_{w_s}(l), \alpha, \eta, A, S, U, H, n)$ is true because $r_{w_s}(l) \in S_H$. By Lemma 4.11 (a), we can find a set $G_l \subseteq r_{w_s}(l) + [H]$ with $\mu_H(G_l) = 1$ such that $\xi(r_{w_s}(l) + y, \alpha, \eta, A, S, U, K, n)$ is true for every $r_{w_s}(l) + y \in G_l$. Set

$$U_s := \bigcap_{l=1}^{k} ((U \cap G_l) - r_{w_s}(l)).$$

Then we have $U_s \subseteq [H]$ and $\mu_H(U_s) = 1$. □ (Claim 3)

Notice that $\delta_H(\bigcap_{i=1}^{s} U_i) > 1 - 1/s$. By Proposition 2.4 we can find $1 \ll I \leq |W_H|$ and

$$U' := \bigcap \{U_s \mid 1 \leq s \leq I\}$$

such that $\delta_H(U') > 1 - 1/I$. Hence $\mu_H(U') = 1$. Applying the induction hypothesis for $\mathbf{L}_1(m-1)(\alpha, 1, H, \tau_H, [H], U')$, we obtain a k–a.p. $\vec{y} \subseteq U'$ with $\vec{y} \oplus [K] \subseteq [H]$, $T'_l \subseteq C_K \cap U'$ with $\mu_{|C_K|}(T'_l) = 1$ and $V'_l \subseteq [K]$ with $\mu_K(V'_l) = 1$ for each $l \geq m-1$, and collections of k–a.p.'s

$$\begin{aligned} \mathcal{P}' &= \bigcup \{\mathcal{P}'_{l,t} \mid t \in T'_l \text{ and } l \geq m-1\} \text{ and} \\ \mathcal{Q}' &= \bigcup \{\mathcal{Q}'_{l,v} \mid v \in V'_l \text{ and } l \geq m-1\} \end{aligned}$$

such that (i) for each $l \geq m-1$ and $t \in T'_l$ we have $\mu_K(\mathcal{P}'_{l,t}) = \alpha^{m-2}/k$ and for each $p \in \mathcal{P}'_{l,t}$ we have $p \sqsubseteq (\vec{y} \oplus [K]) \cap U'$, $p(l') \in \tau_H$ for $l' < m-1$, $p(l) = \vec{y}(l) + t$, and (ii) for each $l \geq m-1$ and $v \in V'_l$ we have $\mu_K(\mathcal{Q}'_{l,v}) \leq \alpha^{m-2}$, and for each $q \sqsubseteq \vec{y} \oplus [K]$ we have $q \in \mathcal{Q}'_{l,v}$ iff $q(l') \in \tau_H$ for every $l' < m-1$ and $q(l) = \vec{y}(l) + v$. For each $l \geq m$, $t \in T_l$, and $v \in V_l$ let

$$E_{l,t} := \{p(m-1) \mid p \in \mathcal{P}'_{l,t}\} \text{ and } F_{l,v} := \{q(m-1) \mid q \in \mathcal{Q}'_{l,v}\}.$$

Then $E_{l,t}, F_{l,v} \subseteq \vec{y}(m-1) + [K]$, $\mu_K(E_{l,t}) = \mu_K(\mathcal{P}'_{l,t}) = \alpha^{m-2}/k$, and $\mu_K(F_{l,v}) = \mu_K(\mathcal{Q}'_{l,v}) \leq \alpha^{m-2}$. Since $\vec{y} \subseteq U'$ we have that for each $l \in 1+[k]$, $(\forall n \in \mathbb{N}_0)\, \xi(r_{w_s}(l) + \vec{y}(l), \alpha, \eta, A, S, U, K, n)$ is true.

Applying Part (iii) of Lemma 4.12 with $R := \{w_s + \vec{y}(m-1) \mid 1 \leq s \leq I\}$ and H being replaced by K we can find $s_0 \in [I]$, $T_l \subseteq T'_l$ with $\mu_{|C_K|}(T_l) = 1$ and $V_l \subseteq V'_l$ with $\mu_K(V_l) = 1$ for each $l \geq m$ such that for each $t \in T_l$ and $v \in V_l$ we have

$$\begin{aligned} &\mu_K((w_{s_0} + E_{l,t}) \cap ((w_{s_0} + \vec{y}(m-1) + [K]) \cap A)) \\ &\qquad = \alpha \mu_K(E_{l,t}) = \alpha(\alpha^{m-2}/k) = \alpha^{m-1}/k \text{ and} \end{aligned} \tag{32}$$

$$\begin{aligned} &\mu_K((w_{s_0} + F_{l,v}) \cap ((w_{s_0} + \vec{y}(m-1) + [K]) \cap A)) \\ &\qquad = \alpha \mu_K(F_{l,t}) \leq \alpha \cdot \alpha^{m-2} = \alpha^{m-1}. \end{aligned} \tag{33}$$

Let $\vec{x} := r_{w_{s_0}} \oplus \vec{y}$. Clearly, we have $\vec{x} \oplus [K] \subseteq [N]$. We also have that $\vec{x} \subseteq U$, $\mu_K((\vec{x}(l) + [K]) \cap S) = \eta$, and $\mu_K((\vec{x}(l) + [K]) \cap A) = \alpha$ because $r_{w_{s_0}} \subseteq S_H$ and $\vec{y} \subseteq U' \subseteq U_{s_0}$. For each $l \geq m$, $t \in T_l$, and $v \in V_l$ let

$$\begin{aligned} \mathcal{P}_{l,t} := \{&r_{w_{s_0}} \oplus p \mid p \in \mathcal{P}'_{l,t} \text{ and} \\ &p(m-1) \in E_{l,t} \cap (((w_{s_0} + \vec{y}(m-1) + [K]) \cap A) - w_{s_0})\}, \\ \mathcal{Q}_{l,v} := \{&r_{w_{s_0}} \oplus q \mid q \in \mathcal{Q}'_{l,t} \text{ and} \\ &q(m-1) \in F_{l,v} \cap (((w_{s_0} + \vec{y}(m-1) + [K]) \cap A) - w_{s_0})\}. \end{aligned}$$

Then $\mu_K(\mathcal{P}_{l,t}) = \alpha^{m-1}/k$ by (32). If $\bar{q} \sqsubseteq \vec{x} \oplus [K]$, then there is a $q \sqsubseteq \vec{y} \oplus [K]$ such that $\bar{q} = r_{w_{s_0}} \oplus q$. If $\bar{q}(l') \in A$ for $l' < m$ and $v \in V_l$ for some $l \geq m$ such that $\bar{q}(l) = \vec{x}(l) + v$, then $q(l') \in \tau_H$ for $l' < m-1$, $v \in V'_l$, and $q(l) = \vec{y}(l) + v$, which imply $q \in \mathcal{Q}'_{l,v}$ by induction hypothesis. Hence we have $q(m-1) \in F_{l,v}$. Clearly, $\bar{q}(m-1) = w_{s_0} + q(m-1) \in A$ implies $q(m-1) \in F_{l,v} \cap (((w_{s_0} + \vec{y}(m-1) + [K]) \cap A) - w_{s_0})$. Thus we have $\bar{q} \in \mathcal{Q}_{l,v}$. Clearly, $\mu_K(\mathcal{Q}_{l,v}) \leq \alpha^{m-1}$ by (33).

Summarizing the argument above we have that for each $r_{w_{s_0}} \oplus p \in \mathcal{P}_{l,t}$

- $r_{w_{s_0}}(l') + p(l') \in r_{w_{s_0}}(l') + \tau_H \subseteq A$ for $l' < m-1$ because $r_{w_{s_0}}(l') \in B_H$,
- $r_{w_{s_0}}(m-1) + p(m-1) = w_{s_0} + p(m-1)$
  $\in (w_{s_0} + E_{l,t}) \cap (w_{s_0} + \vec{y}(m-1) + [K]) \cap A \subseteq A,$

- $r_{w_{s_0}}(l') + p(l') \in (\vec{x}(l') + [K]) \cap U \subseteq U$ for $l' \geq m$ because of $p \subseteq U'$,
- $r_{w_{s_0}}(l) + p(l) = r_{w_{s_0}}(l) + \vec{y}(l) + t = \vec{x}(l) + t$.

For each $\bar{q} \sqsubseteq \vec{x} \oplus [K]$, $\bar{q} \in \mathcal{Q}_{l,v}$ iff there is a $q \sqsubseteq \vec{y} \oplus [K]$ with $\bar{q} = r_{w_{s_0}} \oplus q$ such that

- $r_{w_{s_0}}(l') + q(l') \in r_{w_{s_0}}(l') + \tau_H \subseteq A$ for $l' < m - 1$ because $r_{w_{s_0}}(l') \in B_H$,
- $r_{w_{s_0}}(m-1) + q(m-1) = w_{s_0} + q(m-1) \in A$ which is equivalent to
$$w_{s_0} + q(m-1) \in (w_{s_0} + F_{l,v}) \cap (w_{s_0} + \vec{y}(m-1) + [K]) \cap A \subseteq A,$$
- $r_{w_{s_0}}(l) + q(l) = r_{w_{s_0}}(l) + \vec{y}(l) + v = \vec{x}(l) + v$.

This completes the proof of $\mathbf{L}_1(m)(\alpha, \eta, N, A, S, U)$ as well as $\mathbf{L}(m)$ by Lemma 4.20.
□

Theorem 4.21 (E. Szemerédi, 1975) *Let $k \in \mathbb{N}_0$. If $D \subseteq \mathbb{N}_0$ has positive upper density, then D contains nontrivial k-term arithmetic progressions.*

Proof It suffices to find a nontrivial k–a.p. in $i_0(D)$. Let P be an a.p. such that $|P| \gg 1$ and $\mu_{|P|}(i_0(D) \cap P) = SD(D) = \alpha$. Then $\alpha > 0$ because α is greater than or equal to the upper density of D. Let $A = i_0(D) \cap P$. Without loss of generality, we can assume $P = [N]$ for some $N \gg 1$. We can also assume that $N \in \mathbb{N}_2 \setminus \mathbb{N}_1$ because otherwise replace N by $i_1(N)$ and A by $i_1(A)$. Then we have $\mu_N(A) = SD(A) = \alpha$. Set $U = S = [N]$. Trivially, $\mu_N(S) = SD(S) = \eta = 1$, $A \subseteq S$, and $SD_S(A) = SD(A) = \alpha$. To start with $k' = k+1$ instead of k, we have many k'–a.p.'s $p \in \mathcal{P}$ such that $p(l) \in A$ for $l \leq k' - 1 = k$ in $\mathbf{L}_1(k')$. So there must be many nontrivial k–a.p.'s in $A \subseteq i_0(D)$. By $\mathcal{V}_0 \prec \mathcal{V}_2$, there must be nontrivial k–a.p.'s in D. □

4.4 Exercises

1. Prove that the map $i_{1,1} : \mathcal{V}_1 \to \mathcal{V}_2$ defined in Definition 4.2 is an elementary embedding.
2. Let $s \in \mathbb{N}_0$ and $\Gamma(\overline{x})$ be a countable collection of formulas with parameters from $\mathcal{V}_0$ and s free variables $\overline{x} = (x_1, x_2, \ldots, x_s)$. Prove that there exists a homothetic copy $HC_{\vec{a},d}$ of $[N]^s$ for some $N \in \mathbb{N}_1 \setminus \mathbb{N}_0$ such that for any $0 \leq l, l' < N^s$
$$\mathcal{V}_1 \models \varphi(HC_{\vec{a},d}(l)) \Longleftrightarrow \mathcal{V}_1 \models \varphi(HC_{\vec{a},d}(l'))$$
for every formula φ in Γ.
3. Let $r \in \mathbb{R}_2$ and $|r| < \alpha$ for some $\alpha \in \mathbb{R}_0$. Prove that $st(st_1(r)) = st(r)$.

4. Prove that the set

$$C := \left\{\mu^1_{|P|}(A \cap P) \mid P \text{ is an } a.p. \text{ and } |P| \in \mathbb{N}_2 \setminus \mathbb{N}_1\right\}$$

is an element in $\mathcal{V}_1$ where $A \subseteq \mathbb{N}_2$ is $\mathcal{V}_2$-internal.

References

[1] C. C. Chang and H. J. Keisler, *Model Theory*, 3rd ed., 1990, North-Holland, Amsterdam

[2] M. Di Nasso, *An elementary proof of Jin's Theorem with a bound*, The Electronic Jour. of Combinatorics, 21 (2014), Issue 2.

[3] M. Di Nasso, *Hypernatural numbers as ultrafilters*, in Nonstandard Analysis for the Working Mathematician–Second edition, edited by P. Loeb and M. Wolff, Springer, 2015, 443 – 474.

[4] M. Di Nasso, *Iterated hyper-extensions and an idempotent ultrafilter proof of Rado's Theorem*, Proceedings of the American Mathematical Society, vol. 143 (2015), 1749 – 1761.

[5] M. Di Nasso and L. Luperi Baglini, *Ramsey properties of nonlinear Diophantine equations*, Advances in Mathematics, 324 (2018), 84 – 117.

[6] M. Di Nasso and M. Ragosta, *Central sets and infinite monochromatic exponential patterns*, arXiv: 2211.16269

[7] T. Gowers, *A new proof of Szemerédi's Theorem*, GAFA 11 (2001), 465 – 588.

[8] R. Graham, B. Rothschild, and J. Spencer, *Ramsey Theory*, (2nd edition), Wiley, New York, 1990.

[9] H. Halberstam and K. F. Roth, *Sequences*, Oxford University Press, 1966.

[10] R. Heath-Brown, *Sums of three square-full numbers*, Number Theory, I (Budapest, 1987), Colloq. Math. Soc. János Bolyai, no. 51. 163 - 171.

[11] N. Hindman, *Finite sums from sequences within cells of a partition of* $\mathbb{N}$, Journal of Combinatorial Theory, Series A. (1974) 17 (1): 1 - 11.

[12] R. Jin, *Nonstandard methods for upper Banach density problems*, The Journal of Number Theory, 91 (2001), 20 – 38.

[13] R. Jin, *Density Problems and Freiman's Inverse Problems* in Nonstandard Analysis for the Working Mathematician–Second edition, edited by P. Loeb and M. Wolff, Springer, 2015, 403 – 441.

[14] R. Jin, *Plünnecke's Theorem for other densities*, The Transactions of American Mathematical Society, **363** (2011), 5059 – 5070.

[15] R. Jin, *Density version of Plunnecke inequality – Epsilon-Delta Approach*, Combinatorial and Additive Number Theory, CANT 2011-2012, Springer Proceedings in Mathematics and Statistics, Vol. 101, M. Natheanson edited (2014).

[16] R. Jin, *A Simple Combinatorial Proof of Szemeredi's Theorem via Three Levels of Infinities*, Discrete Analysis, to appear.

[17] A. I. Khinchin, *Three pearls of number theory*, Translated from the 2d (1948) rev. Russian ed. by F. Bagemihl, H. Komm, and W. Seidel, Rochester, N.Y., Graylock Press, 1952.

[18] P. Loeb and M. Wolff, editors, *Nonstandard Analysis for the Working Mathematician*, Second edition, Springer Netherlands, 2015.

[19] W. A. J. Luxemburg, *A general theory of monads*, Applications of Model Theory to Algebra, Analysis and Probability, (W. A. J. Luxemburg, editor, 1969), Holt, Rinehart and Winston, New York, 18 – 86.

[20] M. B. Nathanson, *Additive Number Theory–the Classical Bases*, Springer, 1996.

[21] M. B. Nathanson, *Additive Number Theory–Inverse Problems and the Geometry of Sumsets*, Springer, 1996.

[22] K. Petersen, *Ergodic Theory*, Cambridge University Press, 1989.

[23] A. Robinson, *Nonstandard Analysis*, Princeton University Press, revised edition. Originally published by North-Holland, Amsterdam (1966) and (1974).

[24] A. Robinson and E. Zakon, *A set-theoretical characterization of enlargements*, in: Applications of Model Theory to Algebra, Analysis, and Probability, W. A. J. Luxemburg (ed.), Holt, Rinehart, and Winston, New York, 1969, 109 – 122.

[25] D. Ross, *Loeb Measure and Probability, in Nonstandard Analysis: Theory and Applications*, edited by N. J. Cutland, C. W. Henson, and L. Arkeryd, Kluwer Academic Publishers 1997, 91 – 120.

[26] E. Szemerédi, *On sets of integers containing no k elements in aritheoremetic progression.* Collection of articles in memory of Juriĭ Vladimirovič Linnik, Acta Arith. 27 (1975), 199 – 245.

[27] T. Tao, *Szemerédi's proof of Szemerédi's Theorem*, Acta Math. Hungar. 161 (2020), 443 – 487.
https://terrytao.files.wordpress.com/2017/09/szemeredi-proof1.pdf

[28] T. Tao, *A nonstandard analysis proof of Szemerédi's Theorem*,
https://terrytao.wordpress.com/2015/07/20/
a-nonstandard-analysis-proof-of-szemeredis-theorem/

[29] T. Tao, *Szemeredi's proof of Szemeredi's Theorem*,
https://terrytao.wordpress.com/2017/09/12/
szemeredis-proof-of-szemeredis-theorem/