Ultraproducts and the Foundations of Higher Order Fourier Analysis

Evan Warner
Advisor: Nicolas Templier

Submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Arts
Department of Mathematics
Princeton University

May 2012

I hereby declare that this thesis represents my own work in accordance with University regulations.

Evan Warner

Contents

1. Introduction

1.1. **Higher order Fourier analysis.** The subject of higher order Fourier analysis had its genesis in a seminal paper of Gowers, in which the Gowers norms U^k for functions on finite groups, where k is a positive integer, were introduced. These norms were used to produce an alternative proof of Szemerédi's theorem (see [12] and [13]) on the existence of arbitrarily long arithmetic progressions in sets of integers of positive density.[1] Briefly, while the U^2 norm measures noise from a "Fourier theoretic" point of view — that is, correlations with linear characters — there is substantial amount of structure uncaptured by this perspective. As it turns out, for example, the U^3 norm encapsulates a function's correlations with "quadratic characters," and a function can be structured from this perspective (i.e., possessing a large U^3 norm) while being random from a Fourier-analytic perspective (i.e., possessing a small U^2 norm, which is equivalent to having a small maximum Fourier coefficient).

In the context of Gowers' proof and in additive combinatorics more generally, these higher-order correlations serve to keep track of arithmetic progressions in sets. In the $k = 2$ setting of ordinary Fourier analysis, one can manage progressions of length three, which is reflected in the fact that there is a Fourier-analytic proof of Szemerédi's theorem for progressions of length three (also known as Roth's theorem; see [33] for a proof in this context). Unsurprisingly, in order to manage progressions of arbitrary length, Gowers employed the U^k norms for arbitrarily high k. Although the proof did not use higher order Fourier analysis directly, the famous result of Tao and Green on the existence of arbitrarily large arithmetic progressions of primes [14] was heavily influenced by similar ideas. The subject has subsequently benefited from an influx of ergodic theory, starting with Host and Kra [17].

More recently, Balazs Szegedy has put the subject of higher order Fourier analysis on a more algebraic footing, relying heavily on a novel ultraproduct construction that was first developed in the context of hypergraph regularization (see [8]). In a series of three papers [27], [28], [29], Szegedy laid out a framework based around certain "higher order Fourier" σ-algebras $\mathcal{F}_k$ defined on the ultraproduct of a sequence of finite groups. The ultraproduct construction is employed because a satisfactory algebraic theory of higher order Fourier analysis, in contrast to ordinary Fourier analysis, seems to appear only in the limit of larger and larger ordinary structures. Although this avenue will not be pursued in this paper, one of the great strengths of the ultraproduct theory is that nearly everything proven in the setting of the ultraproduct has a corresponding version in the finite setting, usually at the cost of exactness (for example, replacing precise equalities with ϵ-close approximations). Thus, a development in the ultraproduct setting gives rise to an analogous development in standard models. More recently in [31], Szegedy took the theory in a different direction when he proved that in a certain sense the higher-order structure arises from algebraic structures known as nilspaces.

1.2. **Ultraproduct analysis.** First developed in a model-theoretic context in the 1950s by Tarski, Łoś, and others, the ultraproduct offers a method of taking very general limits objects of first-order logical structures. In a manner analogous to the construction of the reals as equivalence classes of Cauchy sequences modulo the relation of "difference tending to zero," an ultraproduct (or ultrapower, as it is called if all the original structures are equal) is

[1] In reference to the U^k norms in his original paper, Gowers wrote, amusingly, "Although I cannot think of any potential applications, I still feel that it would be interesting to investigate them further."

constructed as the set of equivalence classes of *arbitrary* sequences modulo the much stricter relation of "being equal on an ultrafilter." Thus an infinite ultrapower of $\mathbb{R}$ will contain infinite and infinitesimal element, constructed as, respectively, equivalence classes of sequences of real numbers which tend to infinity and tend to zero (without eventually becoming zero). Models constructed using ultraproducts have two very desirable properties. First, there is a transfer principle, known as Łoś' theorem, that allows one to deduce statements in first-order logic in the ultraproduct from the corresponding statements in its constituent structures, and vice-versa. Second, such models turn out to be saturated, which in the context of this paper manifests itself mostly in a countable compactness result. Ultrapower models with these two properties, especially ultrapowers of the real numbers, are often called nonstandard models. Such models were exploited beginning with Robinson to provide a rigorous underpinning for the naïve manipulation of infinitesimals in calculus (see [24]). To borrow an analogy from [33], the distinction between ultraproduct analysis and nonstandard analysis is similar to the distinction between measure theory and probability theory (or more elementarily, the distinction between matrices and linear transformations); in the former areas one constructs all objects "by hand" in the same context as the base objects, while in the latter one proceeds axiomatically (but largely equivalently, in that results in one context are easily translated into the other context). In this paper, we will proceed exclusively in the framework of ultraproduct analysis, but intuition from nonstandard theory is often helpful.

In 1975, Loeb [21] introduced a useful construction of ordinary measure spaces on ultraproducts, now sometimes known as Loeb measures, that are used to great effect in Szegedy's construction of algebraic higher order Fourier analysis. Besides the current application, Loeb measures have utility in a variety of contexts, including constructing measure spaces with various desirable properties, representing measures (such as Wiener measure) in a more manageable way, and proving existence results in analysis (especially in the field of stochastic differential equations).

Nonstandard or ultraproduct analysis is also often useful in simplifying proofs in analysis or suggesting new ones, especially in arguments requiring the juggling of many "ϵ-small quantities" that disappear or are simplified in the ultraproduct. The aforementioned Green-Tao theorem [14], for example, employs ultraproduct analysis in this context. The book [33] uses ultraproducts to develop results in quantitative algebraic geometry. As a motivating example below, we will show how Szemerédi's theorem can be quickly deduced from the Furstenberg recurrence theorem using simple ultraproduct arguments.

1.3. **Overview of the paper.** This paper contains the following:

(i) A development for the nonspecialist of the background required to understand Szegedy's results in higher order Fourier analysis.
(ii) A careful proof of the first major structure theorem for the Fourier σ-algebras, extending the context from finite groups to all measurable groups that are finite measure spaces, filling out Szegedy's arguments and in some cases correcting them.
(iii) A discussion and proof sketch of the second major structure theorem for Fourier σ-algebras.
(iv) A discussion of possible extensions of these results to the contexts of infinite measure spaces and finitely additive measure spaces.

No prior familiarity with either higher order Fourier analysis or ultraproduct analysis is presumed.

In the second section, we present a unified introduction to ultraproducts and ultraproduct analysis. Two model-theoretic results are relegated to an appendix. No new results are presented, although we include new proofs of known facts (for example, Proposition 2.13). As an illustration of the power of these techniques, we include a proof of Szemerédi's theorem from the Furstenberg recurrence theorem.

In the third section, we construct the main measure space $\mathcal{A}$ on the ultraproduct group $\mathbf{A}$ in a self-contained way, and then detour through a discussion of the integration theory on the ultraproduct with a special focus on under what conditions the integral and ultralimit may be interchanged. This theory has a resemblance to the elementary real analysis question of when a limit and integral commute. The language is rooted in ultraproduct theory rather than nonstandard analysis, in contrast to the context in which these results appear in the literature.

In the fourth and longest section, we introduce various useful measure spaces on the product of ultraproduct spaces and state and prove Keisler's Fubini-type theorem before detouring again through a general discussion of Gowers norms. Here, in recognition of the lack of any discussion in the literature of Gowers norms over infinite spaces, we broaden to this context. Following Szegedy, we, though it makes the language somewhat more difficult to understand, mix ergodic-theoretic language with structures first defined in the context of additive combinatorics. Many results from the setting of finite groups carry over and are presented in the generality of finite measure spaces. Proposition 4.10 is a new result, while some results (such as Lemma 4.1) have novel, simplified proofs. The discussion then shifts to a careful exposition of Szegedy's first structure theorem for higher order Fourier σ-algebras in the expanded context of ultraproducts of groups of finite measure, which states that a function is measurable in $\mathcal{F}_{k-1}$ if and only if it is orthogonal to every L^2 function with a vanishing Gowers norm. The often very sketchy proofs in [27] are fleshed out and in some cases corrected; for example, the original paper contained numerous errors centered around the distinction between the concepts of sets generating a σ-algebra $\mathcal{A}$ and sets generating a σ-algebra $\mathcal{B}$ which is dense in $\mathcal{A}$. Numerous smaller lemmas are added as necessary to ease the presentation.

In the fifth section, we sketch a proof of the second structure theorem in Szegedy's higher order Fourier analysis, which states that $L^2(\mathcal{F}_k)$ can be decomposed into pairwise-orthogonal modules over $L^\infty(\mathcal{F}_{k-1})$, each generated by a single function. This has the corollary that for every function $f \in L^2$ measurable in the ultraproduct and every integer $k > 1$, there is a unique decomposition

$$f = g + \sum_{j=1}^{\infty} f_i$$

in the L^2 sense, where $||g||_{U_k} = 0$ and the f_i are contained in different rank-one modules over $L^\infty(\mathcal{F}_{k-2})$. We then briefly discuss two possible extensions, to spaces of infinite measure and to finitely additive measures.

2. Preliminaries and a motivating example

2.1. **Ultraproducts.** Most of this material can be found in [34]. We start with a definition. An *ultrafilter* on the positive natural numbers $\mathbb{N}$ is a set ω consisting of subsets of $\mathbb{N}$ satisfying the following conditions:

(i) The empty set does not lie in ω.
(ii) If $A \subset \mathbb{N}$ lies in ω, then any subset of $\mathbb{N}$ containing A lies in ω.
(iii) If A and B lie in ω, then the intersection $A \cap B$ lies in ω.
(iv) If $A \subset \mathbb{N}$, then exactly one of A and $\mathbb{N} \setminus A$ lies in ω.

If, furthermore, no finite set lies in ω, we say that ω is a *nonprincipal ultrafilter*.

We would like to show that there exists a nonprincipal ultrafilter. Define a *proper filter* to be a set consisting of subsets of $\mathbb{N}$ that satisfies the first three conditions above. We say that a collection γ of subsets of $\mathbb{N}$ has the *finite intersection property* if for any finite collection $A_1, \ldots, A_n$ of elements of γ, $A_1 \cap \ldots \cap A_n$ is nonempty. Certainly any proper filter has the finite intersection property.

Lemma 2.1. *If a collection γ has the finite intersection property, there is a proper filter ϕ containing γ.*

Proof. Take ϕ to be the set of subsets B of $\mathbb{N}$ such that there exist $A_1, \ldots, A_n$ in γ whose intersection $A_1 \cap \ldots \cap A_n$ is a subset of B. Conditions (i) and (ii) are immediate. If B_1 and B_2 lie in ϕ, then there exist $A_1, \ldots, A_m$ and $A'_1, \ldots, A'_n$ lying in γ such that $A_1 \cap \ldots \cap A_m \subset B_1$ and $A'_1 \cap \ldots \cap A'_n \subset B_2$. Therefore

$$A_1 \cap \ldots \cap A_m \cap A'_1 \cap \ldots \cap A'_n \subset B_1 \cap B_2,$$

so $B_1 \cap B_2$ lies in ϕ and condition (iii) is verified. ☐

Proposition 2.2. *There exists a nonprincipal ultrafilter.*

Proof. The proof requires the use of the axiom of choice in the form of Zorn's lemma. Let γ be the set of all cofinite subsets of $\mathbb{N}$; it is clear that γ is a proper filter. Given any chain of proper filters, ordered by inclusion, take the union of all of them; it is immediate that this union is also a proper filter. Therefore we can apply Zorn's lemma to find a maximal proper filter ω.

Proceed by contradiction and assume that ω does not satisfy condition (iv) in the definition of an ultrafilter, so there is some $A \subset \mathbb{N}$ such that neither A nor $\mathbb{N} \setminus A$ lie in ω. Let

$$\gamma = \{A \cap B : B \in \omega\}.$$

Then ω' has the finite intersection property: if $B_1, \ldots, B_n$ belongs to ω, then $B_1 \cap \ldots \cap B_n$ does as well (and is nonempty). If $A \cap B_1 \cap \ldots \cap B_n$ were empty, then we would have $B_1 \cap \ldots \cap B_n \subset \mathbb{N} \setminus A$, so $\mathbb{N} \setminus A \in \omega$, contrary to hypothesis.

Therefore by the lemma we can extend γ to a filter ω', which must contain A as well as every element of ω. Thus ω was not a maximal proper filter, which is a contradiction. We conclude that ω satisfies condition (iv) and is therefore an ultrafilter. Furthermore, since ω contains γ, by condition (iv) it cannot contain any finite set, so it is a nonprincipal ultrafilter. ☐

Lemma 2.3. *If $A \in \omega$ for some ultrafilter ω, and A can be written as the disjoint union of finitely many sets A_i, then there is exactly one i such that $A_i \in \omega$.*

Proof. We use condition (iv) repeatedly. First, either A_1 or $\mathbb{N} \setminus A_1$ is in ω; if the former, we're done, while if the latter,

$$A \cap (\mathbb{N} \setminus A_1) = \bigsqcup_{i=2}^{n} A_i$$

is in ω. By induction, after a finite number of such steps we can find an i such that $A_i \in \omega$. Only one such i can exist, for otherwise the intersection of two disjoint sets would be in the ultrafilter, contradicting condition (i). $\square$

An *ultraproduct* $\mathbf{X}$ of a sequence of sets $\{X_i\}_{i=1}^{\infty}$ with respect to an ultrafilter ω is defined as follows. First construct the Cartesian product $\prod_{i\in\mathbb{N}} X_i$. Define an equivalence relation $x \sim y$, where $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$, by

$$x \sim y \iff \{i \in \mathbb{N} : x_i = y_i\} \in \omega.$$

Then let $\mathbf{X} = \prod_{i\in\mathbb{N}} X_i/\sim$. One can think of $\mathbf{X}$ as a sort of completion where one can take the limit of arbitrary sequences, rather than just Cauchy sequences: given a sequence $\{x_i\}_{i=1}^{\infty}$, the equivalence class in $\mathbf{X}$ of this sequence will be denoted

$$x = \lim_{i\to\omega} x_i.$$

Thus, in this terminology, we have

$$\lim_{i\to\omega} x_i = \lim_{i\to\omega} y_i$$

if and only if the set of $i \in \mathbb{N}$ such that $x_i = y_i$ is a member of ω. Similarly, for subsets $H_i \subset X_i$ we denote by $\mathbf{H}$ or $\lim_{i\to\omega} H_i$ the set of all elements of the ultraproduct arising from limits of points in the given subsets:

$$\lim_{i\to\omega} H_i = \left\{\lim_{i\to\omega} x_i : x_i \in H_i, i \in \mathbb{N}\right\}$$

If all of the X_i are the same space, the ultraproduct is also called an *ultrapower.*

Ultrapowers with respect to a nonprincipal ultrafilter will be denoted with a prior asterisk; e.g., the ultrapowers of $\mathbb{N}$ and $\mathbb{R}$ are written $^*\mathbb{N}$ and $^*\mathbb{R}$, respectively. The latter object is called the set of *hyperreal numbers*. The order structure carries over into the hyperreals: for real sequences a_i and b_i whose ultralimits are a and b, respectively, by Lemma 2.3 exactly one of the sets $\{i : a_i < b_i\}$, $\{i : a_i = b_i\}$, or $\{i : a_i > b_i\}$ is in ω. In the first case we say $a < b$, in the second $a = b$, and in the third $a > b$.

We will assume basic facts about $^*\mathbb{N}$ and the hyperreals, which can be found in [34] or more completely in [24]: call a hyperreal *standard* if it can be written as $\lim_{i\to\omega} r$ for some constant $r \in \mathbb{R}$; thus the reals can be considered a subset of the hyperreals (and likewise for $^*\mathbb{N}$). The hyperreals are an ordered field with an ordering extending that of the reals. Define an absolute value in the obvious way, by setting

$$\left|\lim_{i\to\omega} r_i\right| = \lim_{i\to\omega} |r_i|,$$

which will be a nonnegative hyperreal. Given $x \in {}^*\mathbb{R}$, we call x *bounded* if $|x| < C$ for some standard C, and we call x *infinitesimal* if $|x| < r$ for all standard r. Hyperreals that are not bounded are called *infinite*. Every bounded x has a unique decomposition

$$x = \mathrm{st}(x) + (x - \mathrm{st}(x))$$

into a standard part $\mathrm{st}(x)$ and an infinitesimal part $x-\mathrm{st}(x)$, where the mapping $x \mapsto \mathrm{st}(x)$ is a homomorphism from the ring of bounded hyperreals to the reals.

Given functions $f_i : X_i \to \mathbb{R}$, we can form an ultralimit $f = \lim_{i\to\omega} f_i : \mathbf{X} \to {}^*\mathbb{R}$ by defining

$$f\left(\lim_{i\to\omega} x_i\right) = \lim_{i\to\omega} f_i(x_i).$$

It is an easy exercise to show that finite Boolean operations are preserved when taking ultraproducts; that is, if $\{G_i\}_{i=1}^{\infty}$ and $\{H_i\}_{i=1}^{\infty}$ are sequences of subsets of the X_i, then we have

$$\lim_{i\to\omega}(G_i \cup H_i) = \left(\lim_{i\to\omega} G_i\right) \cup \left(\lim_{i\to\omega} H_i\right),$$

and similarly for intersections and complements.

The ultraproduct construction is properly a model-theoretic one, and can be defined for arbitrary structures. The investment in added generality is rewarded with two fundamental theorems, which together comprise the core of the utility of ultraproducts in any context: Łoś' theorem (sometimes called the *fundamental theorem of ultraproducts*) and the saturation theorem. To keep the model theory to a minimum, we will briefly define the model-theoretic ultraproduct here and state the two important theorems, but relegate their proofs to the appendix. For additional background, see [2] and [22].

To define the general ultraproduct, let $\mathcal{M}_i$ be a structure with universe M_i in a fixed language $\mathcal{L}$ for each $i \in I$, where I is some index set (for our purposes, I will always be the natural numbers $\mathbb{N}$). Pick an ultrafilter ω on I. Define the ultraproduct $\mathcal{M} = \lim_{i\to\omega} \mathcal{M}_i$ as follows: let the universe of $\mathcal{M}$ be

$$M = \prod_{i\in\mathbb{N}} M_i/\sim,$$

where $\sim$ is the equivalence relation given by the ultrafilter as previously described; we again denote the image in M of a sequence $a_i \in M_i$ by $\lim_{i\to\omega} a_i$. For each constant symbol c of the language, let the interpretation $c^{\mathcal{M}}$ of c in $\mathcal{M}$ be the equivalence class of $(c^{\mathcal{M}_1}, c^{\mathcal{M}_2}, \ldots)$ in $\prod M_i$. For an 1-ary function symbol in the language, if $a = \lim_{i\to\omega} a_i$, define

$$f^{\mathcal{M}}(a) = \lim_{i\to\omega} f^{\mathcal{M}_i}(a_i).$$

It is easy to check that this is well-defined in the sense that it is independent of a particular choice of a_i. A precisely analogous definition is made for n-ary function symbols. Similarly, if R is a one-place relation symbol and $a = \lim_{i\to\omega} a_i$, we interpret

$$R^{\mathcal{M}} = \{a : \{i \in \mathbb{N} : a_i \in R^{\mathcal{M}_i}\} \in \omega\},$$

which is similarly well-defined, and likewise for n-place relation symbols. It will be noted that many of the previous definitions in this section are particular instances of this general construction.

With this out of the way, we can state Łoś' theorem, which states that a first order sentence is true in the ultraproduct if and only if it is true for ω-many of the original structures:

Theorem 2.4 (Łoś' theorem)**.** *Let ω be an ultrafilter on I, let $\mathcal{M}_i$, $i \in I$, be a sequence of structures over the same language, and let $\mathcal{M} = \lim_{i\to\omega} \mathcal{M}_i$. Let ϕ be a first-order formula over σ with n arguments, and let $a^1, \ldots, a^n \in \mathcal{M}$, where $a^k = \lim_{i\to\omega} a_i^k$. Then $\phi(a^1, \ldots, a^n)$*

is true in $\mathcal{M}$ if and only if the set of i for which $\phi(a_i^1, \ldots, a_i^n)$ is true in $\mathcal{M}_i$ is an element of ω.

The following corollary will be useful. By an *internal set* we mean a set, viewed as a subset of $\mathbf{X}$, which is equal to an ultraproduct of subsets of X_i.

Corollary 2.5 (Overspill)**.** *Let $\phi(n)$ be a formula such that $\phi(n)$ is true for all finite $n \in {}^*\mathbb{N}$. Then there exists an infinite m such that $\phi(n)$ is true for all $n \leq m$.*

Proof. First, we show that $\mathbb{N} \subset {}^*\mathbb{N}$ is not an internal set. Note that the complement of an internal set must be internal, so it suffices to show that the complement of $\mathbb{N}$ is not internal. Consider the statement in the language of the natural numbers that there exists a smallest element, which is certainly expressible in first-order logic. This statement is true for every subset of the natural numbers, and therefore by Łoś' theorem, it is true for every internal subset of $\mathbb{N}$. However, the complement of $\mathbb{N}$ has no smallest element: given any element of the complement $d = \lim_{i\to\omega} d_i$, we can find a strictly smaller one given simply $d' = \lim_{i\to\omega} d'_i$, where $d'_i = d_i - 1$ if $d_i > 0$ and $d'_i = 0$ otherwise. Therefore $\mathbb{N}$ is not an internal space.

The set

$$\{m \in {}^*\mathbb{N} \text{ such that } \phi(n) \text{ for all } n < m\}$$

is clearly an internal set, and by hypothesis it contains $\mathbb{N}$. However, $\mathbb{N}$ is not an internal set, so it is a strict inclusion, so there is some infinite m such that $\phi(n)$ is true for all $n \leq m$. $\square$

To state the saturation theorem, we need several definitions from model theory. Let A be a subset of M, the universe of some structure $\mathcal{M}$ in a language $\mathcal{L}$. Let $\mathrm{Th}_A(\mathcal{M})$ be the set of sentences true in $\mathcal{M}$ in the language $\mathcal{L}_A$ obtained by augmenting $\mathcal{L}$ with one constant symbol for each $a \in A$.

An *n-type* p is then a set of $\mathcal{L}_A$-formulas with n free variables such that $p \cup \mathrm{Th}_A(\mathcal{M})$ is satisfiable. A *complete n-type* is an n-type such that if ϕ is a $\mathcal{L}_A$ formula with n free variables, then either $\phi \in p$ or $\neg\phi \in p$. The set of all complete n types in $\mathcal{M}$ over A is denoted $S_n^{\mathcal{M}}(A)$.

If $p \in S_n^{\mathcal{M}}(A)$, we say that $(a_1, \ldots, a_n) \in M^n$ *realizes* p if $\mathcal{M} \models \phi(a_1, \ldots, a_n)$ for all $\phi \in p$. That is, a type p is realized if there is some n-tuple of elements in the model making every formula in p true simultaneously.

Let κ be an infinite cardinal. Say that a structure $\mathcal{M}$ is *κ-saturated* if, for all $A \subset M$ such that $|A| < \kappa$ and all $p \in S_n^{\mathcal{M}}(A)$, p is realized in $\mathcal{M}$. A saturated model is in some sense a particularly large model, containing many realized n-types, relative of course to the given κ.

It is now possible to state the saturation theorem, which requires that the ultrafilter ω be nonprincipal.

Theorem 2.6 (Saturation theorem)**.** *Let ω be a nonprincipal ultrafilter, $\mathcal{M}_i$ a sequence of structures over the same countable language, and $\mathcal{M} = \lim_{i\to\omega} \mathcal{M}_i$. Then $\mathcal{M}$ is $\aleph_1$-saturated.*

Every language used in this paper is countable, so that restriction is not a serious one.

The following is a basic logical fact that will be needed below. We include a proof merely because there exists one that crucially employs the ultraproduct construction:

Theorem 2.7 (Compactness)**.** *A set of sentences $\Gamma = \{\phi_i\}_{i\in I}$ has a model if and only if every finite subset of Γ has a model.*

Proof (sketch). Let J be the set of finite subsets of I. For each $j \in J$, by assumption there is a model $\mathcal{M}_j$ making true $\{\phi_i\}_{i \in j}$. Let $A_j = \{k : j \subseteq k\}$ for each $j \in J$. The collection of all such A_j generates a filter, so there is an ultrafilter ω on J containing A_j. For any formula ϕ, certainly $A_{\{\phi\}} \in \omega$, so the set of all j such that ϕ holds in $\mathcal{M}_j$ is a superset of $A_{\{\phi\}}$ and hence in ω as well. Then by Łoś' theorem ϕ holds in the ultraproduct $\lim_{j \to \omega} \mathcal{M}_j$, so this ultraproduct is a model of Γ, as desired. $\square$

The saturation theorem has several down-to-earth consequences that will be very important. First, it yields a strong "countable compactness" result for internal sets.

Corollary 2.8 (Countable compactness)**.** *Let ω be a nonprincipal ultrafilter. Then any cover of an internal set by countably many internal sets has a finite subcover.*

Proof. We first prove an intermediate result: let $\mathbf{C}_1, \mathbf{C}_2, \ldots$ be a sequence of ultrapdroduct spaces in $\mathbf{X}$ such that any finite collection of these sets has nonempty intersection; that is,

$$\bigcap_{n=1}^{N} \mathbf{C}_n \neq \emptyset, \quad \forall n \in \mathbb{N}.$$

Then we claim that the entire sequence has nonempty intersection; that is,

$$\bigcap_{n=1}^{\infty} \mathbf{C}_n \neq \emptyset.$$

Write

$$\mathbf{C}^n = \lim_{i \to \omega} C_i^n$$

for subsets $C_i^n \subseteq X_i$. Augment the language with countably many one-place relation symbols C^n, which will be interpreted in each of the X_i as the subset C_i^n. By the definition of the ultraproduct, C^n will be interpreted as $\mathbf{C}^n$ in the ultraproduct $\mathbf{X}$. Let p be the set consisting of the following collection of sentences:

$$p(v) = \{C^n(v)\},$$

where the notation is interpreted by letting $C^n(v)$ be true if and only if $v \in C^n$.

We claim that p is an (incomplete) 1-type. By compactness, p is satisfiable if it is finitely satisfiable, which is true because every finite collection has nonempty intersection. Our language is still countable, so we can apply the saturation theorem (combined with Proposition A.1 in the appendix, as p is incomplete) to conclude that p is satisfied in $\mathbf{X}$. Therefore the intermediate result is proven.

It remains to show that this implies countable compactness. Given a countable cover of $\mathbf{X}$ by internal spaces $\{\mathbf{D}_i\}_{i=1}^{\infty}$, let $\mathbf{C}_i$ be the complement of $\mathbf{D}_i$ in $\mathbf{X}$ for each i. We then have, for any set of indices $\{i_1, i_2, \ldots, i_r\}$,

$$\bigcap_{j=1}^{r} \mathbf{C}_{i_j} = \mathbf{X} - \bigcup_{j=1}^{r} \mathbf{D}_{i_j}.$$

Assume there is no finite subcover of the $\mathbf{D}_i$. There is no set of indices such that the right hand side of the above equality is nonempty, so no finite intersection of the $\mathbf{C}_i$ is nonempty. But because the $\mathbf{D}_i$ form a cover, the infinite intersection of all the $\mathbf{C}_i$ is empty. This violates the countable saturation result above. $\square$

It is possible to prove countable compactness "by hand," as follows, although the more general model-theoretic viewpoint will pay off later:

Alternate proof of countable compactness. We follow [34]. It suffices to prove the intermediate result from above, so let notation be as before. Because finite intersections are preserved under ultraproducts,

$$\lim_{i\to\omega}\left(\bigcap_{n=1}^{N} C_n^{(i)}\right) = \bigcap_{n=1}^{N} \mathbf{C}_n.$$

Therefore by assumption the set of i such that $\cap_{n=1}^{N} C_n^{(i)}$ is nonempty, which we denote E_N, is an element of ω. Without loss of generality we can arrange that $E_1 \supset E_2 \supset \dots$ by replacing E_n by the intersection $E_1 \cap \dots \cap E_n$, which must still be a member of the ultrafilter. Additionally, we want to remove any natural number less than N from each E_N, so that $\cap_{N=1}^{\infty} E_N = \emptyset$. That elements of an ultrafilter are still in the ultrafilter after removing finitely many elements requires that the ultrafilter is nonprincipal (this is the only place in the proof where this condition is used); otherwise, if such a set were not in the ultrafilter, the intersection of its complement with the original set would be both finite and a member of the ultrafilter.

For each $i \in E_1$, let N_i be the largest number such that $i \in E_N$ and let x_i be an (arbitrary) element of the set $\cap_{n=1}^{N_i} C_n^{(i)}$, which is nonempty by assumption. By construction, $x_i \in C_n^{(i)}$ for all $i \in E_n$, which is an element of ω. Therefore the object

$$x = \lim_{i\to\omega} x_n$$

lies in $\mathbf{C}_n$ for all n, so the infinite intersection of the $\mathbf{C}_n$ is nonempty. This proves the claim of countable saturation. □

Another form of countable compactness which is often useful is the following:

Corollary 2.9 (Countable comprehension)**.** *Given any sequence $(\boldsymbol{E}_n)_{n\in\mathbb{N}}$ of internal subsets of $\boldsymbol{X}$, there is an internal function with domain ${}^*\mathbb{N}$ (that is, an internal sequence) $(\boldsymbol{E}_n)_{n^*\in\mathbb{N}}$ of subsets of $\boldsymbol{X}$.*

Proof. Let B_m be the set of all sequences $(\mathbf{F}_n)_{n\in\mathbb{N}}$ such that $\mathbf{F}_n = \mathbf{E}_n$ for all $n \leq m$ and $\mathbf{F}_n \subset X$. It is easy to see that any finite collection of the B_m has nonempty intersection. Therefore by the intermediate result of the proof of countable compactness, the countable intersection of all the B_m for finite m is nonempty, so there is some extension of the $\mathbf{E}_n$. □

In this paper, we will need to take ultraproducts of sequences of abelian groups $\{A_i\}_{i=1}^{\infty}$. The group structure on the ultraproduct $\mathbf{A}$ is defined in the obvious way, per the above general definition: we have

$$\lim_{i\to\omega}(a_i + b_i) = \lim_{i\to\omega} a_i + \lim_{i\to\omega} b_i$$

and

$$\left(\lim_{i\to\omega} a_i\right)^{-1} = \lim_{i\to\omega}(a_i^{-1});$$

it is clear both that these operations are well-defined and that they define a group structure on $\mathbf{A}$.

Finite products do not technically commute with ultraproducts: that is, if we have sequences of groups (or more general structures) $\{A_i\}$ and $\{B_i\}$, then the space

$$\mathbf{C} = \lim_{i\to\omega}(A_i \times B_i)$$

is not the same space as

$$\mathbf{A} \times \mathbf{B} = \left(\lim_{i\to\omega} A_i\right) \times \left(\lim_{i\to\omega} B_i\right).$$

However, in light of the following lemma, we will identify them without further concern:

Lemma 2.10. *With notation as above, there is a natural isomorphism*

$$\boldsymbol{C} \simeq \boldsymbol{A} \times \boldsymbol{B}.$$

Proof. An element of $\mathbf{C}$ is an ultralimit of pairs (a_i, b_i), which we can use to define an element

$$\left(\lim_{i\to\omega} a_i\right) \times \left(\lim_{i\to\omega} b_i\right) \in \mathbf{A} \times \mathbf{B};$$

it is clear that this map is well-defined in that it is insensitive to the particular a_i, b_i chosen outside of an element of the ultrafilter. This map has an obvious inverse, and it is a trivial matter to check that it preserves group structure. □

The next lemma and propositions are not strictly necessary for the results that follow, but do they motivate two things. First, Proposition 2.12 explains why our basic construction will require taking a nonprincipal ultrafilter: otherwise, we get no extra structure — and in fact lose almost all information — upon taking the ultraproduct. Second, Proposition 2.13 elucidates why it will be uninteresting to take the ultraproduct of finite groups that are bounded in order.

An ultrafilter which is not nonprincipal is (unsurprisingly) called *principal*.

Lemma 2.11. *If an ultrafilter ω is principal, then it is of the form*

$$\omega_n = \{A \subset \mathbb{N} : n \in A\}$$

for some $n \in \mathbb{N}$.

Proof. It is immediate to verify that ω_n as defined above is an ultrafilter. By definition, ω must contain some finite set $B = \{n_1, \ldots, n_k\}$. By Lemma 2.3, writing B as the union of singleton sets, there is some i such that $\{n_i\} \in \omega$. Because ω contains $\{n\}$, it must contain all subsets of $\mathbb{N}$ containing n; that is, it must contain ω_n. If $A \notin \omega_n$, then $\mathbb{N} \setminus A \in \omega_n$, so $\mathbb{N} \setminus A \in \omega$, so $A \notin \omega$. Therefore $\omega = \omega_n$. □

Proposition 2.12. *The ultraproduct $\boldsymbol{A}$ of groups A_i with respect to the ultrafilter ω_n is isomorphic to A_n.*

Proof. Two elements of the ultraproduct, considered as sequences $a = (a_1, a_2, \ldots)$ and $b = (b_1, b_2, \ldots)$ with $a_i, b_i \in A_i$, are equal if and only if the set

$$\{i \in \mathbb{N} : a_i = b_i\}$$

lies in ω_n, which is true if and only if $a_n = b_n$. Therefore the map $\Phi : \mathbf{A} \to A_n$ given by $\Phi(a) = a_n$ is a bijection. It obviously a homomorphism by the definition of the group structure on the ultraproduct. □

Proposition 2.13. *If A_i is a sequence of groups such that $|A_i|$ is bounded, then there is an i such that the ultraproduct $\boldsymbol{A}$ is isomorphic to A_i. Conversely, if $|A_i|$ is unbounded and ω is nonprincipal, then $\boldsymbol{A}$ is infinite.*

Proof. There are only finitely many isomorphism classes of groups of order less than a given bound. Let the indices i corresponding to a given isomorphism class be grouped into sets. Noting that their union $\mathbb{N}$ is certainly in ω, by Lemma 2.3 one such set I is in the ultrafilter ω; that is, there is an $I \in \omega$ such that $A_i \simeq A_j$ for $i, j \in I$. Let N be the order of the groups in this isomorphism class.

Without loss of generality, we can identify the groups in I via the isomorphisms and label them all as A. Consider the N elements of $\mathbf{A}$ given by the ultralimits of corresponding elements $a \in A$, which by abuse of notation we will denote by $a' = \lim_{i\to\omega} a$, even though the right hand side is defined only on an element of the ultrafilter. The element in the ultralimit is nonetheless well-defined, because any choice for the elements of A_i, $i \notin I$ will lead to an equivalent element in the ultralimit. We can think of this limit as a map $A \to \mathbf{A}$ given by $a \mapsto a'$.

First, we show that this map is injective. If $a \neq b$ in A, then $a' \neq b'$ in $\mathbf{A}$: for otherwise, the set J of indices for which the sequences agree would lie in ω, but I and J would then be disjoint, so their intersection $\emptyset$ would also lie in ω, contrary to definition.

Second, we show that this map is surjective. Consider an arbitrary element $b = \lim_{i\to\omega} b_i$ of $\mathbf{A}$; we will show that it must be equal to one of the N elements a' by another application of Lemma 2.3. For each $a \in A$, consider the set B_a of indices i such that $b_i = a$. We can certainly write I as the finite disjoint union of the B_a, so by the lemma one of the B_a is in ω. Then $\{i : b_i = a\} \in \omega$, so $b = a'$. Therefore the mapping $A \to \mathbf{A}$ described above is bijective, and by the definition of the group structure on the ultraproduct it is an isomorphism.

The converse is straightforward: for any N we can find an index i_N such that $|A_j| > N$ for $j > i_N$. Then it is easy to construct N sequences a_i^k, $1 \leq k \leq N$, that differ on every index greater than j, which is a cofinite set and in the nonprincipal ultrafilter ω. Therefore the elements of the ultraproduct $a^k = \lim_{i\to\omega} a_i^k$ are all distinct. Because we can do this for every N, $\mathbf{A}$ is infinite. □

It should be noted that Proposition 2.13 has an alternate proof as an easy corollary of Łoś' theorem, together with the basic model-theoretic fact that for any finite structure in a finite language, there is a sentence ϕ precisely specifying the structure up to isomorphism.

2.2. **A motivating example: Szemerédi's theorem.** As an example of the utility of ultralimit analysis in additive number theory, we will follow [33] in showing how Szemerédi's theorem can be quickly deduced from the Furstenberg recurrence theorem, which we will assume:

Theorem 2.14 (Furstenberg recurrence theorem)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system, let $A \subset X$ have positive measure, and let k be a positive integer. Then there exists a positive integer r such that*

$$A \cap T^r A \cap \dots T^{(k-1)r} A$$

is nonempty. □

By a measure-preserving system, it is meant that $(X, \mathcal{B}, \mu)$ is a probability space and $T : X \to X$ is a measure-preserving bijection such that both T and its inverse are measurable.

Using ultralimits, we can use the Furstenberg recurrence theorem to prove Szemerédi's thereorem. First, a definition: let A be a set of integers. If

$$\limsup_{n\to\infty} \frac{|A \cap [-n,-n]|}{|[-n,n]|} \geq \delta$$

for some $\delta > 0$, then we say that A has positive upper density.

Theorem 2.15 (Szemérédi's theorem)**.** *Every set of integers of positive upper density contains arbitrarily long arithmetic progressions.*

Proof. Proceed by contradiction, assuming the existence of a set A of positive upper density and no progressions of length k for some $k \geq 1$. Having positive upper density means precisely that there is a sequence of integers of n_i tending to infinity and a $\delta > 0$ such that

$$|A \cap [-n_i, n_i]| \geq \delta |[-n_i, n_i]|$$

for each i. Now consider the ultrapower ${}^*A$ with respect to some nonprincipal ultrafilter ω, and the ultralimit $n = \lim_{i\to\omega} n_i$. The former is a subset of ${}^*\mathbb{N}$, while the latter is in ${}^*\mathbb{N}$ (which can be considered a subset of the hyperreals). We claim that

$$|{}^*A \cap [-n, n]| \geq \delta |[-n, n]|,$$

where the absolute value signs are as defined previously on the hyperreals (and $\delta = \lim_{i\to\omega} \delta$ is considered as a hyperreal as well). This is a simple application of Łoś' theorem; we will include here some of the details to get a sense of how the model-theoretic result is used in practice. Take the signature consisting of the binary operations $+$ and $\cdot$, the binary relation $\geq$, the unary relation (or set) A, and the constant δ. By Łoś' theorem, since the above inequality can be written as a first-order statement and is true for each i, and $\mathbb{N}$ is certainly an element of ω, by taking ultralimits of everything we get a true statement, which is precisely the claim. Additionally, it is true that ${}^*A$ has no progressions of length k, since that property can be expressed in a first-order sentence and is true for A.

Call the space of all finite Boolean combinations of shifts of ${}^*A$ by elements of $\mathbb{Z}$, which includes the empty set and all of ${}^*\mathbb{Z}$, the *definable sets* $\mathcal{D}$. The elements of $\mathcal{D}$ form a field of sets. Define $\mu' : \mathcal{D} \to {}^*\mathbb{R}$ by

$$\mu'(B) = \frac{|B \cap [-n, n]|}{|[-n, n]|},$$

which is a function from definable sets to the set ${}^*[0, 1]$, the ultrapower of the unit interval, such that $\mu'(\emptyset) = 0$ and $\mu'({}^*\mathbb{Z}) = 1$. Additionally, it is finitely additive: given two disjoint definable sets B_1 and B_2,

$$\begin{aligned}
\mu'(B_1 \cup B_2) &= \frac{|(B_1 \cup B_2) \cap [-n, n]|}{|[-n, n]|} \\
&= \frac{|B_1 \cap [-n, n]| + |B_2 \cap [-n, n]|}{|[-n, n]|} \\
&= \mu'(B_1) + \mu'(B_2).
\end{aligned}$$

Finally, it is nearly translation invariant under translation by any standard integer j: we have

$$\begin{aligned}\mu'(B+j) &= \frac{|B\cap[-n-j,n-j]|}{|[-n,n]|}\\ &= \frac{|B\cap[-n,n]|}{|[-n,n]|}+\frac{|B\cap[-n-j,-n-1]|}{|[-n,n]|}-\frac{|B\cap[n-j+1,n]|}{|[-n,n]|}\\ &= \mu'(B)+\epsilon_1+\epsilon_2,\end{aligned}$$

where ϵ_1 and ϵ_2 are both infinitesimal, since both numerators are finite but both denominators are infinite. Define $\mu : \mathcal{D} \to \mathbb{R}$ by

$$\mu(B) = \mathrm{st}(\mu'(B))$$

and recall that st is a homomorphism. It is clear that μ is nonnegative, $\mu(\emptyset) = 1$, $\mu({}^*\mathbb{Z}) = 1$, and μ is finitely additive. Taking standard parts kills infinitesimal quantities, so μ is a finitely additive pre-measure on the field of sets $\mathcal{D}$ that is invariant under shifts by finite integers. By construction, $\mu(A) \geq \delta$.

We will use μ to construct a measure-preserving system on which to apply the Furstenberg recurrence theorem. Let $X = 2^{\mathbb{Z}}$, the set of all subsets of the integers. We can view this space as the set of all doubly-infinite sequences of ones and zeroes. Define a topology on X as follows: first define the *cylinder sets* E_m by

$$E_m = \{B \in 2^{\mathbb{Z}} : m \in B\}.$$

From the viewpoint of the doubly-infinite sequences, E_m consists of all sequences with 1 in the mth position. Let the set of all finite Boolean combinations of the E_m, which is clearly a field of sets, be denoted $\mathcal{E}$. The topology is defined by letting $\mathcal{E}$ be a basis. It then follows that $\mathcal{E}$ consists precisely of the clopen sets of X. Let $T : X \to X$ be a shift-by-one map, shifting each entry of a doubly-infinite sequence to the left. To define a measure ν on this space, we will need to somehow reference our ultraproduct construction.

This reference is provided by a map $\Phi : \mathcal{E} \to \mathcal{D}$, which is defined on the E_m by

$$\Phi(E_m) = {}^*A + m$$

and then extended to all of $\mathcal{E}$ by Boolean combinations. Let $\nu : \mathcal{E} \to \mathbb{R}$ be defined for $E \in \mathcal{E}$ by

$$\nu(E) = \mu(\Phi(E)),$$

the pullback of μ by Φ. It is simple to verify for ν the properties we want: it is nonnegative, we have $\nu(\emptyset) = \mu(\emptyset) = 0$, $\nu(X) = \mu({}^*\mathbb{Z}) = 1$, and it is clearly invariant under the shift T because μ is invariant under shifts by integers. Finite additivity follows because Φ preserves disjointness of sets, which follows because we have for $E_1, E_2 \in \mathcal{E}$ that $E_1 \cap E_2 = \emptyset \implies \Phi(E_1) \cap \Phi(E_2) = \Phi(E_1 \cap E_2) = \emptyset$. Finally, we of course have $\nu(E) \geq \delta$.

By the Carathéodory extension theorem, ν extends to a measure on the σ-algebra $\mathcal{B}$ generated by $\mathcal{E}$ on X. We may now finally apply the Furstenberg recurrence theorem on the measure-preserving system $(X, \mathcal{B}, \nu, T)$ and the set E_0 of positive measure to deduce that there is a positive integer r such that

$$E_0 \cap T^r E_0 \cap \ldots \cap T^{(k-1)r} E_0 \neq \emptyset.$$

By applying Φ, it is immediate that

$${}^*A \cap ({}^*A + r) \cap \ldots \cap ({}^*A + (k-1)r) \neq \emptyset.$$

Therefore ${}^*A$ contains an arithmetic progression of length k, which is a contradiction. $\square$

3. Analysis on ultraproducts

3.1. **The ultraproduct construction.** Let $\{A_i\}_{i=1}^{\infty}$ be a sequence of abelian groups that are also probability spaces, equipped with normalized measures μ_i, such that the measure spaces thus formed are compatible with the group structure in the sense that the action of the group on any measurable set is again measurable. In this situation we say that the σ-algebra in question is invariant with respect to the action of A_i on itself, or simply that the σ-algebra is invariant when there is no danger of confusion. To avoid the trivial case described by Lemma 2.13, we assume that $\sup_i |A_i| = \infty$. Although Szegedy's sequence of papers ([27], [28], [29]) explicitly discusses only the case where the A_i are finite groups (with μ_i therefore the normalized counting measure), his arguments translate into the compact case with little alteration; we will flesh out these arguments in subsequent sections. Later, in the fifth section of this paper, we will discuss extensions of these results to other contexts.

Let $\mathbf{A}$ be the ultraproduct of the A_i with respect to a fixed nonpricipal ultrafilter ω. We will construct a σ-algebra on $\mathbf{A}$ as follows. Say that a subset $\mathbf{H} \subseteq \mathbf{A}$ is an element of $\mathcal{S}(\mathbf{A})$ if we can write it as

$$\mathbf{H} = \lim_{i \to \omega} H_i$$

for some measurable sets $H_i \subseteq A_i$. When it is clear from context, we will often omit writing the name of the ultraproduct group in labeling the σ-algebras on it. Note that $\mathcal{S}(\mathbf{A})$ forms a field of sets because the ultralimit construction respects finite Boolean operations.

Recalling the countable compactness result of Corollary 2.8, one might think at this point that $\mathcal{S}(\mathbf{A})$ is already a σ-algebra (it is certainly closed under complements). However, this fails because not all countable unions of elements of $\mathcal{S}(\mathbf{A})$ are internal sets, so the lemma does not apply.

Remark To clarify this statement, it may be helpful to develop a specific counterexample. Let the A_i be the groups $\mathbb{Z}/i\mathbb{Z}$ equipped with normalized counting measure, and as usual let ω be a nonprincipal ultrafilter. Each element of the resulting ultraproduct $\mathbf{A}$, considered as a singleton set, is trivially an internal space. In $\mathbf{A}$ there is an algebraic copy $\mathbf{N}$ of the natural numbers constructed as follows: for each natural number n, let $n_i = n \in \mathbb{Z}/i\mathbb{Z}$ when $i > n$ and let n_i be the zero element otherwise. The set $\mathbf{N} \subset \mathbf{A}$ is then defined to consist of the ultralimits $\mathbf{n} = \lim_{i \to \omega} n_i$. It is a countable union of internal sets (the singletons $\{\mathbf{n}\}$), but we have already shown in the proof of the overspill principle that it is not an internal set, and therefore not possibly in $\mathcal{S}(\mathbf{A})$. △

We define our main object of study $\mathcal{A}(\mathbf{A})$ to be the completion of the σ-algebra generated by $\mathcal{S}(\mathbf{A})$. The measure constructed here on $\mathcal{A}$ is sometimes called a Loeb measure (for example, by [4], [5], and [6]). The construction can be carried out in two equivalent ways: first, a general construction based on the Carathéodory extension theorem, and second, a hands-on method which gives us a helpful approximation result. By the uniqueness result in the Carathéodory construction, the two extensions are equivalent.

To put a measure μ on $\mathcal{A}$, we first define $\mu' : \mathcal{S} \to {}^*\mathbb{R}$ by

$$\mu'(\mathbf{H}) = \lim_{i \to \omega} \mu_i(H_i)$$

for sets $\mathbf{H} = \lim_{i \to \omega} H_i$. This is well-defined, for if $\mathbf{H}$ is the ultralimit of two sequences of subsets $H_i^{(1)}$ and $H_i^{(2)}$, then the sequences must agree on an element of the ultrafilter, so their measures will agree on an element of the ultrafilter, which uniquely fixes the ultralimit.

As the ultraproduct respects finite Boolean operations, μ' obeys the following: if $\mathbf{G}_j = \lim_{i\to\omega} G_j^{(i)}$, we have

$$\mu'(\mathbf{G}_1 \cap \mathbf{G}_2) = \lim_{i\to\omega} \mu_i \left(G_1^{(i)} \cap G_2^{(i)}\right),$$
$$\mu'(\mathbf{G}_1 \cup \mathbf{G}_2) = \lim_{i\to\omega} \mu_i \left(G_1^{(i)} \cup G_2^{(i)}\right),$$
$$\mu'(\mathbf{G}_1^c) = \lim_{i\to\omega} \mu_i \left(G_1^c\right).$$

It is clear that μ' is nonnegative and $\mu'(\emptyset) = 0$. It is finitely additive: given disjoint sets $\mathbf{G}_1 = \lim_{i\to\omega} G_1^{(i)}$ and $\mathbf{G}_2 = \lim_{i\to\omega} G_2^{(i)}$ in $\mathcal{S}$, the set I of indices i such that $G_1^{(i)} \cap G_2^{(i)} = \emptyset$ must be a member of the ultrafilter. For these indices, we can do the desired manipulation because each of the μ_i are measures, and since $I \in \omega$ the other indices do not matter in the ultralimit:

$$\begin{aligned}
\mu'(\mathbf{G}_1 \cup \mathbf{G}_2) &= \lim_{i\to\omega} \mu_i \left(G_1^{(i)} \cup G_2^{(i)}\right) \\
&= \lim_{i\to\omega} \left(\mu_i \left(G_1^{(i)}\right) + \mu_i \left(G_2^{(i)}\right)\right) \\
&= \lim_{i\to\omega} \mu_i \left(G_1^{(i)}\right) + \lim_{i\to\omega} \mu_i \left(G_2^{(i)}\right) \\
&= \mu'(\mathbf{G}_1) + \mu'(\mathbf{G}_2).
\end{aligned}$$

Now define $\mu : \mathcal{S} \to \mathbb{R}$ by

$$\mu(\mathbf{H}) = \mathrm{st}(\mu'(\mathbf{H})),$$

where the function st is extended so that it is equal to infinity when its argument is an infinite hyperreal. It is clear that μ is nonnegative and $\mu(\emptyset) = 0$, and finite additivity follows because the function st is a homomorphism. In order to prove that μ is a pre-measure on $\mathcal{S}$, it remains to be shown that it is countably additive, not just finitely additive. This follows via the countable compactness result of Corollary 2.8 above: given a countable union $\cup_{j=1}^{\infty} \mathbf{G}_j$ of pairwise disjoint sets in $\mathcal{S}$ which itself lies in $\mathcal{S}$ and is therefore internal, we can write it as a union of finitely many of the sets, which without loss of generality we write as $\cup_{j=1}^{N} \mathbf{G}_j$. By finite additivity,

$$\mu\left(\bigcup_{j=1}^{N} \mathbf{G}_j\right) = \sum_{j=1}^{N} \mu(\mathbf{G}_j),$$

so finite additivity here implies countable additivity. Therefore μ is a pre-measure on $\mathcal{S}$. It is obvious that since the μ_i are probability measures, so will μ.

By the Carathéodory extension theorem, we can extend μ to a measure on the completion $\mathcal{A}$ of the σ-algebra generated by $\mathcal{S}$ (see [25]). See [7] or, especially, [4] for details of the Carathéodory construction in the context of ultraproduct measures.

Lemma 3.1. *The measure μ defined above on $\mathcal{A}$ is invariant under the group action.*

Proof. First consider $\mu' : \mathcal{S} \to {}^*\mathbb{R}$. By the definition of the group structure, for $\mathbf{G} \in \mathcal{S}$ and $a \in \mathbf{A}$ we have

$$\mu'(\mathbf{G} + a) = \lim_{i\to\omega} \mu_i(G_i + a_i) = \lim_{i\to\omega} \mu_i(G_i) = \mu'(\mathbf{G}).$$

Taking standard parts, then, μ restricted to $\mathcal{S}$ is invariant under the group action. It is clear that the Carathéodory construction does not affect invariance with respect to the group action, for the outer measure (which restricts to the Carathéodory measure) is defined as

$$\mu_*(\mathbf{E}) = \inf\left\{\sum_{j=1}^{\infty}\mu(\mathbf{E}_j) : \mathbf{E} \subset \bigcup_{j=1}^{\infty}\mathbf{E}_j, \text{where } \mathbf{E}_j \in \mathcal{S}(\mathbf{A}) \text{ for all } j\right\}. \qquad \square$$

For the second method of construction, we proceed as follows. We need one preliminary lemma, which shows us that "ultralimit convergence" gives us some control over actual convergence, at least for bounded sequences and for a set of indices in the ultrafilter:

Lemma 3.2. *Let $a = \lim_{i\to\omega} a_i$, where each $a_i \in \mathbb{R}$ and the a_i are uniformly bounded. For all real $\epsilon > 0$, considering $\mathbb{R}$ as a subset of ${}^*\mathbb{R}$,*

$$\{i \in \mathbb{N} : a_i \in [\mathrm{st}(a) - \epsilon, \mathrm{st}(a) + \epsilon]\} \in \omega.$$

Proof. Since the a_i are bounded, $\mathrm{st}(a)$ is finite and differs from a only by an infinitesimal. Without loss of generality, it suffices to assume that $\mathrm{st}(a) = 0$, because we can then prove the result for any other sequence simply by translating; thus, a is infinitesimal and therefore in $[-\epsilon, \epsilon]$. Consider the language of the real numbers extended by the constant ϵ. By Łoś' theorem in this language, $a \in [-\epsilon, \epsilon]$ implies that $\{i \in \mathbb{N} : a_i \in [-\epsilon, \epsilon]\} \in \omega$, as this formula is certainly expressible in first-order language. $\square$

Construct $\mu : \mathcal{S}(\mathbf{A}) \to \mathbb{R}$ exactly as before. Define a *null set* to be any subset $\mathbf{E}$ of $\mathbf{A}$ such that for each $\epsilon > 0$, there is a set $\mathbf{F} \in \mathcal{S}(\mathbf{A})$ such that $\mathbf{E} \subset \mathbf{F}$ and $\mu(\mathbf{F}) < \epsilon$. Clearly, if $\mathbf{E}$ is a null set, then any subset of $\mathbf{E}$ is also a null set. It turns out that $\mathcal{S}$ is almost a σ-algebra in a sense made precise by the following lemma:

Lemma 3.3. *Let $\boldsymbol{G}_1, \boldsymbol{G}_2, \ldots$ be an increasing family of sets in $\mathcal{S}(\boldsymbol{A})$ and let*

$$\boldsymbol{E} = \bigcup_{j=1}^{\infty} \boldsymbol{G}_j.$$

Then there is a set $\boldsymbol{G} \in \mathcal{S}(\boldsymbol{A})$ such that $\boldsymbol{E} \subset \boldsymbol{G}$, $\mu(\boldsymbol{G}) = \lim_{j\to\infty}\mu(\boldsymbol{G}_j)$, and $\boldsymbol{G} \setminus \boldsymbol{E}$ is a null set.

Proof. Let the sets $G_j^i \in A_i$ be given such that

$$\mathbf{G}_j = \lim_{i\to\omega} G_j^i,$$

and let

$$S_j = \left\{i \in \mathbb{N} : |\mu_i(G_j^i) - \mu(\mathbf{G}_j)| \le \frac{1}{2^j}\right\},$$

a set of indices for which the measures of the G_j^i are particularly close to that of their ultralimit. By Lemma 3.2, $S_j \in \omega$ for each j.

Let $G^i \in A_i$ be defined as follows: if $i \in S_1, S_2, \ldots, S_m$ but $i \notin S_{m+1}$, let $G^i = \bigcup_{j=1}^m G_j^i$. As one might expect, if i is in all of the S_j, let $G^i = \bigcup_{j=1}^{\infty} G_j^i$. Then define

$$\mathbf{G} = \lim_{i\to\omega} G^i.$$

We have to check three things, the third of which is an immediate consequence of the first two, so it suffices to verify that $\mathbf{E} \subset \mathbf{G}$ and that $\mu(\mathbf{G}) = \lim_{j\to\infty}\mu(\mathbf{G}_j)$. First, fix j.

The set of indices i for which $G^i_j \subset G^i$ is a superset of the intersection $S_1 \cap S_2 \cap \ldots \cap S_j$, which is in ω, so

$$\{i \in \mathbb{N} : G^i_j \subset G^i\} \in \omega,$$

which implies that $\mathbf{G}_j \subset \mathbf{G}$ for each j, which implies that $\mathbf{E} \subset \mathbf{G}$. The second fact follows by summing the measures over j and using the definition of S_j. $\square$

Lemma 3.4. *The set of null sets is closed under countable union.*

Proof. Suppose that $\mathbf{E}_1, \mathbf{E}_2, \ldots$ are null sets, and choose any $\epsilon > 0$. Let $\mathbf{F}^\epsilon_j$ be a set in $\mathcal{S}(\mathbf{A})$ containing $\mathbf{E}_j$ and such that

$$\mu(\mathbf{F}^\epsilon_j) < \frac{\epsilon}{2^j}.$$

Define

$$\mathbf{G}^\epsilon_j = \bigcup_{i=1}^{j} \mathbf{F}^\epsilon_i,$$

so the $\mathbf{G}^\epsilon_j$ form an increasing family of sets in $\mathcal{S}$. By the previous lemma, there is a set $\mathbf{G}^\epsilon \in \mathcal{S}$ containing all of the $\mathbf{G}^\epsilon_i$ (and therefore all of the $\mathbf{F}^\epsilon_i$) and such that

$$\mu(\mathbf{G}^\epsilon) \leq \sum_{j=1}^{\infty} \frac{\epsilon}{2^j} = \epsilon,$$

so we are done. $\square$

Finally, define a set $\mathbf{E} \subset \mathbf{A}$ to be measurable if there is a set $\mathbf{F} \in \mathcal{S}(\mathbf{A})$ such that the symmetric difference $\mathbf{E} \triangle \mathbf{F}$ is a null set, and define $\mu(\mathbf{E})$ in that case to be equal to $\mu(\mathbf{F})$. It is easy to check that this is well-defined, for if we have two sets $\mathbf{F}_1$ and $\mathbf{F}_2$ such that both symmetric differences are null sets, their symmetric difference will be a null set in $\mathcal{S}(\mathbf{A})$; that is, a set of measure zero, so their measures will agree. In particular, all null sets are measurable, with measure zero.

Let the measurable sets be denoted $\mathcal{A}(\mathbf{A})$. It follows from the following theorem and the uniqueness of Carathéodory extensions for σ-finite spaces that we have constructed "by hand" the same measure as before.

Theorem 3.5. *So defined, $\mathcal{A}(\boldsymbol{A})$ is a σ-algebra with a probability measure μ, and this measure space is complete.*

Proof. Call two measurable sets equivalent, with the symbol $\equiv$, if their symmetric difference is a null set. Two equivalent sets clearly have the same measure. Let $\mathbf{E}_1, \mathbf{E}_2, \ldots$ be measurable sets, so there exist sets $\mathbf{F}_1, \mathbf{F}_2, \ldots$ in $\mathcal{S}$ such that $\mathbf{E}_j \equiv \mathbf{F}_j$ for each j. It is easy to check that

$$\mathbf{A} \setminus \mathbf{E}_1 \equiv \mathbf{A} \setminus \mathbf{F}_1,$$
$$\mathbf{E}_1 \cup \mathbf{E}_2 \equiv \mathbf{F}_1 \cup \mathbf{F}_2,$$
$$\mathbf{E}_1 \cap \mathbf{E}_2 \equiv \mathbf{F}_1 \cap \mathbf{F}_2.$$

This demonstrates that $\mathcal{A}$ is closed under finite Boolean operations and μ is a finitely additive probability measure on it. To prove that $\mathcal{A}$ is closed under countable unions and μ

is countably additive, it is enough to prove that if the $\mathbf{E}_j$ are disjoint, there is a $\mathbf{E} \in \mathcal{S}$ such that

$$\bigcup_{j=1}^{\infty} \mathbf{E}_j \equiv \mathbf{E}$$

and

$$\sum_{j=1}^{\infty} \mu(\mathbf{E}_j) = \mu(\mathbf{E}).$$

(This works because being closure under countable disjoint unions implies closure under arbitrary countable unions).

As before, define

$$\mathbf{G}_j = \bigcup_{i=1}^{j} \mathbf{E}_i,$$

so the $\mathbf{G}_j^{\epsilon}$ form an increasing family of sets in $\mathcal{S}$, and

$$\bigcup_{j=1}^{\infty} \mathbf{G}_j = \bigcup_{j=1}^{\infty} \mathbf{E}_j.$$

By Lemma 3.3, there exists an $\mathbf{E} \in \mathcal{S}$ containing that union such that

$$\mathbf{E} \setminus \bigcup_{j=1}^{\infty} \mathbf{E}_j$$

is a null set and

$$\mu(\mathbf{E}) = \lim_{j \to \infty} \mu(\mathbf{G}_j) = \sum_{j=1}^{\infty} \mu(\mathbf{E}_j),$$

so we are done.

Completeness is clear; if $\mathbf{E}$ is a subset of a measure-zero set, then it certainly differs from that set by a null set and is therefore in $\mathcal{A}$. □

Remark It is worth discussing why one has to deal with the σ-algebra $\mathcal{A}$ instead of a topology on $\mathbf{A}$. To define a topology on the ultraproduct, we would certainly want all ultraproducts of open sets to be open, so we could simply take these sets to be a basis for this topology. However, this topology will not have nice analytic properties: for example, if each A_i has the discrete topology (as would be natural on, for example, $\mathbb{Z}$ or a finite group), then every point of $\mathbf{A}$ will be made an open set; hence $\mathbf{A}$ would be endowed with the discrete topology. This is unhelpful for our purposes: we are interested in structures beyond the classical Fourier theory, so insofar as linear characters on locally compact groups (such as any discrete group) already span the relevant function spaces they do not give any new information.

A partial way around this problem is by defining, instead of a topology, a σ-topology, which is a collection τ of subsets of $\mathbf{A}$ obeying the following axioms:

(i) The empty set and $\mathbf{A}$ are in τ.
(ii) The family τ is closed under finite intersection.
(iii) The family τ is closed under countable union.

Thus a σ-topology is a weakening of the usual notion of a topology, allowing only countably many unions rather than arbitrary unions. We can define the standard σ-topology $\mathcal{O}(\mathbf{A})$ as the collection of countable unions of ultraproducts of measurable sets; the axioms above are immediate since ultraproducts respect finite Boolean operations.

With this loosening, however, the group structure does not behave well. One would perhaps like the above-defined measure to be the σ-topological equivalent of a Haar measure, but this is not achievable. Firstly, the group addition function $+ : \mathbf{A} \times \mathbf{A} \to \mathbf{A}$ will not be continuous with respect to $\mathcal{O}(\mathbf{A}) \times \mathcal{O}(\mathbf{A})$, as would be required for a "σ-topological group." Secondly (if we temporarily relax our assumption that the A_i are probability spaces), not all compact sets have finite measure: for example, take $A_i = \mathbb{Z}_i$, the cyclic group of order i, with the discrete topology, and let μ_i be the counting measure on A_i so that $\mu_i(A_i) = i$. All of the A_i are obviously compact. By Tychonoff's theorem, the product $\prod_i A_i$ is compact, and as an arbitrary quotient of compact groups is compact, the ultraproduct $\mathbf{A}$ is compact as well. But

$$\mu(\mathbf{A}) = \operatorname{st}\left(\lim_{i\to\omega} \mu_i(A_i)\right) = \operatorname{st}\left(\lim_{i\to\omega} i\right) = \infty.$$

Finally, we have nothing close to a result like the one on uniqueness of Haar measures; given topological groups A_i and their product $\mathbf{A}$, there are many possibilities for the measure on $\mathbf{A}$. For example, if we take all of the μ_i as probability measures, then $\mu(\mathbf{A}) = 1$, which is clearly not related to the measure of the above example by a rescaling.

These considerations necessitate a more measure-theoretic, less topological approach than in the case of classical Fourier analysis. △

3.2. **Measurable functions on the ultraproduct.** We will need some results about measurable functions on the ultraproduct space $\mathbf{A}$. Let $f_i : A_i \to \mathbb{R}$ be a sequence of measurable functions on the A_i, and let $f : \mathbf{A} \to \overline{\mathbb{R}}$ be given by $f = \operatorname{st}(\lim_{i\to\omega} f_i)$. Since we are not assuming that the f_i are bounded, f is a function to the extended real numbers $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$.

Lemma 3.6. *So defined, f is measurable on $\mathbf{A}$.*

Proof. Throughout, let $\{a_i\}_i^\infty$, $a_i \in A_i$ denote a sequence whose ultralimit is $a \in A$. For measurability, it suffices to prove that

$$f_{[-\infty,d]} = \{a \in \mathbf{A} : -\infty \le f(a) \le d\} = \left\{a \in \mathbf{A} : -\infty \le \operatorname{st}\left(\lim_{i\to\omega} f_i(a_i)\right) \le d\right\}$$

is a measurable set for every $d \in \mathbb{R}$.

If b is a hyperreal, it is easy to see that $-\infty \le \operatorname{st}(b) \le d$ if and only if, for every real $\epsilon > 0$, $b \le d + \epsilon$, or equivalently for all $n \in \mathbb{N}$, $b \le d + 1/n$. Therefore $a \in A$ is in the above set if and only if for every $n \in \mathbb{N}$,

$$\{i \in \mathbb{N} : f_i(a_i) \le d + 1/n\} \in \omega.$$

Let

$$P_{n,i} = \{a_i : f_i(a_i) \le d + 1/n\}.$$

Letting $P_n = \lim_{i\to\omega} P_{n,i}$, it is clear that

$$f_{[-\infty,d]} = \bigcap_{n=1}^{\infty} P_n.$$

But each $P_{n,i}$ is measurable because f_i is, so P_n and hence $f_{[c,d]}$ is as well. □

In the other direction, we can ask to what extent we can represent a given μ-measurable function on $\mathbf{A}$ as an ultraproduct of μ_i-measurable functions on the A_i. The explicit construction of μ suggests that, as every μ-measurable set is "almost" (i.e., up to a null set) an ultraproduct of measurable sets in A_i, each μ-measurable function should be "almost" an ultraproduct of measurable functions. This is indeed the case.

First, we need a lemma. For a function h into $\mathbb{R}$ or ${}^*\mathbb{R}$ and $n \in {}^*\mathbb{R}$ let

$$(h \wedge n)(a) = \begin{cases} h(a) & \text{if } h(a) < n, \\ n & \text{otherwise,} \end{cases}$$

so we have both $(h \wedge n) \leq h$ and $(h \wedge n) \leq n$.

Lemma 3.7. *If $f_i : A_i \to \mathbb{R}$ is a sequence of functions and $f = \mathrm{st}(\lim_{i\to\omega} f_i)$, then*

$$f \wedge n = \mathrm{st}\left(\lim_{i\to\omega}(f_i \wedge n)\right).$$

Proof. This is a straightforward verification. Let $a = \lim_{i\to\omega} a_i$. First presume that $f(a) < n$. Then on the one hand, $(f \wedge n)(a) = f(a)$. On the other hand, by the definition of the standard part and Łoś' theorem,

$$f(a) < n \implies \left(\lim_{i\to\omega} f_i\right)(a) \leq n \implies \{i \in \mathbb{N} : f_i(a_i) \leq n\} \in \omega,$$

so we have that

$$\{i \in \mathbb{N} : f_i(a_i) = (f_i \wedge n)(a_i)\} \in \omega,$$

which implies that

$$f(a) = \mathrm{st}\left(\lim_{i\to\omega} f_i\right)(a) = \mathrm{st}\left(\lim_{i\to\omega}(f_i \wedge n)\right)(a).$$

Therefore the equality is verified if $f(a) < n$.

Otherwise, we have $f(a) \geq n$, so on the one hand we have $(f \wedge n)(a) = n$. On the other hand, for every real $\epsilon > 0$, we have

$$f(a) \geq n \implies \left(\lim_{i\to\omega} f_i\right)(a) \geq n - \epsilon \implies \{i \in \mathbb{N} : f_i(a_i) \geq n - \epsilon\} \in \omega,$$

so we have that

$$\{i \in \mathbb{N} : |f_i(a_i) - (f_i \wedge n)(a_i)| < \epsilon\} \in \omega$$

for every real $\epsilon > 0$. This implies that

$$\left|\left(\lim_{i\to\omega} f_i\right)(a) - \left(\lim_{i\to\omega}(f_i \wedge n)\right)(a)\right| < \epsilon$$

for every real $\epsilon > 0$, so if we take standard parts we get equality:

$$f(a) = \mathrm{st}\left(\lim_{i\to\omega} f_i\right)(a) = \mathrm{st}\left(\lim_{i\to\omega}(f_i \wedge n)\right)(a).$$

The equality is therefore verified. □

Proposition 3.8. *For every measurable function $g : \boldsymbol{A} \to \overline{\mathbb{R}}$, there exists a sequence of measurable functions $f_i : A_i \to \mathbb{R}$ such that if $f = \mathrm{st}(\lim_{i\to\omega} f_i)$, then $f = g$ almost everywhere with respect to μ. Furthermore, if g is bounded, then the f_i can be chosen so as to be uniformly bounded (above or below) with the same bound.*

Proof. Let $(q_n)_{n\in\mathbb{N}}$ be an enumeration of all standard rationals, and let

$$\mathbf{E}_n = \{a \in \mathbf{A} : g(a) \le q_n\},$$

which is measurable because g is. These sets form an "increasing sequence" in the sense that $\mathbf{E}_n \subseteq \mathbf{E}_m$ whenever $q_n \le q_m$. By the construction of $\mathcal{A}(\mathbf{A})$, there are sets $\mathbf{F}_n \in \mathcal{S}(\mathbf{A})$ such that $\mu(\mathbf{E}_n \triangle \mathbf{F}_n) = 0$, and we can certainly choose the $\mathbf{F}_n$ to be increasing in the same sense as well.

By countable comprehension, extend the sequence of $\mathbf{F}_n$ to an internal sequence $(\mathbf{F}_n)_{n\in{}^*\mathbb{N}}$. We can express the fact that $\mathbf{F}_n \subseteq \mathbf{F}_m$ if $q_n \le q_m$ as a first-order sentence, so by the overspill principle there is some infinite K such that this nesting property holds for all $n, m \le K$.

Order the set $\{q_n\}_{n\le K}$ as $q_{i_1} < q_{i_2} < \ldots < q_{i_K}$. Let $f' : \mathbf{A} \to {}^*\mathbb{R}$ be given by

$$f'(a) = \begin{cases} q_{i_j} & \text{if } a \in \mathbf{F}_{i_j} \setminus \mathbf{F}_{i_{j-1}} \\ q_{i_K} + 1 & \text{if } a \notin \mathbf{F}_{i_K}. \end{cases},$$

where we have made the obvious convention $\mathbf{F}_{i_0} = \emptyset$. So defined, f is clearly the ultralimit of functions $f_i : A_i \to \mathbb{R}$ because it is a simple function defined with internal sets. Furthermore, outside the set

$$\bigcup_{n\in\mathbb{N}} (\mathbf{E}_n \triangle \mathbf{F}_n),$$

which is the countable union of null sets and therefore itself null, we have $f'(a) \le q_n$ if and only $g(a) \le q_n$ for all $n \in \mathbb{N}$, hence if $f = \mathrm{st}(f')$ we have $f = g$ almost everywhere.

If g is bounded above, say by n, then by Lemma 3.7 the sequence $f_i \wedge n$ is a lifting of g. The bounded below case follows analogously. □

Given a measurable $g : \mathbf{A} \to \mathbb{R}$, we will call a sequence $f_i : A_i \to \mathbb{R}$ given by the above proposition a *lifting* of g. A lifting will be highly nonunique in general.

3.3. **Integration theory on the ultraproduct.** Standard measure theory gives rise to an integral defined on μ-measurable functions from the ultraproduct group $\mathbf{A}$. For ultralimits of measurable functions, we can also integrate first on the A_i with respect to the measures μ_i, then take the ultralimit. The following proposition shows that for uniformly bounded functions, it does not matter these two operations — ultralimit and integral — are taken.

Proposition 3.9. *If the f_i are uniformly bounded and $f = \mathrm{st}(\lim_{i\to\omega} f_i)$,*

$$\int_{\boldsymbol{A}} f \, d\mu = \mathrm{st}\left(\lim_{i\to\omega} \int_{A_i} f_i \, d\mu_i\right).$$

Proof. Fix N and approximate by simple functions: for each f_i and integer $j \ge 0$, let $g_i(a_i) = j/N$ precisely when

$$\frac{j}{N} \le f_i(a_i) < \frac{j+1}{N},$$

while for each integer $j < 0$, let $g_i(a_i) = (j+1)/N$ precisely when this condition is satisfied. Then g_i is a measurable function on A_i, $|g_i(a_i)| \le| f_i(a_i)|$ for every a_i (so in particular the g_i are uniformly bounded), and $|f_i - g_i| \le 1/N$ uniformly. Since we are integrating over probability spaces,

$$\left|\int_{A_i} f_i - \int_{A_i} g_i\right| \le \frac{1}{N}.$$

Consider $g = \mathrm{st}(\lim_{i\to\omega} g_i)$, which by the above lemma is measurable. It is clear that $|f - g| < 1/N$ uniformly, so again because $\mathbf{A}$ is a probability space

$$\left| \int_{\mathbf{A}} f - \int_{\mathbf{A}} g \right| \leq \frac{1}{N}.$$

By taking N large, therefore, it suffices to check the proposition for bounded simple functions, and by linearity (of both the ultralimit and the integral) it suffices to check the proposition for characteristic function of measurable sets.

So let $f_i = \chi_{B_i}$; then it is easy to check that $f = \chi_B$ where $B = \lim_{i\to\omega} B_i$. We want to show that $\mu(B) = \mathrm{st}\left(\lim_{i\to\omega} \mu_i(B_i)\right)$. But this is precisely the definition of the measure μ. □

Corollary 3.10. *If $g : \boldsymbol{A} \to \overline{\mathbb{R}}$ is μ-measurable and bounded, for any bounded lifting g_i the following equality holds:*

$$\int_{\boldsymbol{A}} g \; d\mu = \mathrm{st}\left(\lim_{i\to\omega} \int_{A_i} g_i \; d\mu_i \right).$$

Proof. By definition, a the ultraproduct of a lifting is μ-a.e. equal to the original function, so the result follows immediately when combined with the second part of Proposition 3.8. □

Without a uniform bound, we cannot expect to preserve the above equality. For a simple couterexample, let $A_i = [0,1]$ equipped with the usual Lebesgue measure and define

$$f_i(x) = \begin{cases} i & \text{if } x \leq \frac{1}{i} \\ 0 & \text{otherwise.} \end{cases}$$

The standard part of the ultraproduct, $f : {}^*[0,1] \to \overline{\mathbb{R}}$, is certainly equal to zero on all non-infinitesimal elements of ${}^*[0,1]$ (that is, all elements not infinitely close to zero). The set $\mathbf{I}$ of infinitesimal elements is a null set because $\mathbf{I} \subset [0,\epsilon] \subset {}^*[0,1]$ for all ϵ, and by trivial calculation $\mu([0,\epsilon]) = \epsilon$. Therefore f is zero except on a set of measure zero, so

$$\int_{{}^*[0,1]} f \; d\mu = 0.$$

On the other hand, it is clear that

$$\mathrm{st}\left(\lim_{i\to\omega} \int_{[0,1]} f_i \; d\mu_i \right) = \mathrm{st}\left(\lim_{i\to\omega} 1 \right) = 1.$$

For arbitrary positive μ-measurable functions, the best we can do is the following, which is reminiscent of Fatou's Lemma:

Proposition 3.11. *Let f_i and f be as above, where no boundedness is assumed but we do have $f \geq 0$. Then*

$$\int_{\boldsymbol{A}} f \; d\mu \leq \mathrm{st}\left(\lim_{i\to\omega} \int_{A_i} f_i \; d\mu_i \right).$$

Proof. We proceed somewhat analogously to the proof of Fatou's lemma. Suppose g is a bounded measurable function on $\mathbf{A}$ with $0 \leq g \leq f$. By Proposition 3.8, there is a lifting $g_i : A_i \to \mathbb{R}$ such that

$$g = \mathrm{st}\left(\lim_{i\to\infty} g_i \right)$$

μ-almost everywhere. Since we are ultimately only considering integrals of g, we can identify the two functions without harm. By the definition of the order relation on ${}^*\mathbb{R}$ and the monotonicity of the integral, we have the following chain of inferences:

$$\begin{aligned} &g \le f \\ \Longrightarrow &\{i \in \mathbb{N} : g_i \le f_i\} \in \omega \\ \Longrightarrow &\{i \in \mathbb{N} : \int_{A_i} g_i \, d\mu_i \le \int_{A_i} f_i \, d\mu_i\} \in \omega \\ \Longrightarrow &\lim_{i\to\omega} \int_{A_i} g_i \, d\mu_i \le \lim_{i\to\omega} \int_{A_i} f_i \, d\mu_i. \end{aligned}$$

Taking the standard part certainly preserves $\le$, so we actually have

$$\mathrm{st}\left(\lim_{i\to\omega} \int_{A_i} g_i \, d\mu_i\right) \le \mathrm{st}\left(\lim_{i\to\omega} \int_{A_i} f_i \, d\mu_i\right).$$

By Corollary 3.10, as g is bounded, the left hand side above is equal to the integral of g on the ultraproduct; that is,

$$\int_{\mathbf{A}} g \, d\mu \le \mathrm{st}\left(\lim_{i\to\omega} \int_{A_i} f_i \, d\mu_i\right).$$

Taking the supremum over all μ-measurable g such that $0 \le g \le f$ yields the desired result. □

We want to be able to interchange integrals and ultralimits for a broader class of functions than L^∞. In order to get results in this direction, we introduce the following definition, which characterizes a large class of functions for which the result of Proposition 3.9 holds. Let $f_i : A_i \to \mathbb{R}$ be a sequence of μ_i-measurable functions. Call this sequence *S-integrable* if the following two conditions hold:

(i) $\lim_{i\to\omega} \int_{A_i} |f_i| \, d\mu_i$ is finite,
(ii) if $\mathbf{E} = \lim_{i\to\omega} E_i \subset A_i$ is in $\mathcal{S}(\mathbf{A})$ and $\mu(\mathbf{E}) = 0$, then

$$\mathrm{st}\left(\lim_{i\to\omega} \int_{E_i} |f_i| \, d\mu_i\right) = 0.$$

We are essentially requiring that the functions do not concentrate too much mass in an infinitesimal neighborhood.

The following equivalence holds:

Proposition 3.12. *Let $f_i : A_i \to \mathbb{R}$ be a sequence of μ_i-measurable functions. Then the following are equivalent:*

(i) the f_i are S-integrable,
(ii) if $f = \mathrm{st}(\lim_{i\to\omega} f_i)$,

$$\int_{\mathbf{A}} f \, d\mu = \mathrm{st}\left(\lim_{i\to\omega} \int_{A_i} f_i \, d\mu_i\right).$$

Proof. If f_i^+ and f_i^- are the positive and negative parts of the f_i, respectively, then it is clear that $\{f_i\}$ is S-integrable if and only if both $\{f_i^+\}$ and $\{f_i^-\}$ are. Therefore it suffices to assume that $f_i \ge 0$ for each i.

First assume (i). By the definition of S-integrability, the right-hand side is finite, so by Proposition 3.11 so is

$$I = \int_{\mathbf{A}} f \, d\mu.$$

By Lemma 3.7 and Proposition 3.9, for all $n \in \mathbb{N}$ we have

$$\int_{\mathbf{A}} (f \wedge n) \, d\mu = \operatorname{st}\left(\lim_{i \to \omega} \int_{A_i} f_i \wedge n \, d\mu_i \right).$$

Certainly, then, we have

$$\begin{aligned} \lim_{i \to \omega} \int_{A_i} f_i \wedge n \, d\mu_i &\leq \int_{\mathbf{A}} (f \wedge n) \, d\mu + \frac{1}{n} \\ &\leq I + \frac{1}{n}. \end{aligned}$$

By overflow, there is an infinite K such that, if $K = \lim_{i \to \omega} k_i$,

$$\lim_{i \to \omega} \int_{A_i} f_i \wedge k_i \, d\mu_i \leq I + \frac{1}{K}.$$

Consider the set $\mathbf{E} = \{a : \{f_i(a_i) > k_i\} \in \omega \text{ and } a = \lim_{i \to \omega} a_i\} \subset \mathbf{A}$. By measurability, $\mathbf{E} \in \mathcal{A}$, and by construction of $\mathcal{A}$ there is an $\mathbf{F} \in \mathcal{S}$ with $\mu(\mathbf{E} \triangle \mathbf{F}) = 0$. Let

$$\mathbf{F} = \lim_{i \to \omega} F_i.$$

We certainly have the following:

$$\lim_{i \to \omega} \int_{A_i} f_i \, d\mu_i \leq \lim_{i \to \omega} \int_{F_i} f_i \, d\mu_i + \lim_{i \to \omega} \int_{A_i} f_i \wedge k_i \, d\mu_i \leq \lim_{i \to \omega} \int_{F_i} f_i \, d\mu_i + I + \frac{1}{K}.$$

Since $f(a) < \infty$ almost everywhere as I is finite, $\mu(\mathbf{E}) = \mu(\mathbf{F}) = 0$. By S-integrability,

$$\operatorname{st}\left(\lim_{i \to \omega} \int_{F_i} f_i \, d\mu_i \right) = 0,$$

so taking standard parts of the above inequality,

$$\operatorname{st}\left(\lim_{i \to \omega} \int_{A_i} f_i \, d\mu_i \right) \leq I = \int_{\mathbf{A}} f \, d\mu.$$

When combined with Proposition 3.11, this gives (ii).

Now assume (ii). It is immediate that $\lim_{i \to \omega} \int_{A_i} f_i \, d\mu_i$ is finite. Let $\mathbf{E} = \lim_{i \to \omega} E_i$ be in $\mathcal{S}$ and such that $\mu(\mathbf{E}) = 0$. Then

$$\begin{aligned} \operatorname{st}\left(\lim_{i \to \omega} \int_{A_i} f_i \, d\mu_i \right) &= \operatorname{st}\left(\lim_{i \to \omega} \int_{A_i \setminus E_i} f_i \, d\mu_i \right) + \operatorname{st}\left(\lim_{i \to \omega} \int_{E_i} f_i \, d\mu_i \right) \\ &\leq \operatorname{st}\left(\lim_{i \to \omega} \int_{A_i \setminus E_i} f_i \, d\mu_i \right) \\ &\leq \int_{\mathbf{A} \setminus \mathbf{E}} f \, d\mu \\ &= \int_{\mathbf{A}} f \, d\mu. \end{aligned}$$

The third line comes from Proposition 3.11 applied to the set $\mathbf{A}_i \setminus \mathbf{E} = \lim_{i\to\omega}(A_i \setminus E_i)$, while the fourth line comes from the fact that $\mu(\mathbf{E}) = 0$. By (ii), all of the inequalities are equalities, so in particular

$$\mathrm{st}\left(\lim_{i\to\omega}\int_{E_i} f_i \, d\mu_i\right) = 0,$$

so (i) is proved. □

The S-integrability condition gives a necessary and sufficient condition that a μ-measurable function be integrable.

Proposition 3.13. *A μ-measurable function $f : \boldsymbol{A} \to \overline{\mathbb{R}}$ is μ-integrable if and only if it has an S-integrable lifting.*

Proof. If f has an S-integrable lifting $\{f_i\}$, then

$$f = \mathrm{st}\left(\lim_{i\to\omega} f_i\right)$$

almost everywhere, and by Proposition 3.12 the latter is integrable.

Conversely, assume f is integrable. We can consider f^+ and f^- separately, which means it suffices to assume $f \geq 0$. Take a lifting $\{f_i\}$ such that $f_i \geq 0$, which we can do by the last part of Proposition 3.8. By Lemma 3.7 and Corollary 3.10, for each finite n,

$$\mathrm{st}\left(\lim_{i\to\omega}\int_{A_i} (f_i \wedge n)\, d\mu_i\right) = \int_{\mathbf{A}} (f \wedge n)\, d\mu,$$

so in particular

$$\left(\lim_{i\to\omega}\int_{A_i} (f_i \wedge n)\, d\mu_i\right) \leq \int_{\mathbf{A}} (f \wedge n)\, d\mu + \frac{1}{n} \leq \int_{\mathbf{A}} f\, d\mu + \frac{1}{n}.$$

By overspill, there is an infinite $K = \lim_{i\to\omega} k_i$ such that

$$\left(\lim_{i\to\omega}\int_{A_i} (f_i \wedge k_i)\, d\mu_i\right) \leq \int_{\mathbf{A}} f\, d\mu + \frac{1}{K}.$$

We claim that that $\{f_i \wedge k_i\}$ is a lifting of f: certainly

$$\mathrm{st}\left(\lim_{i\to\omega}(f_i \wedge k_i)\right)(a) = \infty \iff \mathrm{st}\left(\lim_{i\to\omega} f_i\right)(a) = \infty$$

precisely because $\lim_{i\to\infty} k_i$ is infinite. In the finite case, we also have equality, because the set of i such that k_i is greater than any finite number is an element of the ultrafilter, so we can disregard the remainder.

Taking standard parts of the above inequality, we get

$$\mathrm{st}\left(\lim_{i\to\omega}\int_{A_i} (f_i \wedge k_i)\, d\mu_i\right) \leq \int_{\mathbf{A}} f\, d\mu = \int_{\mathbf{A}} \mathrm{st}\left(\lim_{i\to\omega}(f_i \wedge k_i)\right)\, d\mu_i$$

because by definition the standard part of the ultraproduct of a lifting is almost everywhere equal to the original function. By Proposition 3.11, the opposite inequality is also true, so

$$\mathrm{st}\left(\lim_{i\to\omega}\int_{A_i} (f_i \wedge k_i)\, d\mu_i\right) = \int_{\mathbf{A}} \mathrm{st}\left(\lim_{i\to\omega}(f_i \wedge k_i)\right)\, d\mu_i.$$

By Proposition 3.12, $\{f_i \wedge k_i\}$ is an S-integrable lifting, as desired. □

Finally, we have a result, first realized by Lindström in [20], that is especially important in the case $p = 2$:

Proposition 3.14. *If $f_i : A_i \to \mathbb{R}$ is a sequence of μ_i-measurable functions with*

$$\mathrm{st}\left(\lim_{i\to\omega}\int_{A_i} |f_i|^p \, d\mu_i\right) < \infty$$

for some $p > 1$, then the f_i are S-integrable.

Proof. Because the A_i are all probability spaces, $\int_{A_i} |f_i| \, d\mu_i$ is bounded in terms of $\int_{A_i} |f_i|^p \, d\mu_i$. Specifically, let $B_i = \{a_i : |f_i(a_i)| < 1\}$ and let $C_i = \{a_i : |f_i(a_i)| \geq 1\}$. Then

$$\int_{A_i} |f_i| \, d\mu_i = \int_{B_i} |f_i| \, d\mu_i + \int_{C_i} |f_i| \, d\mu_i \leq 1 + \int_{A_i} |f_i|^p \, d\mu_i,$$

which immediately implies that

$$\mathrm{st}\left(\lim_{i\to\omega}\int_{A_i} |f_i| \, d\mu_i\right) < \infty.$$

To verify the second condition for S-integrability, consider some $\mathbf{E} = \lim_{i\to\omega} E_i$ such that $\mu(\mathbf{E}) = 0$. By Hölder's inequality, we have

$$\int_{E_i} |f_i| \, d\mu_i \leq \left(\int_{A_i} I_{E_i} \, d\mu_i\right)^{1/q} \left(\int_{A_i} |f_i|^p \, d\mu_i\right)^{1/p},$$

where $1/p + 1/q = 1$. Taking the ultralimit of both sides, we get

$$\begin{aligned}\lim_{i\to\omega}\left(\int_{E_i} |f_i| \, d\mu_i\right) &\leq \lim_{i\to\omega}\left[\left(\int_{A_i} I_{E_i} \, d\mu_i\right)^{1/q}\right] \lim_{i\to\omega}\left[\left(\int_{A_i} |f_i|^p \, d\mu_i\right)^{1/p}\right] \\ &= \lim_{i\to\omega} \mu_i(E_i)^{1/q} \lim_{i\to\omega}\left(\int_{A_i} |f_i|^p \, d\mu_i\right)^{1/p}.\end{aligned}$$

But $\mu(\mathbf{E})$ is zero and the second factor is finite by assumption, so taking standard parts

$$\mathrm{st}\left(\lim_{i\to\omega}\int_{E_i} |f_i| \, d\mu_i\right) = 0.$$

Therefore the f_i are S-integrable. □

4. Product and sub−σ−algebras

4.1. **Definitions and the Fubini theorem.** So far products of ultraproduct σ-algebras have not been discussed, but they play a crucial role in the theory.

The first observation is that we have two obvious ways of putting a σ-algebra on the product $\mathbf{A}^k$. By Lemma 2.10, the product $\mathbf{A}^k$ is unambiguously defined in that we can assume that finite products and ultraproducts commute. However, in general the σ-algebras $\mathcal{A}(\mathbf{A}^k)$, defined by starting with the product σ-algebras on A_i^k and performing the above procedure on their ultraproduct, will be different from the product σ-algebra

$$\mathcal{A}(\mathbf{A}) \times \ldots \times \mathcal{A}(\mathbf{A}).$$

In fact, much of the structure of the theory developed by Szegedy can be traced to the noncommutativity

$$\mathcal{A}(\mathbf{A}) \times \ldots \times \mathcal{A}(\mathbf{A}) \neq \mathcal{A}(\mathbf{A}^k).$$

We have the following inclusion:

Lemma 4.1. *Let $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ be ultraproduct spaces equipped with measures μ_1, μ_2 defined as above. Then*

$$\mathcal{A}(\boldsymbol{A}_1) \times \mathcal{A}(\boldsymbol{A}_2) \subseteq \mathcal{A}(\boldsymbol{A}_1 \times \boldsymbol{A}_2)$$

and the measure $\mu_1 \times \mu_2$ is unambiguously defined on the smaller σ-algebra.

Proof. That the product measure is unambiguously defined simply requires a retracing of the steps in the definition, using Lemma 2.10 wherever necessary.

By definition, $\mathcal{A}(\mathbf{A}_1) \times \mathcal{A}(\mathbf{A}_2)$ is generated by rectangles $\mathbf{E}_1 \times \mathbf{E}_2$, where $\mathbf{E}_1 \in \mathcal{A}(\mathbf{A}_1)$ and $\mathbf{E}_2 \in \mathcal{A}(\mathbf{A}_2)$. It suffices to show that each such rectangle is in $\mathcal{A}(\mathbf{A}_1 \times \mathbf{A}_2)$.

By construction, for $j = 1, 2$ there is a $\mathbf{F}_j$ that is the ultraproduct of finite Boolean combinations of measurable sets such that

$$\mu_j(\mathbf{E}_j \triangle \mathbf{F}_j) = 0.$$

Tracing the construction backwards, it is certainly the case that $\mathbf{F}_1 \times \mathbf{F}_2 \in \mathcal{A}(\mathbf{A}_1 \times \mathbf{A}_2)$. We have the following, for purely set-arithmetical reasons as well as by the definition of the product measure:

$$\begin{aligned}
&(\mu_1 \times \mu_2)((\mathbf{E}_1 \times \mathbf{E}_2) \triangle (\mathbf{F}_1 \times \mathbf{F}_2)) \\
&\qquad \leq \mu_1(\mathbf{E}_1 \triangle \mathbf{F}_1)\mu_2(\mathbf{E}_2 \cup \mathbf{F}_2) + \mu_1(\mathbf{E}_1 \cup \mathbf{F}_1)\mu_2(\mathbf{E}_2 \triangle \mathbf{F}_2) \\
&\qquad = 0.
\end{aligned}$$

Therefore $\mathbf{E}_1 \times \mathbf{E}_2 \in \mathcal{A}(\mathbf{A}_1 \times \mathbf{A}_2)$, as desired. □

In fact, the same proof gives that $\overline{\mathcal{A}(\mathbf{A}_1) \times \mathcal{A}(\mathbf{A}_2)} \subseteq \mathcal{A}(\mathbf{A}_1 \times \mathbf{A}_2)$, where the bar indicates the completion of a measure.

Several iterations of this lemma give that

$$\mathcal{A}(\mathbf{A}) \times \ldots \times \mathcal{A}(\mathbf{A}) \subseteq \mathcal{A}(\mathbf{A}^k).$$

In general the inclusion is proper, as we will see.[2]

[2]See [5] page 557 for a fairly down-to-earth example.

We need some notation for discussing such products. Let $[k]$ denote the set $\{1, \dots, k\}$ and let $S \subseteq [k]$. Let

$$A_{i,S} = \bigoplus_{j \in S} A_i,$$

and let

$$\mathbf{A}^S = \lim_{i \to \omega} A_{i,S}$$

be the ultraproduct of these sets, equipped with σ-algebra $\mathcal{A}(\mathbf{A}^S)$ as described in the previous section.

There is a natural projection map

$$p_S : \mathbf{A}^k \to \mathbf{A}^S$$

which we can use to define σ-algebras on $\mathbf{A}^k$: define

$$\mathcal{A}_S(\mathbf{A}^k) = p_S^{-1}(\mathcal{A}(\mathbf{A}^S)),$$

the σ-algebra of measurable sets that depend only on the coordinates in S. By this definition $\mathcal{A}_{[k]} = \mathcal{A}$, while $\mathcal{A}_\emptyset$ is the trivial σ-algebra on $\mathbf{A}^k$. It is obvious that

$$\mathcal{A}_T \subseteq \mathcal{A}_S, \qquad \text{if } T \subseteq S.$$

Let μ_S denote the measure on $\mathbf{A}^S$; by Lemma 4.1 it does not matter with respect to precisely what σ-algebra we are referring.

Finally, let S^* denote the system of sets $\{T : T \subset S, |T| = |S| - 1\}$, and let $\mathcal{A}_{S^*}$ be the σ-algebra generated by the algebras $\mathcal{A}_T$ such that $T \in S^*$. (Note that it is equivalent to define S^* to consist of all proper subsets of S.) For example, in the case $k = 2$, we simply have that

$$\mathcal{A}_{[k]^*}(\mathbf{A}) = \mathcal{A}(\mathbf{A}) \times \mathcal{A}(\mathbf{A}),$$

while for higher k the σ-algebra $\mathcal{A}_{[k]^*}$ lies somewhere in between the product σ-algebra and $\mathcal{A} = \mathcal{A}_{[k]}$ itself, inclusion-wise.

We can use Fubini's theorem to get results about interchanging the order of integration for functions in product σ-algebras. Because we do not in general have equality in Lemma 4.1, however, we cannot *a priori* say anything about interchanging the order of integration for $\mathcal{A}(\mathbf{A}^k)$-measurable functions. Fortunately, we do have a strengthened Fubini's theorem, developed by Keisler. In order to prove it, we need to develop a lemma about product null sets that is manifestly false outside of the ultraproduct setting, but gives us the extra control needed for the full theorem:

Lemma 4.2. *Let $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ be ultraproduct spaces of the kind constructed above, with measures μ_1 and μ_2, and let $\boldsymbol{E} \subset \boldsymbol{A}_1 \times \boldsymbol{A}_2$. For each $y \in \boldsymbol{A}_2$ define*

$$\boldsymbol{E}_y = \{x \in \boldsymbol{A}_1 : (x, y) \in \boldsymbol{E}\},$$

the slice of $\boldsymbol{E}$ at y. Then the following are equivalent:

(i) $(\mu_1 \times \mu_2)(\boldsymbol{E}) = 0$.

(ii) For each $n \in \mathbb{N}$ there is a set $\boldsymbol{E}_n \in \mathcal{S}(\boldsymbol{A}_1 \times \boldsymbol{A}_2)$ containing $\boldsymbol{E}$ such that $(\mu_1 \times \mu_2)(\boldsymbol{E}_n) < 1/n$.

(iii) For each $n \in \mathbb{N}$ there is a set $\boldsymbol{E}_n \in \mathcal{S}(\boldsymbol{A}_1 \times \boldsymbol{A}_2)$ containing $\boldsymbol{E}$ such that

$$\mu_2 \left(\left\{ y \in \boldsymbol{A}_2 : \mu_1(\boldsymbol{E}_{n,y}) < \frac{1}{n} \right\} \right) \geq 1 - \frac{1}{n}.$$

(iv) For almost all $y \in \boldsymbol{A}_2$, $\mu_1(\boldsymbol{E}_y) = 0$.

Proof (sketch). The implication (i) to (ii) is clear; if $\mathbf{E}$ is itself in $\mathcal{S}(\mathbf{A}_1 \times \mathbf{A}_2)$ then we can take $\mathbf{E}_n = \mathbf{E}$ for all n, while otherwise we can use Lemma 3.3 to find $\mathbf{E}_n$. The implication (ii) to (iii) follows by using Fubini's theorem on the $A_{1,i} \times A_{2,i}$, which works precisely because $\mathbf{E}_n$ is an internal set. The implication (iii) to (iv) follows by the following argument: if we let $\mathbf{L}_n \subset \mathbf{A}_2$ be the set

$$\left\{ y \in \mathbf{A}_2 : \mu_1(\mathbf{E}_{n,y}) \geq \frac{1}{n} \right\},$$

then by the Borel-Cantelli lemma almost all y can be contained in only finitely many of the $\mathbf{L}_n$. For these y, $\mu_1(\mathbf{E}_y) = 0$. The implication (iv) to (i) is obvious. □

Theorem 4.3 (Keisler's Fubini theorem)**.** *Let $S, T \subseteq [k]$, and let S^c denote the complement of S in $[k]$. If $f : \boldsymbol{A}^k \to \overline{\mathbb{R}}$ and $y \in \boldsymbol{A}^{S^c}$, let $f_y : \boldsymbol{A}^S \to \overline{\mathbb{R}}$ be the function obtained from f by holding the S^c coordinates constant at y.*

(i) If $f : \boldsymbol{A}^k \to \overline{\mathbb{R}}$ is measurable with respect to $\mathcal{A}_T(\boldsymbol{A}^k)$ and $y \in \boldsymbol{A}^{S^c}$, then f_y is measurable with respect to $\mathcal{A}_{S\cap T}(\boldsymbol{A}^k)$ for almost all y.

(ii) If $f : \boldsymbol{A}^k \to \overline{\mathbb{R}}$ is integrable with respect to $\mathcal{A}(\boldsymbol{A}^k)$, then

$$\int_{\boldsymbol{A}^k} f \, d\mu_{[k]} = \int_{\boldsymbol{A}^{S^c}} \left(\int_{\boldsymbol{A}^S} f_y(x) \, d\mu_S(x) \right) d\mu_{S^c}(y).$$

Proof. Although the theorem is stated for many coordinates, it is easy to see that it suffices to prove the corresponding version for two coordinates. Namely, let $\mathbf{A}_1$ and $\mathbf{A}_2$ be two ultraproduct spaces, and if $f : \mathbf{A}_1 \times \mathbf{A}_2 \to \overline{\mathbb{R}}$ and $y \in \mathbf{A}_2$ let $f_y : \mathbf{A}_1 \to \overline{\mathbb{R}}$ be the function obtained from f by holding the second coordinate constant at y. Then the simplified theorem states:

(i) If $f : \mathbf{A}_1 \times \mathbf{A}_2 \to \overline{\mathbb{R}}$ is measurable with respect to $\mathcal{A}(\mathbf{A}_1 \times \mathbf{A}_2)$ and $y \in \mathbf{A}^2$, then f_y is measurable with respect to $\mathcal{A}(\mathbf{A}_1)$ for almost all y.

(ii) If $f : \mathbf{A}_1 \times \mathbf{A}_2 \to \overline{\mathbb{R}}$ is integrable with respect to $\mathcal{A}(\mathbf{A}_1 \times \mathbf{A}_2)$, then

$$\int_{\mathbf{A}_1 \times \mathbf{A}_2} f \, (d\mu_1 \times d\mu_2) = \int_{\mathbf{A}_2} \left(\int_{\mathbf{A}_1} f_y(x) \, d\mu_1(x) \right) d\mu_2(y).$$

Implicit in the second statement is the assertion that the inner function (of y) is measurable on $\mathbf{A}_2$, which needs to be proven. Let $\mathbf{A}_1 = \lim_{i \to \omega} A_{1,i}$ and $\mathbf{A}_2 = \lim_{i \to \omega} A_{2,i}$.

Let $\{f_i\}$ be a lifting of f. Then by definition the set

$$\mathbf{E} = \{(x, y) : \operatorname{st} \left(\lim_{i \to \omega} f_i \right) (x, y) \neq f(x, y)\}$$

has measure zero. By the equivalence of (i) and (iv) in Lemma 4.2 and the existence of liftings, the functions

$$f_{i,y}(x) = f_i(x, y)$$

are liftings of f_y for almost all y, so the first part is established.

For the second part, let

$$g(y) = \int_{\mathbf{A}_1} f(x, y) \, d\mu_1(x).$$

Since f is integrable we can pick an S-integrable lifting $\{f_i\}$ and proceed the same way. Let

$$g_i(y) = \int_{A_{1,i}} f_i(x, y)\; d\mu_{1,i}(x).$$

Whenever $f_{i,y}$ is a lifting of f_y it is easy to check that it is an S-integrable lifting as well, so by Proposition 3.12 we have that

$$\mathrm{st}\left(\lim_{i\to\omega} g_i\right)(y) = g(y).$$

Therefore $\{g_i\}$ is a lifting of g whenever $f_{i,y}$ is a lifting of f_y; that is, almost everywhere, so in particular g is measurable.

Now we use Proposition 3.12 twice (once in the first line and once in the fourth), together with the usual Fubini theorem in the second line, to get the following chain of equalities:

$$\begin{aligned}
\int_{\mathbf{A}_1\times\mathbf{A}_2} f\;(d\mu_1 \times d\mu_2) &= \mathrm{st}\left(\lim_{i\to\omega}\int_{A_{1,i}\times A_{2,i}} f_i\;(d\mu_{1,i}\times d\mu_{2,i})\right)\\
&= \mathrm{st}\left(\lim_{i\to\omega}\int_{A_{2,i}}\left(\int_{A_{1,i}} f_i\; d\mu_{1,i}\right)\; d\mu_{2,i}\right)\\
&= \mathrm{st}\left(\lim_{i\to\omega}\int_{A_{2,i}} g_i\; d\mu_{2,i}\right)\\
&= \int_{\mathbf{A}_2} g\; d\mu_2,
\end{aligned}$$

as desired. □

4.2. **Gowers (semi-)norms.** The Gowers norms U^k are now common tools in arithmetic combinatorics and are intimately connected with higher order Fourier analysis. We present a definition and a few properties for the Gowers norms of functions $A \to \mathbb{C}$, where A is an arbitrary abelian group and $(A, \mathcal{B}, \mu)$ is a measure space such that:

(i) The σ-algebra $\mathcal{B}$ is closed under the group action,
(ii) The measure μ is invariant under the group action, and
(iii) Fubini's theorem is true in $(A, \mathcal{B}, \mu)$.

The measures spaces $(\mathbf{A}, \mathcal{A}, \mu)$ constructed above satisfy these three criteria (recall Lemma 3.1), so everything mentioned here applies to that case. The greater generality here is justifiable because of the dearth of sources for this material in the literature; it seems probable that no investigation has thus far been made in comparable generality. For further development in the finite case when A is finite and μ is the counting measure, see [33] or [1]. Throughout this section, all integrals are taken over the entire group A unless otherwise marked.

By $c \in \{0,1\}^k$, we mean c is a length-k sequence of zeroes and ones $c = (c_1, \dots, c_k)$. For f measurable and $k \geq 1$ a natural number, define

$$||f||_{U^k}^{2^k} = \iint\dots\int \prod_{c\in\{0,1\}^k} C^{|c|}[f(x + c_1h_1 + \ldots + c_kh_k)]\; d\mu(x)\; d\mu(h_1)\ldots d\mu(h_k),$$

where $Cf = \overline{f}$ is the complex conjugation operator and $|c| = c_1 + \ldots + c_k$. Essentially, we are integrating (or, in the case of a probability measure, averaging) the function over all parallelepiped structures of dimension k. Define the multiplicative derivative Δ_h on functions $A \to \mathbb{C}$ by

$$\Delta_h f(x) = f(x+h)\overline{f(x)}.$$

It is clear that we can rewrite the Gowers norms as

$$||f||_{U^k}^{2^k} = \int\int \ldots \int \Delta_{h_1} \ldots \Delta_{h_k} f(x)\; d\mu(x)\; d\mu(h_1) \ldots d\mu(h_k).$$

For example, for $k = 2$ we have

$$||f||_{U^2} = \left(\int\int\int f(x)\overline{f(x+h_1)f(x+h_2)} f(x+h_1+h_2)\; d\mu(x)\; d\mu(h_1)\; d\mu(h_2) \right)^{1/4}.$$

while if $k = 1$ the norm reduces to

$$\begin{aligned} ||f||_{U^1} &= \left(\int\int f(x)\overline{f(x+h)}\; d\mu(x)\; d\mu(h) \right)^{1/2} \\ &= \left(\int f(x) \left[\int \overline{f(h)}\; d\mu(h) \right]\; d\mu(x) \right)^{1/2} \\ &= \left| \int f(x)\; d\mu(x) \right|, \end{aligned}$$

the absolute value of the integral of f, where we have made the obvious change of variables $h \mapsto h - x$ in the inner integral.

Let $G^k(A, \mu)$ be the space of functions for which $||f||_{U^k}$ is defined and finite (that is, f is measurable and if we replace f by $|f|$, all integrals are finite). Clearly, U^1 is only a seminorm on $G^1(A, \mu) = L^1(A, \mu)$, and although we will refer to the U^k seminorms as norms, they are in general only seminorms as well, with a nonzero kernel. By an obvious change of variables, the U^k norms are invariant under the group action. It obviously remains to be shown that the U^k are indeed seminorms on $G^k(A, \mu)$. To do this, we will prove several intermediate results.

Lemma 4.4. *For $k \geq 2$, the following inductive formula holds:*

$$||f||_{U^k}^{2^k} = \int ||\Delta_h f||_{U^{k-1}}^{2^{k-1}}\; d\mu(h).$$

Proof. This is a trivial verification. □

In particular, the inductive formula shows that the U^k map to nonnegative reals for all k; this is certainly true for $k = 1$ by the above calculation of U^1 and it is therefore true for all higher k by induction.

We can put more structure on $G^k(A, \mu)$ by defining a sort of inner product taking 2^k arguments instead of two. Given 2^k functions $f_c \in G^k(A, \mu)$, $c \in \{0,1\}^k$, define the U^k product of these functions by

$$\langle (f_c)_{c \in \{0,1\}^k} \rangle_{U^k} = \int\int \ldots \int \prod_{c \in \{0,1\}^k} C^{|c|}[f_c(x + c_1 h_1 + \ldots + c_k h_k)]\; d\mu(x)\; d\mu(h_1) \ldots d\mu(h_k).$$

It is clear that this product is linear in all entries. Furthermore, if $f_c = f$ for all c, we have

$$\langle (f_c)_{c \in \{0,1\}^k} \rangle_{U^k} = ||f||_{U^k}^{2^k}.$$

Proposition 4.5 (Gowers-Cauchy-Schwarz inequality)**.** *Let $f_c \in G^k(A,\mu)$, $c \in \{0,1\}^k$ be 2^k functions, where $k \geq 1$. Then*

$$\left|\langle (f_c)_{c\in\{0,1\}^k}\rangle_{U^k}\right| \leq \prod_{c\in\{0,1\}^k} ||f_c||_{U^k}.$$

Proof. Writing $y = x + h_1$, we have

$$\langle (f_c)_{c\in\{0,1\}^k}\rangle_{U^k} = \int \dots \int \left(\int \prod_{\substack{c\in\{0,1\}^k \\ c_1=0}} C^{|c|}[f_c(x + c_2h_2 + \ldots + c_kh_k)]\, d\mu(x) \right)$$
$$\left(\int \prod_{\substack{c\in\{0,1\}^k \\ c_1=1}} C^{|c|}[f_c(y + c_2h_2 + \ldots + c_kh_k)]\, d\mu(y) \right) d\mu(h_2)\dots d\mu(h_k).$$

Temporarily denote the function of the variables $h_2, \ldots, h_k$ in the first set of parentheses by $s(h_2, \ldots, h_k)$ and the function in the second set of parentheses by $t(h_2, \ldots, h_k)$, so we end up with the expression

$$\int \dots \int s \cdot t\, d\mu(h_2)\dots d\mu(h_k).$$

We will apply Cauchy-Schwarz over A^{k-1} to these two functions. We do not know *a priori* that the quantities involved are finite; however, we will shortly show that the right hand side is finite, which implies that the left hand side is. In all,

$$|\langle (f_c)_{c\in\{0,1\}^k}\rangle_{U^k}| \leq$$
$$\left| \int \dots \int |s|^2\, d\mu(h_2)\dots d\mu(h_k)\right|^{1/2} \left| \int \dots \int |t|^2\, d\mu(h_2)\dots d\mu(h_k)\right|^{1/2}.$$

Consider the first term, which upon changing the dummy variable from x to h_1 in the second term and putting the conjugation operator inside the integral immediately expands to

$$\int \dots \int\int \prod_{\substack{c\in\{0,1\}^k \\ c_1=0}} C^{|c|} f_c(x + c_2h_2 + \ldots + c_kh_k)\, d\mu(x)$$
$$\int \prod_{\substack{c\in\{0,1\}^k \\ c_1=0}} C^{|c|} \overline{f_c(h_1 + c_2h_2 + \ldots + c_kh_k)}\, d\mu(h_1)\, d\mu(h_2)\dots d\mu(h_k)$$
$$= \int \dots \int \prod_{\substack{c\in\{0,1\}^k \\ c_1=0}} C^{|c|} f_c(x + c_2h_2 + \ldots + c_kh_k)\overline{f_c(h_1 + c_2h_2 + \ldots + c_kh_k)}\, d\mu(x)\, d\mu(h_1)\dots d\mu(h_k)$$

By making the change of variables $h_1 \mapsto x + h_1$ and letting $c' = (c_2, \ldots, c_k)$ denote the projection of c along the first coordinate, this can be rewritten as

$$\int \dots \int \prod_{c\in\{0,1\}^k} C^{|c|} f_{0,c'}(x + c_1h_1 + \ldots + c_kh_k)\, d\mu(x)\dots d\mu(h_k) = \langle (f_{0,c'})_{c\in\{0,1\}^k}\rangle_{U^k}$$

The same can be done to the second term in the above inequality, except with the projection to $f_{1,c'}$. In all, we have

$$|\langle (f_c)_{c\in\{0,1\}^k}\rangle_{U^k}| \leq |\langle (f_{0,c'})_{c\in\{0,1\}^k}\rangle_{U^k}|^{1/2}|\langle (f_{1,c'})_{c\in\{0,1\}^k}\rangle_{U^k}|^{1/2}.$$

We now go through precisely the same procedure for each other coordinate. For example, if we denote $(c_3,\ldots,c_k)$ by c'' in the same way as above we can show that

$$|\langle (f_{0,c'})_{c\in\{0,1\}^k}\rangle_{U^k}| \leq |\langle (f_{0,0,c''})_{c\in\{0,1\}^k}\rangle_{U^k}|^{1/2}|\langle (f_{0,1,c''})_{c\in\{0,1\}^k}\rangle_{U^k}|^{1/2}.$$

Repeating $2^k - 1$ times, we arrive at the inequality

$$|\langle (f_c)_{c\in\{0,1\}^k}\rangle_{U^k}| \leq \prod_{c\in\{0,1\}^k} ||f_c||_{U^k}.$$

Since all of the f_c were assumed in $G^k(A,\mu)$, the right hand side is finite, which justifies our earlier use of Cauchy-Schwarz. □

In much the same way that the triangle inequality for L^2 norms follows from the Cauchy-Schwarz inequality, the triangle inequality for Gowers norms can be deduced from the Gowers-Cauchy-Schwarz inequality.

Proposition 4.6. *The U^k are seminorms on $G^k(A,\mu)$.*

Proof. It is obvious that the U^k are positively homogeneous (that is, if a is a complex number, $||af||_{U^k} = |a|\cdot||f||_{U^k}$) and we have already shown that U^k maps to the nonnegative reals. The triangle inequality is all that remains to be seen.

By expanding $||f+g||_{U^k}^{2^k}$, we get a sum of 2^{2^k} different terms, each of which can be written as a U^k product of fs and gs. Each possible configuration of U^k products of fs and gs is represented exactly once in this sum. Letting j_c, $c \in \{0,1\}^k$ represent one such configuration (so either $j_c = f$ or $j_c = g$ for every $c \in \{0,1\}^k$), by the Gowers-Cauchy-Schwarz inequality

$$\langle (j_c)_{c\in\{0,1\}^k}\rangle_{U^k} \leq \prod_{c\in\{0,1\}^k} ||j_c||_{U^k}.$$

Group together all terms with d copies of f and $2^k - d$ copies of g. It is clear that there are precisely 2^k choose d such terms. All in all,

$$||f+g||_{U^k}^{2^k} \leq \sum_{d=0}^{2^k} \binom{2^k}{d} ||f||_{U^k}^{d}||g||_{U^k}^{2^k-d} = (||f||_{U^k} + ||g||_{U^k})^{2^k},$$

so upon taking 2^kth roots we arrive at the triangle inequality. □

In the finite case (that is, when $\mu(A) < \infty$), the U^k are bounded in terms of the U^{k+1}:

Proposition 4.7. *If $\mu(A) < \infty$, for every $k \geq 1$ and $f \in G^{k+1}(A,\mu)$,*

$$||f||_{U^k} \leq C||f||_{U^{k+1}}$$

for some constant C independent of f.

Proof. For $c \in \{0,1\}^{k+1}$, let $f_c = f$ when $c_{k+1} = 0$ and $f_c = 1$ otherwise. By the Gowers-Cauchy-Schwarz inequality for $k+1$, we have

$$|\langle (f_c)_{c\in\{0,1\}^{k+1}}\rangle_{U^{k+1}}| \leq \prod_{c\in\{0,1\}^{k+1}} ||f_c||_{U^{k+1}}$$

The left hand side is equal to

$$\left| \int \cdots \int \prod_{c \in \{0,1\}^k} C^{|c|} f(x + c_1 h_1 + \ldots + c_k h_k) \, d\mu(x) \ldots d\mu(h_k) \int d\mu(h_{k+1}) \right|$$
$$= ||f||_{U^k}^{2^k} \cdot \mu(A),$$

while the right hand side is equal to $||f||_{U^{k+1}}^{2^k} ||1||_{U^{k+1}}^{2^k}$. It is a simple calculation that

$$||1||_{U^{k+1}}^{2^k} = \left(\int \cdots \int d\mu(x) \, d\mu(h_1) \ldots d\mu(h_{k+1}) \right)^{1/2} = \mu(A)^{k/2+1}.$$

Plugging back into the Gowers-Cauchy-Schwarz inequality, we get

$$||f||_{U^k}^{2^k} \cdot \mu(A) = ||f||_{U^{k+1}}^{2^k} \cdot \mu(A)^{k/2+1},$$

which reduces to

$$||f||_{U^k} \leq \mu(A)^{k/2^{k+1}} ||f||_{U^{k+1}},$$

thus proving the theorem. □

The form of C immediately gives the following:

Corollary 4.8. *If μ is a probability measure, then the Gowers norms are nondecreasing with increasing k.* □

As another easy corollary, we have the following important fact, which implies that in the finite case the kernels of the Gowers seminorms are decreasing as k increases:

Corollary 4.9. *If $\mu(A) < \infty$ and $k \geq 1$, if $||f||_{U^{k+1}} = 0$, then $||f||_{U^k} = 0$.* □

It is also clear that, if $\mu(A) < \infty$,

$$L^1(A, \mu) = G^1(A, \mu) \supset G^2(A, \mu) \supset G^3(A, \mu) \supset \ldots.$$

The form of the above proposition suggests that no similarly useful inequality should hold in the infinite case. In fact, it is easy to construct an explicit counterexample, taking advantage of the extra (multiplicative) structure of the real numbers and the fact that the Gowers norms do not scale nicely with this structure. Let $A = \mathbb{R}$ and let $f_n : \mathbb{R} \to \mathbb{R}$ for $n > 0$ be given by

$$f_n(x) = e^{-nx^2},$$

which is in $G^k(\mathbb{R}, \mu)$ for all k. Then a tedious but elementary calculation shows that

$$||f_n||_{U^k} = 2^{\frac{2^{k+1}}{k(k-1)}} \left(\frac{n}{\pi} \right)^{\frac{2^{k+1}}{k+1}}.$$

Therefore for any C, by letting

$$c_k = \frac{2^{k+1}}{k+1} - \frac{2^{k+2}}{k+2}, \quad d_k = \frac{2^{k+2}}{(k+1)k} - \frac{2^{k+1}}{k(k-1)},$$

and choosing

$$n > \left(C \pi^{c_k} 2^{d_k} \right)^{1/c_k},$$

simple algebra shows that

$$||f_n||_{U^k} > C ||f_n||_{U^{k+1}}.$$

Therefore there is no constant entirely independent of f such that Proposition 4.7 holds when $A = \mathbb{R}$.

The only situation in which such a bound can be recovered in the infinite case is if f has compact support on a set S and there exists a measurable set B such that the Minkowski sum/difference $B + S - S$ is contained in a set C of finite measure. That is, there is some measurable B such that

$$B + S - S = \{x : x = b + s_1 - s_2; b \in B, s_1, s_2 \in S\}$$

is contained in some C such that $\mu(C) < \infty$.[3] This condition is not automatic: for example, consider the measure space on $\mathbb{R}$ consisting of all Lebesgue measurable sets, where the measure of a set is either 0 if that set is Lebesgue-null or ∞ otherwise. Then if C is the Cantor set, which has zero measure, its sumset $2C$ contains an interval and therefore has infinite measure.

Proposition 4.10. *Let $f \in G^{k+1}(A, \mu)$ be supported on a set S, and let there exist a set B such that $B + S - S$ is contained in a set of finite measure. Then*

$$||f||_{U^k} \leq C||f||_{U^{k+1}}$$

for some constant C depending on S, B, and k.

Proof. Similarly to the above proof, for $c \in \{0,1\}^{k+1}$, let $f_c = f$ when $c_{k+1} = 0$. Denote the set of finite measure containing $B + S - S$ by C. Letting $c' = (0, 0, \ldots, 0, 1)$, define $f_{c'} = \chi_B$ and $f_c = \chi_C$ when $c_{k+1} = 1$ and $c \neq c'$.

By Gowers-Cauchy-Schwarz,

$$|\langle (f_c)_{c \in \{0,1\}^{k+1}} \rangle_{U^{k+1}}| \leq \prod_{c \in \{0,1\}^{k+1}} ||f_c||_{U^{k+1}}.$$

The left-hand side is equal to the absolute value of

$$\int \cdots \int \prod_{c \in \{0,1\}^k} C^{|c|} f(x + c_1 h_1 + \ldots + c_k h_k)$$
$$\left(\int \chi_B(h_{k+1}) \prod_{\substack{c \in \{0,1\}^k \\ c \neq c'}} \chi_C(c_1 h_1 + \ldots + c_k h_k + h_{k+1}) \, d\mu(h_{k+1}) \right) d\mu(x) \ldots d\mu(h_k).$$

where in the inner integral we have made the change of variables $h_{k+1} \mapsto h_{k+1} - x$.

Because f is supported on S, we can narrow down the domain of integration considerably due to the first product; i.e., for a particular set of coordinates $(x, h_1, \ldots, h_k)$ to contribute to the integration, all sums $x + c_1 h_1 + \ldots + c_k h_k$ must be in S. In particular, since x must be in S, subtracting off x, all sums of the form $c_1 h_1 + \ldots + c_k h_k$ must be in $S - S$. Looking specifically at the inner integral, we must have $h_{k+1} \in B$ to contribute to the integral. Therefore every sum $c_1 h_1 + \ldots + c_k h_k + h_{k+1}$ is in $B + S - S$, so the inner integral is greater than or equal to

$$\int \chi_B(h_{k+1}) \, d\mu(h_{k+1}) = \mu(B),$$

[3]This condition is slightly more permissible than the more obvious condition that $\mu(B + S - S) < \infty$, in recognition of the fact that if S is measurable, $S - S$ may not be. See [3].

and the left hand side as a whole is greater than or equal to

$$||f||_{U^k}^{2^k} \cdot \mu(B).$$

Now consider the right hand side, which is equal to $||f||_{U^{k+1}}^{2^k} ||\chi_B||_{U^{k+1}} ||\chi_C||_{U^{k+1}}^{2^k-1}$. For any characteristic function on a set, we can very roughly estimate the Gowers norm as follows. We can eliminate characteristic functions in the integrand at the cost of possibly increasing the integral. Therefore

$$||\chi_B||_{U^{k+1}}^{2^{k+1}} \leq \int \dots \int \chi_B(x)\chi_B(x+h_1)\dots\chi_B(x+h_{k+1})\, d\mu(x)\dots d\mu(h_{k+1}).$$

By pulling in each $d\mu(h_i)$, making the change of variables $h_i \mapsto h_i - x$ in each term, and evaluating, we get

$$||\chi_B||_{U^{k+1}}^{2^{k+1}} \leq \mu(B)^{k+2} \implies ||\chi_B||_{U^{k+1}} \leq \mu(B)^{(k+2)/2^{k+1}}.$$

Likewise

$$||\chi_C||_{U^{k+1}}^{2^{k+1}} \leq \mu(C)^{k+2} \implies ||\chi_C||_{U^{k+1}}^{2^k-1} \leq \mu(C)^{(k+2)(2^k-1)/2^{k+1}}.$$

Putting this all together with the original inequality,

$$||f||_{U^k}^{2^k} \cdot \mu(B) \leq ||f||_{U^{k+1}}^{2^k} \cdot \mu(B)^{(k+2)/2^{k+1}} \cdot \mu(C)^{(k+2)(2^k-1)/2^{k+1}},$$

so (modulo some rearrangement) we are done. □

Another important fact about Gowers norms is the following, which characterizes U_2 in terms of Fourier analysis in the compact case:

Proposition 4.11. *If A is compact, μ normalized Haar measure on A, and $f \in G^2(A, \mu)$, then*

$$||f||_{U^2} = \left(\int_{\hat{A}} |\hat{f}(\xi)|^4 \, d\hat{\mu}(\xi) \right)^{1/4},$$

where $\hat{A}$ is the Pontryagin dual of A and $\hat{\mu}$ is the dual measure of μ (chosen so that the Fourier inversion formula holds).

Proof. If A is compact and μ normalized Haar measure, its Pontryagin dual $\hat{A}$ is discrete and $\hat{\mu}$ is counting measure, so the above integral becomes a sum, and we have the Fourier expansion

$$f(x) = \sum_{\alpha \in \hat{A}} \hat{f}(\alpha)\chi_\alpha(x).$$

The verification is then straightforward, if irritating: we have

$$\begin{aligned}
||f||_{U^2}^4 &= \int\int\int f(x)f(x+y)f(x+z)f(x+y+z)\, d\mu(x)\, d\mu(y)\, d\mu(z) \\
&= \int\int\int f(x)f(x+y) \sum_{\alpha,\beta \in \hat{A}} \hat{f}(\alpha)\hat{f}(\beta)\chi_\beta(y)\chi_{\alpha-\beta}(x+z)\, d\mu(x)\, d\mu(y)\, d\mu(z) \\
&= \int\int f(x)f(x+y) \sum_{\alpha,\beta \in \hat{A}} \hat{f}(\alpha)\hat{f}(\beta)\chi_\beta(y) \left(\int \chi_{\alpha-\beta}(x+z)\, d\mu(z) \right) d\mu(x)\, d\mu(y).
\end{aligned}$$

The quantity in parenthesis is zero unless $\alpha = \beta$, in which case it is equal to one, so this simplifies to

$$\begin{aligned}||f||_{U^2}^4 &= \int\int f(x)f(x+y) \sum_{\alpha \in \hat{A}} \hat{f}(\alpha)^2 \chi_\alpha(y)\, d\mu(x)\, d\mu(y) \\ &= \sum_{\alpha \in \hat{A}} \hat{f}(\alpha)^2 \int\int f(x)f(x+y)\chi_\alpha(y)\, d\mu(x)\, d\mu(y) \\ &= \sum_{\alpha \in \hat{A}} \hat{f}(\alpha)^2 \int\int \hat{f}(\beta)\hat{f}(\gamma)\chi_{\beta-\gamma}(x)\chi_{\alpha-\gamma}(y)\, d\mu(x)\, d\mu(y) \\ &= \sum_{\alpha \in \hat{A}} \hat{f}(\alpha)^2 \hat{f}(\beta)\hat{f}(\gamma) \int \chi_{\beta-\gamma}(x)\, d\mu(x) \int \chi_{\alpha-\gamma}(y)\, d\mu(y).\end{aligned}$$

Again, the integrals cancel unless $\alpha = \beta = \gamma$, so we are left with

$$||f||_{U^2}^4 = \sum_{\alpha \in \hat{A}} \hat{f}(\alpha)^4,$$

as desired. □

This lemma has the following corollary:

Corollary 4.12. *If A is compact and μ normalized Haar measure on A, then the U^k seminorms on (A, μ) are norms.*

Proof. Let f be in the kernel of U^k. By Corollary 4.8, f must also be in the kernel of U^2. By the above proposition, this implies $\hat{f} = 0$ a.e., which implies that $f = 0$ a.e., so we are done. □

The failure of the U^k to be norms in the ultraproduct case, which we will see below, therefore reflects the fact that none of the measure spaces constructed in this manner are compact topological groups.

4.3. **Octahedral (semi-)norms and quasirandomness.** In order to establish the basic equivalence result between the kernels of the Gowers seminorms and certain σ-algebras, we need to make a detour into a multidimensional version. For a group A and measure μ as above, let $f : A^k \to \mathbb{C}$ denote a measurable function with respect to the product measure on A^k. Then define the *octahedral norm* of f to be

$$||f||_{O^k}^{2^k} = \int\int \cdots \int \prod_{c \in \{0,1\}^k} C^{|c|}[f(x_{1,c_1}, x_{2,c_2}, \ldots, x_{k,c_k})]\, d\mu(x_{1,0})\, d\mu(x_{1,1}) \ldots d\mu(x_{k,1}).$$

It is easy to check that if $g : A \to \mathbb{C}$ is a measurable function and

$$g_k(x_1, \ldots, x_k) = g\left(\sum_{j=1}^k x_j\right),$$

then we have

$$||g||_{U^k} = ||g_k||_{O^k}.$$

The importance of the octahedral norms comes from their connection to the σ-algebra $\mathcal{A}_{[k]^*}$ on $\mathbf{A}^k$ defined above.

First recall that a subspace H_1 of a Hilbert space H_2 is itself a Hilbert space if and only if it is closed. We can uniquely define a projection operator onto closed subspaces; we will denote the projection onto H_1 by P_{H_1}. If $f \in H_1$ and $g \in H_2$, the projection operator obeys

$$\langle f, g\rangle = \langle f, P_{H_1}(g)\rangle.$$

Next, recall the definition of the conditional expectation. Given a probability space $(X, \mathcal{A}, \mu)$, a sub-σ-algebra $\mathcal{B}$, and an $\mathcal{A}$-measurable function $f : X \to \mathbb{C}$, the *conditional expectation* of f given $\mathcal{B}$ is a $\mathcal{B}$-measurable function $E(f|\mathcal{B}) : X \to \mathbb{C}$ such that

$$\int_B E(f|\mathcal{B})\, d\mu = \int_B f\, d\mu$$

for each $B \in \mathcal{B}$. Although this is not a constructive definition, it is a classical result (which crucially employs the Radon-Nikodym theorem) that if f is integrable, $E(f|\mathcal{B})$ exists and is unique up to measure-zero change. As it turns out, conditional expectation coincides with Hilbert space projection:

Proposition 4.13. *Let $(X, \mathcal{A}, \mu)$ be a probability space and $\mathcal{B} \subseteq \mathcal{A}$ a sub-σ-algebra. Let $H_1 = L^2(X, \mathcal{A})$ and $H_2 = L^2(X, \mathcal{B})$; obviously both are Hilbert spaces and $H_1 \subseteq H_2$. If $f \in H_1$, then*

$$P_{H_2}(f) = E(f|\mathcal{B})$$

almost everywhere.

Proof. Pick some $B \in \mathcal{B}$ and let χ_B be the indicator function of B. For any $f \in H_1$, by the projection property we have

$$\langle f, \chi_B\rangle = \langle P_{H_2}(f), \chi_B\rangle,$$

so

$$\int_B f\, d\mu = \int_X f\overline{\chi_B}\, d\mu = \langle f, \chi_B\rangle = \langle P_{H_2}(f), \chi_B\rangle = \int_B P_{H_2}(f)\, d\mu.$$

Therefore by the definition of the conditional expectation,

$$\int_B E(f|\mathcal{B})\, d\mu = \int_B P_{H_2}(f)\, d\mu,$$

so since both functions are $\mathcal{B}$-measurable, we have

$$P_{H_2}(f) = E(f|\mathcal{B})$$

almost everywhere. □

Now consider a function $f \in L^2(\mathbf{A}^k, \mathcal{A}(\mathbf{A}^k))$. We say that f is *quasirandom* if

$$E(f|\mathcal{A}_{[k]^*}) = 0.$$

Keeping in mind that conditional probability coincides with Hilbert space projection for L^2 functions, this is equivalent to the statement that if f is orthogonal to every $g \in L^2(\mathbf{A}^k, \mathcal{A}_{[k]^*})$.

Lemma 4.14. *The set of quasirandom functions on $\mathbf{A}^k$ forms a Hilbert space.*

Proof. That they form a vector space is obvious. By basic Hilbert space theory, orthogonal complements are closed. □

The main connection is the following surprising fact:

Proposition 4.15. *If $f \in L^2(\boldsymbol{A}^k, \mathcal{A}(\boldsymbol{A}^k))$, then f is quasirandom if and only if $||f||_{O^k} = 0$.*

Proof (sketch). This proposition is proved using a straightforward induction on k with a formula similar to that for the Gowers norm in Lemma 4.4; details can be found (in a slightly different setting) in [17]. However, in the interest of explanation, the following sketch explains why the statement is true.

We have to show that the space of functions with zero octahedral norm is orthogonal to the space of quasirandom functions. It can be shown that the former space is a Hilbert space as well, so the question reduces to a statement about the orthogonality of two Hilbert spaces, the space of functions with zero octahedral norm and $L^2(\mathbf{A}^k, \mathcal{A}_{[k]^*})$. By a closure argument, it suffices to prove for a dense subset of both spaces, so by linearity we just have to consider characteristic functions of measurable sets. Furthermore, by the construction of $\mathcal{A}_{[k]^*}$, it suffices to let the function in $L^2(\mathbf{A}^k, \mathcal{A}_{[k]^*})$ be $\chi_{\mathbf{E}\times\mathbf{F}}$, where $\mathbf{E} \in \mathcal{A}(\mathbf{A}^{k-1})$ and $\mathbf{F} \in \mathcal{A}(\mathbf{A})$. Let f be given by χ_G, where $G \in \mathcal{A}(\mathbf{A}^k)$.

The statement of the proposition then boils down to

$$0 = \int_{\mathbf{A}^k} \chi_{\mathbf{E}\times\mathbf{F}}\chi_{\mathbf{G}} \, d\mu = \mu((\mathbf{E}\times\mathbf{F}) \cap \mathbf{G})$$

if and only if

$$||\chi_G||_{O^k} = 0.$$

The integrand of the latter equation is nonzero (and equal to one) if and only if all points of the rectangular prism described by the $2k$ coordinates $x_{1,0}, \ldots, x_{k,1}$ are in $\mathbf{G}$. The octahedral norm therefore measures the total measure of rectangular prisms with all vertices in $\mathbf{G}$. The first statement, on the other hand, is approximately a statement that $\mathbf{G}$ has no positive-measure intersection with any cylinder set. It should be clear that these statements are "morally" similar (they both state in some sense that $\mathbf{G}$ is nowhere dense) and it is likely that an alternate proof of the proposition could be developed along these lines. □

4.4. **A few Hilbert space results.** Associated to each measure space $(X, \mathcal{B}, \mu)$ is the corresponding Hilbert space $L^2(X, \mathcal{B}, \mu)$ of square-integrable $\mathcal{B}$-measurable functions $X \to \mathbb{C}$ with inner product given by

$$\langle f, g\rangle = \int_X f\overline{g} \, d\mu.$$

Consideration of this space requires that we consider complex-valued functions rather than real-valued functions, but this is not a substantial increase in generality; the results of Sections 3.2 and 3.3 still apply where relevant by taking real and imaginary parts separately. In particular, the standard part of a hypercomplex number is defined by taking the standard parts of the real and imaginary parts. Because the ultraproduct σ-algebras we are considering are too big - specifically, they are nonseparable - the corresponding Hilbert spaces will be nonseparable as well.

First, we have a lemma that will be useful several times throughout the remainder of this paper:

Lemma 4.16. *Let $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ be ultraproduct spaces with measures μ_1 and μ_2 constructed as above, and let $H \subseteq L^2(\boldsymbol{A}_1, \mathcal{A}(\boldsymbol{A}_1), \mu_1)$ be a Hilbert space. Let $f : \boldsymbol{A}_1 \times \boldsymbol{A}_2 \to \mathbb{C}$ be an $\mathcal{A}(\boldsymbol{A}_1 \times \boldsymbol{A}_2)$-measurable function such that the function $g_y : \boldsymbol{A}_1 \to \mathbb{C}$ defined by*

$g_y(x) = f(x,y)$ is in H for every $y \in \boldsymbol{A}_2$. Then $g : \boldsymbol{A}_1 \to \mathbb{C}$ defined by

$$g(x) = \int_{\boldsymbol{A}_2} f(x,y)\, d\mu_2(y)$$

is in H.

Proof. By Theorem 4.3, we can freely interchange integrals. Using the above fact about projection operators twice, we get

$$\begin{aligned}
||g||_2^2 &= \int_{\mathbf{A}_1}\int_{\mathbf{A}_2} f(x,y)\overline{g(x)}\, d\mu_1(y)\, d\mu_2(x)\\
&= \int_{\mathbf{A}_2}\int_{\mathbf{A}_1} f(x,y)\overline{g(x)}\, d\mu_2(x)\, d\mu_1(y)\\
&= \int_{\mathbf{A}_2}\int_{\mathbf{A}_1} f(x,y)\overline{P_H(g(x))}\, d\mu_2(x)\, d\mu_1(y)\\
&= \int_{\mathbf{A}_1}\int_{\mathbf{A}_2} f(x,y)\overline{P_H(g(x))}\, d\mu_1(y)\, d\mu_2(x)\\
&= \int_{\mathbf{A}_1} g(x)\overline{P_H(g(x))}\, d\mu_1(x)\\
&= \int_{\mathbf{A}_1} P_H(g(x))\overline{P_H(g(x))}\, d\mu_1(x)\\
&= ||P_H(g)||_2^2.
\end{aligned}$$

A Hilbert space projection is a always a norm-decreasing function except when it acts as the identity. Therefore $||g||_2^2 = ||P_H(g)||_2^2$ implies that $g = P_H(g)$, which implies that $g \in H$. □

Let $\mathbf{A}$ be an ultraproduct space with measure μ as constructed above, and let $\mathcal{B} \subseteq \mathcal{A}(\mathbf{A})$ be a sub-σ-algebra. Consider $L^\infty(\mathcal{B})$ as a ring; that is, the ring of bounded $\mathcal{B}$-measurable complex-valued functions, equipped with pointwise multiplication. A Hilbert space $H \subseteq L^2(\mathbf{A})$ is an $L^\infty(\mathcal{B})$-module if it is a module over this ring via pointwise multiplication; that is, for each $h \in H$ and $f \in L^\infty(\mathcal{B})$ we should have $hf \in H$.

The following lemma shows that the module operation commutes with projections.

Lemma 4.17. *Let H_1 and H_2 be two $L^\infty(\mathcal{B})$-modules. Then for every $v \in H_1$ and $f \in L^\infty(\mathcal{B})$, we have*

$$P_{H_2}(vf) = fP_{H_2}(v).$$

Proof. Because the module operation is just pointwise multiplication, we can move elements of $L^\infty(\mathcal{B})$ freely from one side of an inner product to the other, provided we take the complex conjugate. Freely using the aforementioned property of projection operators, for every $w \in H_2$ we have

$$\langle P_{H_2}(vf), w\rangle = \langle vf, w\rangle = \langle v, w\overline{f}\rangle = \langle P_{H_2}(v), w\overline{f}\rangle = \langle fP_{H_2}(v), w\rangle.$$

Then set $w = P_{H_2}(vf) - fP_{H_2}(v) \in H_2$, so the above equality yields $\langle w, w\rangle = 0$. Therefore $w = 0$, which is the desired result. □

The *rank* of H, denoted $\mathrm{rk}(H)$, is then defined to be the minimal cardinality of a subset S of H such that the $L^\infty(\mathcal{B})$-module generated by S is dense in H.

Lemma 4.18. *Let $H_1 \subseteq H_2$ be two $L^\infty(\mathcal{B})$-modules. Then* $\mathrm{rk}(H_1) \leq \mathrm{rk}(H_2)$.

Proof. Let S be a subset of H_2 such that the $L^\infty(\mathcal{B})$-module generated by S is dense in H_2. Let

$$S' = \{P_{H_1}(s) : s \in S\}.$$

We claim that the $L^\infty(\mathcal{B})$-module generated by S' is dense in H_1.

Consider some $h \in H_1$ and some $\epsilon > 0$. By assumption, there are elements $s_1, \dots, s_r \in S$ and $f_1, \dots, f_r \in L^\infty(\mathcal{B})$ such that

$$\left\| h - \sum_{i=1}^{r} f_i s_i \right\|_2 \leq \epsilon.$$

The projection P_{H_1} cannot increase distances, so applying P_{H_1} we certainly have

$$\left\| h - \sum_{i=1}^{r} P_{H_1}(f_i s_i) \right\|_2 \leq \epsilon.$$

By Lemma 4.17 above, this implies that

$$\left\| h - \sum_{i=1}^{r} f_i P_{H_1}(s_i) \right\|_2 \leq \epsilon,$$

so the $L^\infty(\mathcal{B})$-module generated by S' is dense in H_1. □

4.5. **Weak orthogonality and coset σ-algebras.** Let H_1, H_2 be two Hilbert spaces both contained in a larger Hilbert space H. We say that H_1 and H_2 are *weakly orthogonal* if for any $h \in H_1$, $P_{H_2}(h) \in H_1$.

Lemma 4.19. *Weak orthogonality for Hilbert spaces is symmetric; that is, if H_1 is weakly orthogonal to H_2, then H_2 is weakly orthogonal to H_1.*

Proof. It is easy to see that that weak orthogonality means precisely that the orthogonal complements of $H_1 \cap H_2$ in H_1 and H_2 are orthogonal, which is clearly a symmetric relation. □

There is an analogous notion for σ-algebras. Given two σ-algebras $\mathcal{B}_1$ and $\mathcal{B}_2$ both contained in a larger σ-algebra, we say that $\mathcal{B}_1$ and $\mathcal{B}_2$ are *weakly orthogonal* if for any $f \in L^2(X, \mathcal{B}_1)$, the conditional expectation $E(f|\mathcal{B}_2)$ is measurable in $\mathcal{B}_1$. By the connection between Hilbert space projections and conditional expectations, we have the following:

Lemma 4.20. *Weak orthogonality for σ-algebras is symmetric.*

Proof. Weak orthogonality for σ-algebras when restricted to L^2 functions is, by Proposition 4.13, equivalent to weak orthogonality on the corresponding Hilbert spaces. □

Back in the setting of the ultraproduct group $\mathbf{A}$, consider an internal subgroup $\mathbf{H}$; that is, a subgroup of the form

$$\mathbf{H} = \lim_{i \to \omega} H_i,$$

where each H_i is a subgroup of A_i, or equivalently that ω-many of the H_i are subgroups of the corresponding A_i. Note that not every subgroup of $\mathbf{A}$ is of this form; for example, $\mathbb{Z} \subset {}^*\mathbb{R}$ is a subgroup but as already discussed not an internal set. If $\mathbf{H}$ is an internal set that is a subgroup, however, the following lemma shows us that it is an internal subgroup as defined above:

Lemma 4.21. *If* $\boldsymbol{H} = \lim_{i\to\omega} H_i$ *is a subgroup of* $\boldsymbol{A}$*, then*

$$\{i \in \mathbb{N} : H_i \text{ is a subgroup of } A_i\} \in \omega.$$

Proof. Łoś' theorem. □

The *coset* σ*-algebra* $\mathcal{A}(\mathbf{A}, \mathbf{H}) \subseteq \mathcal{A}(\mathbf{A})$ is defined to be the subalgebra of $\mathcal{A}$ consisting of unions of cosets of $\mathbf{H}$. For example, if $\mathbf{H}$ is the trivial subgroup then $\mathcal{A}(\mathbf{A}, \mathbf{H}) = \mathcal{A}(\mathbf{A})$, while if $\mathbf{H} = \mathbf{A}$ then $\mathcal{A}(\mathbf{A}, \mathbf{H}) = \{\emptyset, \mathbf{A}\}$ is the trivial σ-algebra.

The importance of coset σ-algebras is the following proposition. An *invariant* σ-algebra $\mathcal{B}$ is one such that if $\mathbf{E} \in \mathcal{B}$ and $a \in \mathbf{A}$ then the translation $\mathbf{E} + a \in \mathcal{B}$.

Proposition 4.22. *A coset* σ*-algebra on* $\boldsymbol{A}$ *is weakly orthogonal to any invariant* σ*-algebra.*

Proof. Let $\mathbf{H}$ be the coset in question and let $\mathcal{B}$ be an invariant σ-algebra. We can define an averaging operator that we will show coincides with the projection operator into $L^2(\mathcal{A}(\mathbf{A}, \mathbf{H}))$. Let $f \in L^2(\mathcal{B})$. Define the operator T on $L^2(\mathcal{B})$ by

$$T(f)(x) = \int_{\mathbf{H}} f(x + h)\, d\mu_{\mathbf{H}},$$

where $d\mu_{\mathbf{H}}$ is the probability measure associated with $\mathbf{H}$. Since $T(f)$ is obviously constant on each coset of $\mathbf{H}$, it is in $L^2(\mathcal{A}(\mathbf{A}, \mathbf{H}))$.

If $w \in L^2(\mathcal{A}(\mathbf{A}, \mathbf{H}))$ then w is constant on each coset of $\mathbf{H}$. Therefore with such w we have, making the change of variables $x \mapsto x - h$, noting that w is unchanged under such a transformation, and using the fact that $d\mu_{\mathbf{H}}$ is a probability measure, we get

$$\begin{aligned}
\langle T(f)(x), w\rangle &= \int_{\mathbf{A}} \left(\int_{\mathbf{H}} f(x + h)\, d\mu_{\mathbf{H}}(h) \right) \overline{w(x)}\, d\mu(x) \\
&= \int_{\mathbf{A}} \left(\int_{\mathbf{H}} f(x)\, d\mu_{\mathbf{H}}(h) \right) \overline{w(x)}\, d\mu(x) \\
&= \int_{\mathbf{A}} f(x)\overline{w(x)}\, d\mu(x) \\
&= \langle f, w\rangle.
\end{aligned}$$

Therefore $T(f)$ is (up to measure zero) the projection operator $E(f|\mathcal{A}(\mathbf{A}, \mathbf{H}))$, so it remains to be shown that $T(f) \in L^2(\mathcal{B})$. For this, use Lemma 4.16 with $\mathbf{A}_2 = \mathbf{H}$. For every $x \in \mathbf{H}$, $f(x+h)$ as a function of h is certainly in $L^2(\mathcal{B})$. Therefore by the result of the lemma, $T(f)$ is also in $L^2(\mathcal{B})$, which implies weak orthogonality. □

For what follows, we will need a proposition about weakly orthogonal σ-algebras, which requires a lemma and a theorem whose proofs we exclude because they do not have any direct bearing on the matter at hand. For a proof of the lemma, see [15], while for the theorem see [7] or the fuller version [8]. If we have two σ-algebras $\mathcal{B}_1$ and $\mathcal{B}_2$ on a single space, then let $\langle \mathcal{B}_1, \mathcal{B}_2\rangle$ denote the σ-algebra generated by both.

Lemma 4.23. *Let* $\mathcal{B}_1$ *and* $\mathcal{B}_2$ *be* σ*-algebras, and let* $W \in \langle \mathcal{B}_1, \mathcal{B}_2\rangle$*. Then for every* ϵ *there exists a set* W' *of the form*

$$W' = \bigcup_{j=1}^{n} (B_j^1 \cap B_j^2),$$

where $B_j^1 \in \mathcal{B}_1$ *and* $B_j^2 \in \mathcal{B}_2$*, such that*

$$\mu(W \triangle W') \leq \epsilon,$$

and in fact $\{B_j^2\}_{j=1}^n$ *is a partition of the space.* □

If $(X, \mathcal{A}, \mu)$ is a probability measure and $\mathcal{B}$ and $\mathcal{C}$ are sub-σ-algebras, we call $\mathcal{B}$ and $\mathcal{C}$ *independent* if

$$\mu(B \cap C) = \mu(B)\mu(C)$$

for all $B \in \mathcal{B}$ and $C \in \mathcal{C}$. We say that $\mathcal{C}$ is an *independent complement* of $\mathcal{B}$ if it is independent from $\mathcal{B}$ and $\langle \mathcal{B}, \mathcal{C} \rangle$ is dense in $\mathcal{A}$ in the following sense: if $A \in \mathcal{A}$ is any set and $\epsilon > 0$, then there is a set $D_\epsilon \in \langle \mathcal{B}, \mathcal{C} \rangle$ such that

$$\mu(A \triangle D_\epsilon) \leq \epsilon.$$

Theorem 4.24. *Let* $(\boldsymbol{A}, \mathcal{A}, \mu)$ *be an ultraproduct space and let* $\mathcal{B}_1$ *and* $\mathcal{B}_2$ *be a sub-*σ*-algebras. Then there is an independent complement* $\mathcal{C}$ *of* $\mathcal{B}_1$ *in* $\langle \mathcal{B}_1, \mathcal{B}_2 \rangle$*.* □

We say that a function f measurable in a σ-algebra $\mathcal{A}$ is *almost measurable* in a sub-σ-algebra $\mathcal{B}$ if it can be approximated arbitrarily well in an L^1 sense by functions measurable in $\mathcal{B}$; that is, for every $\epsilon > 0$ there is a f_ϵ measurable in $\mathcal{B}$ such that

$$||f - f_\epsilon||_1 \leq \epsilon.$$

Proposition 4.25. *Let* $\mathcal{B}_1, \mathcal{B}_2 \subseteq \mathcal{A}$ *be two weakly orthogonal* σ*-algebras on a probability space* $(X, \mathcal{A}, \mu)$*, and let* W *be a set measurable in* $\langle \mathcal{B}_1, \mathcal{B}_2 \rangle$*. Let* $\mathcal{B}_3$ *be the* σ*-algebra generated by all the functions*

$$E(\chi_W \cdot t | \mathcal{B}_1), \quad t \in L^\infty(\mathcal{B}_2).$$

Then χ_W *is almost measurable in* $\langle \mathcal{B}_2, \mathcal{B}_3 \rangle$*.*

Proof. First assume that $\mathcal{B}_1$ and $\mathcal{B}_2$ are independent. By Lemma 4.23, there exist sets $B_j^1 \in \mathcal{B}_1$ and $B_j^2 \in \mathcal{B}_2$, $1 \leq j \leq n$, such that $\{B_j^2\}$ is a partition of the space and if we let

$$W' = \bigcup_{j=1}^n (B_j^1 \cap B_j^2),$$

we have

$$\mu(W \triangle W') \leq \epsilon.$$

By construction,

$$\chi_{W'}(x)\chi_{B_j^2}(x) = \chi_{B_j^1 \cap B_j^2}(x).$$

The projection of this function onto $\mathcal{B}_1$ can be easily calculated: we claim that

$$E(\chi_{W'}\chi_{B_j^2} | \mathcal{B}_1) = \mu(B_j^2)\chi_{B_j^1}.$$

Checking the claim: for all $B_1 \in \mathcal{B}_1$, using the independence of $\mathcal{B}_1$ and $\mathcal{B}_2$, we have

$$\begin{aligned}
\int_{B^1} \chi_{B_j^1 \cap B_j^2}(x)\, d\mu(x) &= \mu(B_j^1 \cap B_j^2 \cap B^1) \\
&= \mu(B_j^1 \cap B_j^1)\mu(B_j^2) \\
&= \int_{B^1} \mu(B_j^2)\chi_{B_j^1}(x)\, d\mu(x).
\end{aligned}$$

By the almost everywhere uniqueness of conditional expectations, therefore, we have found the correct projection.

Using the partition $\{B_j^2\}$ and the above calculation, we have

$$\chi_{W'} = \sum_{j=1}^{n} \chi_{B_j^1}\chi_{B_j^2} = \sum_{j=1}^{n} \frac{\chi_{B_j^2}}{\mu(B_j^2)} E(\chi_{W'}\chi_{B_j^2}|\mathcal{B}_1).$$

Define the function g by

$$g = \sum_{j=1}^{n} \frac{\chi_{B_j^2}}{\mu(B_j^2)} E(\chi_W\chi_{B_j^2}|\mathcal{B}_1),$$

replacing the W' in the above expansion by W. Clearly, $g \in \langle \mathcal{B}_2, \mathcal{B}_3 \rangle$. Then, using the linearity of conditional probability and the independence of $\mathcal{B}_1$ and $\mathcal{B}_2$,

$$\begin{aligned}
||\chi_{W'} - g||_1 &= \left|\left| \sum_{j=1}^{n} \frac{\chi_{B_j^2}}{\mu(B_j^2)} E((\chi_{W'} - \chi_W)\chi_{B_j^2}|\mathcal{B}_1) \right|\right|_1 \\
&\leq \sum_{j=1}^{n} \frac{\mu(B_j^2)}{\mu(B_j^2)} \left|\left| E(|\chi_{W'} - \chi_W| \cdot \chi_{B_j^2}|\mathcal{B}_1) \right|\right|_1 \\
&= \sum_{j=1}^{n} \left|\left| |\chi_{W'} - \chi_W| \cdot \chi_{B_j^2} \right|\right|_1 \\
&\leq \epsilon.
\end{aligned}$$

We also clearly have that

$$||\chi_{W'} - \chi_W||_1 \leq \epsilon.$$

Combining the two, therefore, we get

$$||g - \chi_W||_1 \leq 2\epsilon,$$

so χ_W is almost measurable in $\langle \mathcal{B}_2, \mathcal{B}_3 \rangle$.

In the case when $\mathcal{B}_1$ and $\mathcal{B}_2$ are not independent, use Theorem 4.24 to find a σ-algebra $\mathcal{C}$ that is an independent complement of $\mathcal{B}_1$ in $\langle \mathcal{B}_1, \mathcal{B}_2 \rangle$ and such that W is measurable in $\langle \mathcal{B}_1, \mathcal{C} \rangle$. The previous argument then proves the proposition in this case. □

4.6. **Relative separability and essential σ-algebras.** Let $\mathcal{B}_1 \subseteq \mathcal{B}_2$ be two σ-algebras. Say that $\mathcal{B}_2$ is *relative separable* over $\mathcal{B}_1$ if we can choose countably many elements of $\mathcal{B}_2$ such that the σ-algebra generated by $\mathcal{B}_1$ and these elements is dense in $\mathcal{B}_2$. As usual for σ-algebras, we mean dense in the sense of symmetric difference; that is, one σ-algebra $\mathcal{C}$ is dense in another $\mathcal{D}$ if for every $D \in \mathcal{D}$ and $\epsilon > 0$, there is a $C \in \mathcal{C}$ such that $\mu(C \triangle D) \leq \epsilon$.

In the language of Hilbert space modules above, relative separability has a precise interpretation:

Lemma 4.26. *If $\mathcal{B}_1 \subseteq \mathcal{B}_2$ are σ-algebras, then $\mathcal{B}_2$ is relative separable over $\mathcal{B}_1$ if and only if the rank of $L^2(\mathcal{B}_2)$ over the ring $L^\infty(\mathcal{B}_1)$ is at most countable.*

Proof. First assume that $\mathcal{B}_2$ is relative separable over $\mathcal{B}_1$, where the countably many extra elements are denoted $\mathbf{F}_i$, $i \in \mathbb{N}$. Then the $L^\infty(\mathcal{B}_1)$-module generated by the characteristic functions $\chi_{\mathbf{F}_i}$ is easily seen to be dense in $L^2(\mathcal{B}_2)$.

In the other direction, if the rank of $L^2(\mathcal{B}_2)$ over $L^\infty(\mathcal{B}_1)$ is countable, then there are countably many $f_i \in L^2(\mathcal{B}_2)$ such that the $L^\infty(\mathcal{B}_1)$-module generated by the f_i is dense in $L^2(\mathcal{B}_2)$. Each of the f_i can be approximated arbitrarily well by a $\mathcal{B}_2$-measurable stepfunction, so if we take all of the sets in $\mathcal{B}_2$ used in these stepfunctions (which is still a countable set), we generate a σ-algebra that is dense in $\mathcal{B}_2$. □

For Proposition 4.28 below, we want a notion of relative separability defined even when $\mathcal{B}_1$ is not contained in $\mathcal{B}_2$. In this case, say that $\mathcal{B}_2$ is relative separable over $\mathcal{B}_1$ if $\mathcal{B}_1$ together with at most countably extra elements together generate a σ-algebra that is dense in $\langle \mathcal{B}_1, \mathcal{B}_2 \rangle$.

Likewise, with the same notation, define a *relative atom* of $\mathcal{B}_2$ over $\mathcal{B}_1$ to be an element $\mathbf{E}$ of $\mathcal{B}_2$ with $\mu(\mathbf{E}) > 0$ such that for every subset $\mathbf{F} \subseteq \mathbf{E}$ there is a $\mathbf{G} \in \mathcal{B}_1$ such that

$$\mu((\mathbf{E} \cap \mathbf{G}) \triangle \mathbf{F}) = 0.$$

If $\mathcal{B}_1$ is the trivial σ-algebra, then this reduces to the statement that the measure space with σ-algebra $\mathcal{B}_2$ is atomless.

These notions are relevant because relative separable, relative atomless extensions have a particularly nice structure. In particular, we have the following measure-theoretic fact, whose proof would take us too far astray:

Lemma 4.27. *Let $\mathcal{B}_1 \subseteq \mathcal{B}_2$ be two σ-algebras on a space X such that $\mathcal{B}_2$ is relative separable and relative atomless over $\mathcal{B}_1$. Then there is a system of functions $g_i : X \to \mathbb{C}$ such that*

$$E(g_i \overline{g_j} | \mathcal{B}_1) = \delta_{i,j}$$

for all i, j in a countable set. □

We wish to study the slices of subsets of $\mathbf{A}^k$ in the last coordinate. Let $\mathbf{E} \subset \mathbf{A}^k$ be an arbitrary subset, and for $x \in \mathbf{A}$ define

$$\mathbf{E}_x = \{(x_1, \ldots, x_{k-1}) | (x_1, \ldots, x_{k-1}, x) \in \mathbf{E}\}$$

to be the x-slice of $\mathbf{E}$. Let $\mathcal{S}(\mathbf{E})$ be the σ-algebra on $\mathbf{A}^{k-1}$ generated by all x-slices.

The following proposition connects this notion back with the product σ-algebras defined earlier. Recall that $\mathcal{A}_{[k]^*}$ is a σ-algebra on $\mathbf{A}^k$ generated by the inverse images of the natural σ-algebras $\mathcal{A}(\mathbf{A}^{k-1})$ of all projections onto one fewer dimension. Likewise, $\mathcal{A}_{[k-1]^*}$ is similarly defined with respect to $\mathbf{A}^{k-1}$.

Proposition 4.28. *The set $\boldsymbol{E}$ is in $\mathcal{A}_{[k]^*}(\boldsymbol{A}^k)$ if and only if $\mathcal{S}(\boldsymbol{E})$ is relative separable over $\mathcal{A}_{[k-1]^*}(\boldsymbol{A}^{k-1})$.*

Proof. In one direction, assume that $\mathbf{E} \in \mathcal{A}_{[k]^*}$. Any particular element of a σ-algebra is certainly generated by countably many elements of that σ-algebra (by transfinite induction, if you like), so by the definition of $\mathcal{A}_{[k]^*}$ there are countably many sets $\mathbf{E}_{i,j} \subseteq \mathbf{A}^k$ such that $\mathbf{E}_{i,j} \in \mathcal{A}_{[k] \setminus \{j\}}$, with $i \in \mathbb{N}$ and $1 \leq j \leq k$, such that $\mathbf{E}$ is in the σ-algebra generated by the $\mathbf{E}_{i,j}$. By definition, if $j < k$, then

$$\mathcal{S}(\mathbf{E}_{i,j}) \in \mathcal{A}_{[k-1]^*},$$

so we don't have to worry about those elements. If $j = k$, then $\mathbf{E}_{i,k} \in \mathcal{A}_{[k-1]}$, so $\mathbf{E}_{i,k}$ is constant in the k coordinate, so all slices in that coordinate are the same, so $\mathcal{S}(\mathbf{E}_{i,k})$ is generated by that one element. Therefore

$$\mathcal{S}(\mathbf{E}) \subseteq \langle \mathcal{S}(\mathbf{E}_{i,j}) | i \in \mathbb{N}, 1 \leq j \leq k \rangle$$

is generated by $\mathcal{A}_{[k-1]^*}$ together with at most countably many extra elements, so it is relative separable over $\mathcal{A}_{[k-1]^*}$.

In the other direction, let $\mathcal{S}(\mathbf{E})$ be relative separable over $\mathcal{A}_{[k-1]^*}$. We want to use Lemma 4.27, but $\mathcal{S}(\mathbf{E})$ may not be relative atomless over $\mathcal{A}_{[k-1]^*}$. To remedy this, construct a σ-algebra $\mathcal{B}$ on $\mathbf{A}^k$ such that

$$\mathcal{A}_{[k-1]^*} \subseteq \mathcal{S}(\mathbf{E}) \subseteq \mathcal{B} \subseteq \mathcal{A}(\mathbf{A}^{k-1})$$

and the extension $\mathcal{B}$ over $\mathcal{A}_{[k-1]^*}$ is relative separable and relative atomless. We can construct such a $\mathcal{B}$ by adding countably many generators to $\mathcal{S}(\mathbf{E})$, "filling in" all possible relative atoms. Now use Lemma 4.27 to find a relative basis $g_1, g_2, \ldots$ of $\mathcal{B}$ over $\mathcal{A}_{[k-1]^*}$. Then, untangling the definitions, we have

$$\chi_{\mathbf{E}}(x_1, \ldots, x_k) = \sum_{j=1}^{\infty} g_i(x_1, \ldots, x_{k-1}) E(\chi_{\mathbf{E}_{x_k}}(x_1, \ldots, x_{k-1}) \cdot \overline{g_j(x_1, \ldots, x_{k-1})} | \mathcal{A}_{[k-1]^*}),$$

where as we recall $\mathbf{E}_{x_k}$ is the x_k-slice of $\mathbf{E}$. This formula shows that $\chi_{\mathbf{E}}$ is measurable in $\mathcal{A}_{[k]^*}$, so $\mathbf{E}$ is a measurable set in the same σ-algebra. □

With this definition, $\mathcal{S}(\mathbf{E})$ may depend on a change of $\mathbf{E}$ on a set of measure zero. To eliminate this possibility, it is convenient to introduce the somewhat more involved notion of the *essential* σ-algebra $\mathcal{E}(\mathbf{E})$ on $\mathbf{A}^{k-1}$. Let $[k]'$ denote the system of sets $\{T : T \subset [k], k \in T, |T| = k-1\}$. This is similar to the system $[k]^*$, but with the added constraint that the last coordinate must always be included. Denote the σ-algebra generated by all of the algebras $\mathcal{A}_T$ such that $T \in [k]'$ by $\mathcal{A}_{[k]'}$. Now define $\mathcal{E}(\mathbf{E})$ to be the smallest σ-algebra containing $\mathcal{A}_{[k-1]^*}$ in which all of the functions

$$g(x_1, \ldots, x_{k-1}) = \int_{\mathbf{A}} \chi_{\mathbf{E}}(x_1, \ldots, x_k) \cdot t(x_1, \ldots, x_k)\, d\mu(x_k)$$

are measurable, where $t \in L^{\infty}(\mathbf{A}^k, \mathcal{A}_{[k]'})$ is allowed to vary.

The following proposition characterizes the essential σ-algebras.

Proposition 4.29. *The following are satisfied by $\mathcal{E}(\boldsymbol{E})$, where $p : \boldsymbol{A}^k \to \boldsymbol{A}^{k-1}$ denotes the projection canceling the last coordinate:*

(i) Whenever $\mu(\boldsymbol{E} \triangle \boldsymbol{E}') = 0$, $\mathcal{E}(\boldsymbol{E}) = \mathcal{E}(\boldsymbol{E}')$ (insensitivity to measure-zero change).
(ii) If $\boldsymbol{E} \in \mathcal{A}_{[k]^*}$, then $\mathcal{E}(\boldsymbol{E})$ is relative separable over $\mathcal{A}_{[k-1]^*}$.
(iii) If $\boldsymbol{E} \in \mathcal{A}_{[k]^*}$, then the characteristic function $\chi_{\boldsymbol{E}}$ is almost measurable in $\langle p^{-1}(\mathcal{E}(\boldsymbol{E})), \mathcal{A}_{[k]'} \rangle$.
(iv) If $\boldsymbol{E} \in \mathcal{A}_{[k]^*}$, then the functions

$$\chi_{\boldsymbol{E},x_k}(x_1, \ldots, x_{k-1}) = \chi_{\boldsymbol{E}}(x_1, \ldots, x_k),$$

are measurable in $\mathcal{E}(\boldsymbol{E})$ for almost all $x_k \in \boldsymbol{A}$.

Proof. The first property is obvious, since $\mathcal{E}(\mathbf{E})$ is defined in terms of an integral; the functions we get with a measure-zero change will be precisely the same as the functions we got originally.

From Lemma 4.16, since the characteristic functions of the slices are in the Hilbert L^2 space corresponding to $\mathcal{S}(\mathbf{E})$ by definition, we conclude that $L^2(\mathbf{E}(\mathbf{E})) \subseteq L^2(\mathbf{S}(\mathbf{E}))$. Therefore $\mathcal{E}(\mathbf{E}) \subseteq \mathcal{S}(\mathbf{E})$. By Proposition 4.28, since $\mathbf{E} \in \mathcal{A}_{[k]^*}$, $\mathcal{S}(\mathbf{E})$ is relative separable over $\mathcal{A}_{[k-1]^*}$. Using the characterization of relative separability in terms of Hilbert spaces modules in Lemma 4.26, by Lemma 4.18, since $\mathcal{S}(\mathbf{E})$ is relative separable over $\mathcal{A}_{[k-1]^*}$, so is $\mathcal{E}(\mathbf{E})$.

For the third property, let

$$\begin{aligned}
\mathcal{B}_1 &= p^{-1}(\mathcal{A}(\mathbf{A}^{k-1})) = \mathcal{A}_{[k-1]}(\mathbf{A}^k),\\
\mathcal{B}_2 &= \mathcal{A}_{[k]'},\\
\mathcal{B}_3 &= p^{-1}(\mathcal{E}(\mathbf{E})),
\end{aligned}$$

be three σ-algebras on $\mathbf{A}^k$. It is easy to see that $\mathcal{B}_1$ and $\mathcal{B}_2$ are weakly orthogonal, and by definition, $\mathcal{B}_3$ is the σ-algebra generated by all the functions $E(\chi_{\mathbf{E}} \cdot t|\mathcal{B}_1)$, where $t \in L^\infty(\mathcal{B}_2)$, since projection onto $\mathcal{B}_1$ is clearly just integrating over the last coordinate. Therefore by Proposition 4.25, $\mathbf{E}$ is almost measurable in $\langle p^{-1}(\mathcal{E}(\mathbf{E})), \mathcal{A}_{[k]'}\rangle$, as desired.

The fourth part follows from Fubini's theorem on the ultraproduct, together with the third part. $\square$

4.7. **Higher order Fourier σ-algebras.** Here we finally define the kth-order Fourier σ-algebras on the ultraproduct group $\mathbf{A}$ and prove the first major result about them.

Let $\mathcal{B}$ be a σ-algebra on $\mathbf{A}$ that is contained in $\mathcal{A}$ and invariant under the group action, and let $\mathbf{G}$ be an element of $\mathcal{A}$. We say $\mathbf{G}$ is *relatively separable* with respect to $\mathcal{B}$ if there is a set of countably many elements $a_i \in \mathbf{A}$, $i \in \mathbb{N}$ such the set of translates

$$\{\mathbf{G} + a_i : i \in \mathbb{N}\},$$

together with $\mathcal{B}$, is dense the σ-algebra generated by all the translates, together with $\mathcal{B}$. Here density is in the usual sense, where sets are arbitrarily well-approximated if they are arbitrarily well-approximated in terms of symmetric differences. In other words, as σ-algebras, countably many translates are dense in all the translates.

We have the following:

Lemma 4.30. *If $\mathcal{B}_1 \subseteq \mathcal{B}_2$ are invariant σ-algebras and $\mathcal{B}_2$ is relative separable over $\mathcal{B}_1$ as a σ-algebra, then each element of $\mathcal{B}_2$ is relative separable over $\mathcal{B}_1$.*

Proof. Pick an element $\mathbf{F} \in \mathcal{B}_2$, and let $\mathcal{C}$ be the σ-algebra generated by $\mathcal{B}_1$ and all of the translates of $\mathbf{F}$. As $\mathcal{C} \subseteq \mathcal{B}_2$, there are countably many elements of $\mathcal{C}$ such that they, together with $\mathcal{B}_1$, generate a σ-algebra that is dense in $\mathcal{C}$. By construction of $\mathcal{C}$, it is easy to see that we can pick these elements to be themselves translates of $\mathbf{F}$. $\square$

Denote the set of relatively separable elements over $\mathcal{B}$ by $\Upsilon(\mathcal{B})$. Using the monotone class theorem, $\Upsilon(\mathcal{B})$ must also be a σ-algebra, and it is obvious that $\Upsilon(\mathcal{B})$ is invariant under the group action if $\mathcal{B}$ is. In a failure of terminology, the set of relatively separable elements over a σ-algebra will *not* be relative separable over that σ-algebra. In particular, even if $\mathcal{B}$ is the trivial σ-algebra, $\Upsilon(\mathcal{B})$ will be nonseparable.

The characterization of the higher order Fourier σ-algebras (once we define them) in this section will require one additional lemma, allowing us to preserve relative separability of sets when passing to coset σ-algebras.

Lemma 4.31. *Let $\boldsymbol{H}$ be an internal subgroup of $\boldsymbol{A}$, $\mathcal{B}$ an invariant sub-σ-algebra of $\mathcal{A}(\boldsymbol{A})$, and $\boldsymbol{G} \in \mathcal{A}(\boldsymbol{A}, \boldsymbol{H})$ a relative separable element over $\mathcal{B}$. Then $\boldsymbol{G}$ is also relative separable over $\mathcal{B} \cap \mathcal{A}(\boldsymbol{A}, \boldsymbol{H})$.*

Proof. Note that it is a trivial verification that the intersection of two σ-algebras is also a σ-algebra, so $\mathcal{B} \cap \mathcal{A}(\mathbf{A}, \mathbf{H})$ is indeed a σ-algebra and the statement of the lemma makes sense.

As $\mathbf{G}$ is relative separable over $\mathcal{B}$, there are a countable set $a_i \in \mathbf{A}$ such that all translates of $\mathbf{G}$ are arbitrarily well-approximated (in the sense of symmetric differences) in the σ-algebra generated by $\mathcal{B}$ and the sets $T = \{\mathbf{G} + a_i\}$. Therefore if $a \in \mathbf{A}$ and $\epsilon > 0$, by Lemma 4.23, there are sets $B_j^1 \in \langle T \rangle$ and $B_j^2 \in \mathcal{B}$, $1 \leq j \leq n$, such that

$$\mathbf{G}_\epsilon = \bigcup_{j=1}^{n} (B_j^1 \cap B_j^2)$$

is such that

$$\mu((\mathbf{G} + a) \triangle \mathbf{G}_\epsilon) \leq \epsilon.$$

By the definition of the coset σ-algebras, it is clear that

$$E(\chi_{\mathbf{G}_\epsilon} | \mathcal{A}(\mathbf{A}, \mathbf{H})) = \sum_{j=1}^{n} \chi_{B_j^1} E(\chi_{B_j^2} | \mathcal{A}(\mathbf{A}, \mathbf{H})).$$

No L^2 norms can increase under projection, and the projection onto $\mathcal{A}(\mathbf{A}, \mathbf{H})$ obviously leaves $\mathbf{G} + a$ unchanged, so $\mu((\mathbf{G} + a) \triangle \mathbf{G}_\epsilon) \leq \epsilon$ implies that

$$||E(\chi_{\mathbf{G}_\epsilon} | \mathcal{A}(\mathbf{A}, \mathbf{H})) - \chi_{\mathbf{G}}||_2 \leq \epsilon.$$

By Lemma 4.22, $\mathcal{B}$ and $\mathcal{A}(\mathbf{A}, \mathbf{H})$ are weakly orthogonal. Therefore all of the functions

$$E(\chi_{B_j^2} | \mathcal{A}(\mathbf{A}, \mathbf{H}))$$

are measurable in $\mathcal{B} \cap \mathcal{A}(\mathbf{A}, \mathbf{H})$, which is sufficient to show that $\mathbf{G} + a$ is arbitrarily well-approximated by elements of that intersection and countably many translates, as desired. □

It is now possible to define and characterize the basic objects of study of this subject, the Fourier σ-algebras $\mathcal{F}_i$ on $\mathbf{A}$. The definition is surprisingly straightforward; in terms of the operator Υ above, we just define $\mathcal{F}_i$ recursively, letting $\mathcal{F}_0 = \{\emptyset, \mathbf{A}\}$ be the trivial σ-algebra and defining

$$\mathcal{F}_i = \Upsilon(\mathcal{F}_{i-1})$$

for $i \geq 1$. Of course, all of the $\mathcal{F}_i$ are invariant under the group action.

Using the results developed above, we can link the $\mathcal{F}_i$ back to the Gowers norms, via a detour through product spaces. Let $\mathbf{D}_k$ denote the diagonal subgroup of $\mathbf{A}^k$, consisting of elements $(x_1, \ldots, x_k)$ with $\sum_{j=1}^{k} x_j = 0$. Let $\tau_k : \mathbf{A}^k \to \mathbf{A}$ be given by

$$\tau_k(x_1, \ldots, x_k) = \sum_{j=1}^{k} x_j.$$

Clearly, τ_k descends to a homomorphism from the quotient group $\mathbf{A}^k / \mathbf{D}_k$, which we will also denote by τ_k.

Lemma 4.32. *The map $\tau_k : \boldsymbol{A}^k/\boldsymbol{D}_k \to \boldsymbol{A}$ is a group isomorphism and gives rise to a measure-preserving equivalence between $\mathcal{A}(\boldsymbol{A}^k, \boldsymbol{D}_k)$ and $\mathcal{A}(\boldsymbol{A})$.*

Proof. That the map is an isomorphism is true generally for any abelian group.

For the second part, note that $\mathcal{A}(\mathbf{A}^k, \mathbf{D}_k)$ is a σ-algebra on $\mathbf{A}^k$, so we are using the map $\tau_k : \mathbf{A}^k \to \mathbf{A}$. By a measure-preserving equivalence, we mean that images and preimages of measurable sets are measurable and the measure is unchanged by the map. We will show that we can factor τ_k through an intermediate measure space to get two evident measure-preserving equivalences.

Let $\mathbf{C}_k$ be the subgroup of $\mathbf{A}_k$ with the kth coordinate is always equal to zero; that is,

$$\mathbf{C} = \{(x_1, \ldots, x_k) \in \mathbf{A}^k : x_k = 0\}.$$

Let $\alpha_k : \mathbf{A}^k \to \mathbf{A}^k$ be the map given by

$$\alpha_k(x_1, \ldots, x_{k-1}, x_k) = \left(x_1, \ldots, x_{k-1}, \sum_{j=1}^{k} x_j \right).$$

Note that α_k is an invertible affine linear transformation with determinant one. By construction of $\mathcal{A}(\mathbf{A}^k)$, the measure is invariant under such transformations (in much the same way that the Lebesgue measure on $\mathbb{R}^d$ is, for example). Furthermore, α_k takes $\mathbf{D}_k$ to $\mathbf{C}_k$, so it is a measure-preserving equivalence between $\mathcal{A}(\mathbf{A}, \mathbf{D}_k)$ and $(\mathbf{A}, \mathbf{C}_k)$.

Now let $\beta_k : \mathbf{A}^k \to \mathbf{A}$ be the projection onto the kth coordinate. It is obvious that, by the definition of the coset σ-algebras, β_k is an equivalence between $\mathcal{A}(\mathbf{A}^k, \mathbf{C}_k)$ and $\mathcal{A}(\mathbf{A})$ (the elements of the former are precisely the unions of cosets of $\mathbf{C}_k$ which are measurable in $\mathcal{A}(\mathbf{A}^k)$; that is, cylinder sets in the kth coordinate). That measure is preserved follows from the definition of the product σ-algebra.

Since $\tau_k = \beta_k \circ \alpha_k$ is the composition of two measure-preserving equivalences, it itself is a measure-preserving equivalence. □

The following proposition ties together much of the above work in giving an understandable lifting of $\mathcal{F}_{k-1}$ to the product.

Proposition 4.33. *The map $\tau_k : \boldsymbol{A}^k/\boldsymbol{D}_k \to \boldsymbol{A}$ gives a measure-preserving equivalence between the σ-algebra $\mathcal{A}(\boldsymbol{A}^k/\boldsymbol{D}_k) \cap \mathcal{A}_{[k]^*}$ and $\mathcal{F}_{k-1}$.*

Proof. The proof is by induction on k.

If $k = 1$, the diagonal subgroup $\mathbf{D}_1$ is trivial, so $\mathcal{A}(\mathbf{A}^1, \mathbf{D}_1) = \mathcal{A}(\mathbf{A})$, while $\alpha_{[1]^*}$ is the trivial σ-algebra on $\mathbf{A}$. By definition, $\mathcal{F}_0$ is the trivial σ-algebra on $\mathbf{A}$ as well, and τ_1 is the identity map. The base case is therefore verified.

Assume that the proposition holds for $k-1$, and let $\mathbf{H}$ be a set in $\mathcal{A}(\mathbf{A})$. Let

$$\mathbf{H}_k = \tau_k^{-1}(\mathbf{H})$$

be the corresponding subset of $\mathbf{A}^k$, which is by Lemma 4.32 an element of $\mathcal{A}(\mathbf{A}^k)$. Therefore it suffices to prove that $\mathbf{H}$ is measurable in $\mathcal{F}_{k-1}$ if and only if $\mathbf{H}_k$ is measurable in $\mathcal{A}_{[k]^*}(\mathbf{A}^k)$. For convenience, let $f = \chi_{\mathbf{H}}$, the characteristic function of $\mathbf{H}$. Then if we let

$$f_k(x_1, \ldots, x_k) = f\left(\sum_{j=1}^{k} x_j \right),$$

as in the above discussion of octahedral norms, it is clear that $f_k = \chi_{\mathbf{H}_k}$.

In one direction, assume that $\mathbf{H}$ is measurable in $\mathcal{F}_{k-1}$, so $\mathbf{H}$ is relative separable over $\mathcal{F}_{k-2}$. The slice σ-algebra $\mathcal{S}(\mathbf{H}_k)$ on $\mathbf{A}^{k-1}$ consists of translates of the slices in the kth coordinate of $\mathbf{H}_k$, which are all the same and equal to $\mathbf{H}_{k-1}$ (to see this, it is perhaps easiest to consider the respective characteristic functions). Thus $\mathcal{S}(\mathbf{H}_k)$ is relative separable over $\tau_k^{-1}(\mathcal{F}_{k-2})$. Therefore by the inductive hypothesis, $\mathcal{S}(\mathbf{H}_k)$ is relative separable over $\mathcal{A}_{[k-1]^*}$. By Proposition 4.28, $\mathbf{H}_k$ is measurable in $\mathcal{A}_{[k]^*}$.

In the other direction, assume that $\mathbf{H}_k \in \mathcal{A}_{[k]^*}$. We claim that the essential σ-algebra $\mathcal{E}(\mathbf{H}_k)$ is an invariant σ-algebra on $\mathbf{A}^{k-1}$. This is a routine verification from the definition. By the fourth property in Proposition 4.29, almost all of the slices of $\mathbf{H}_k$ are measurable in $\mathcal{E}(\mathbf{H}_k)$. But as previously mentioned, all the slices of $\mathbf{H}_k$ are translates of $\mathbf{H}_{k-1}$, so as $\mathcal{E}(\mathbf{H}_k)$ is invariant, $\mathbf{H}_{k-1}$ itself must be in $\mathcal{E}(\mathbf{H}_k)$. By the second property in Proposition 4.29, $\mathcal{E}(\mathbf{H}_k)$ is relative separable over $\mathcal{A}_{[k-1]^*}$, so using Lemma 4.30, $\mathbf{H}_{k-1}$ is relative separable over $\mathcal{A}_{[k-1]^*}$. By Lemma 4.31, $\mathbf{H}_{k-1}$ is relative separable over $\mathcal{A}_{[k-1]^*} \cap \mathcal{A}(\mathbf{A}^{k-1}, \mathbf{D}_{k-1})$. By the induction hypothesis, therefore, by mapping via τ_k^{-1} we get that $\mathbf{H}$ is relative separable over $\mathcal{F}_{k-2}$. □

As a corollary, we get the first structure theorem in the subject, justifying the above definition of of the Fourier σ-algebras:

Theorem 4.34. *Let f be a function in $L^2(\boldsymbol{A}, \mathcal{A})$, and define*

$$f_k(x_1, \ldots, x_k) = f\left(\sum_{j=1}^{k} x_j\right).$$

Then the following are equivalent:

(i) The function f is measurable in $\mathcal{F}_{k-1}$.
(ii) The function f is orthogonal to every $g \in L^2(\boldsymbol{A}, \mathcal{A})$ such that $||g||_{U^k} = 0$.
(iii) The function f_k is measurable in $\mathcal{A}_{[k]^*}$.
(iv) The function f_k is orthogonal to every $g \in L^2(\boldsymbol{A}^k, \mathcal{A})$ such that $||g||_{O^k} = 0$.

Proof. By Proposition 4.33, we immediately have the equivalence of (i) and (iii) and (ii) and (iv). Equivalence of (iii) and (iv) is given by Proposition 4.15. □

5. Further results and extensions

5.1. **The second structure theorem.** As mentioned in the introduction, Szegedy's second structure theorem is the following statement:

Theorem 5.1. *The Hilbert space $L^2(\boldsymbol{A}, \mathcal{F}_k)$ is generated as a module by rank-one modules over $L^\infty(\boldsymbol{A}, \mathcal{F}_{k-1})$ which are pairwise orthogonal.*

The following corollary is immediate:

Corollary 5.2. *For each $f \in L^2(\boldsymbol{A}, \mathcal{A})$ and each $k > 1$, there is a unique decomposition*

$$f = g + \sum_{j=1}^{\infty} f_j,$$

where $||g||_{U^k} = 0$ and each f_j is contained in a different rank-one module over $L^\infty(\mathcal{F}_{k-2})$.

Proof. By Theorem 4.34, we can uniquely decompose

$$f = g + f',$$

where $||g||_{U^k} = 0$ and $f' \in L^2(\mathcal{F}_{k=1})$. By Theorem 5.1 and the definition of a Hilbert space module (which requires only that generators span a dense subspace of the module), we have a further decomposition

$$f' = \sum_{j=1}^{\infty} f_j,$$

where each f_j is contained in a different rank-one module over $L^\infty(\mathcal{F}_{k-2})$. Furthermore, each rank-one module is generated by a function $\alpha : \mathbf{A} \to \mathbb{C}$ such that $|\alpha| = 1$ and $\alpha(x)\overline{\alpha(x+t)}$ is measurable in $\mathcal{F}_{k-1}$. □

Although we will not prove Theorem 5.1 in close detail, in what follows we will outline the pieces one needs to do so and give some indication of the proof. First, using language first employed in [17], we will discuss the cubic structure on powers of $\mathbf{A}$.

For a positive integer k, let V_k denote the set of all subsets of $[k] = \{1, \ldots, k\}$. We can think of the 2^k elements of V_k as vertices of a k-dimensional cube. For $d < k$, a d-dimensional face of V_k is defined in the natural way; that is, it is a maximal set of subsets of $[k]$ that is constant on $k - d$ coordinates; that is, for each of the chosen coordinates $j \in [k]$, either all of the sets of the face contain j or all omit j. Define the subgroup $\mathbf{B}_k \subset \mathbf{A}^{V_k}$ as the collection of elements $\{a_i\}_{i \in V_k}$ satisfying all possible equations of the form

$$a_p - a_q + a_r - a_s = 0,$$

where $\{p, q\}$ is a one-dimensional face and $\{p, q, r, s\}$ is a two-dimensional face. If $v \in V_k$ is a vertex, let $S(v)$ refer to v together with its k neighbors obtained by including an extra element of $[k]$ or omitting an element (if V_k is pictured as a k-dimensional cube, then the elements of $S(v)$ are the literal geometrically closest neighbors of v). Somewhat fancifully, $S(v)$ is referred to as the *spider* of v.

Viewing V_k as a cube, there is enough symmetry so that any given vertex is interchangeable with any other. For definitiveness, therefore, we will isolate one, the zero vertex that is represented by the empty set in $[k]$, and refer to it as 0. The following lemma, whose proof is in the same vein as Lemma 4.33, would apply equally well to any other vertex.

Lemma 5.3. *Label the elements of the spider $S(0)$ as $0, 1, 2 \ldots, k$, where 0 is the zero vertex and j is the set $\{j\} \subset [k]$ for $j \neq 0$. Let $v \in V_k$ and let $\delta_k : \boldsymbol{A}^{S(0)} \to \boldsymbol{A}^{V_k}$ denote the group homomorphism given by*

$$\delta_k(a_0, a_1, \ldots, a_k)_v = a_0 + \sum_{i \in v} a_i.$$

That is, the vth coordinate of δ_k is calculated by adding a_0 to the sum of the a_i such that $\{i\} \subset v$. Then the action of δ_k on the measure space $\mathcal{A}(\boldsymbol{A}^{k+1})$ gives a measure-preserving equivalence of the groups $\boldsymbol{A}^{k+1}$ and $\boldsymbol{B}_k$. □

For a vertex $v \in V_k$, let τ_v denote the projection $\mathbf{B}_k \to \mathbf{A}$ to the vth coordinate. We can pull back the σ-algebra $\mathcal{A}(\mathbf{A})$ via this map to get a σ-algebra on $\mathbf{B}_k$ which will be denoted $\mathcal{A}_v(\mathbf{B}_k)$. Using the equivalence of $\mathbf{B}_k$ and $\mathbf{A}^{k+1}$, we can define a similar map $\nu_v : \mathbf{A}^{k+1} \to \mathbf{A}$ by the formula

$$\nu_v = \tau_v \circ \delta_k.$$

The following lemma is a trivial verification, expressing the Gowers norm in terms of the cubic structure.

Lemma 5.4. *If $f : \boldsymbol{A} \to \mathbb{C}$ is a $\mathcal{A}$-measurable function, then*

$$||f||_{U^k}^{2^k} = \int_{\boldsymbol{B}_k} \prod_{v \in V_k} C^{|v|}[f(\tau_v(x))] \, d\mu(x),$$

where $|v|$ is the literal size of v as a subset of $[k]$ and $Cf = \overline{f}$ is the complex conjugation operator. □

Finally, a third lemma relates the cubic structure with the Fourier σ-algebras. Its proof uses the previous lemma's characterization of the Gowers norm together with Theorem 4.34.

Lemma 5.5. *For every $v \in V_k$,*

$$\mathcal{A}_v \cap \langle \{\mathcal{A}_w | w \neq v\} \rangle = \tau_v^{-1}(\mathcal{F}_{k-1}).$$

on $\boldsymbol{B}_k$. □

An automorphism of a measure space is a measure-preserving equivalence from that measure space to itself. Let $\mathcal{H}$ be an $\mathbf{A}$-invariant σ-algebra on $\mathbf{A}$ sandwiched between two Fourier σ-algebras, as follows:

$$\mathcal{F}_{k-1} \subseteq \mathcal{H} \subseteq \mathcal{F}_k.$$

We will call an automorphism of $\mathcal{H}$ a *k-automorphism* if

(i) it commutes with the action of $\mathbf{A}$ on itself, and
(ii) it fixes $\mathcal{F}_{k-1}$.

Let $\mathrm{Aut}_k(\mathcal{H})$ denote the set of k-automorphisms of $\mathcal{H}$. By transferring in the usual way between a σ-algebra and the L^2 space over it, a k-automorphism σ induces an action on $L^2(\mathcal{H})$, which we will also denote by σ, such that

$$E(\sigma(f)|\mathcal{F}_{k-1}) = E(f|\mathcal{F}_{k-1})$$

for $f \in L^2(\mathcal{H})$. Since the underlying measure spaces are finite, $L^\infty \subset L^2$, so σ acts on $L^\infty(\mathcal{H})$ as well.

Each k-automorphism of $\mathcal{H}$ gives rise to a set of actions on the direct sum of 2^k copies of $L^\infty(\mathcal{H})$ in the following way. Let $T \subset V_{k+1}$ be a set of vertices. Define the action $l_{T,\sigma}$ on this direct sum by

$$l_{T,\sigma}\left(\bigoplus_{v\in V_{k+1}} f_v\right) = \bigoplus_{v\in V_{k+1}} \sigma_v f_v,$$

where $\{f_v\}_{v\in V_{k+1}}$ is a collection of functions each in $L^\infty(\mathcal{H})$ and $\sigma_v = \sigma$ whenever $v \in T$ and is the identity otherwise. If T is a face of V^{k+1}, this action is called a *face action.*

We define a second functional $\tilde{U}^k$ on collections of functions on $\mathbf{A}$, one for each $v \in V_k$:

$$||\{f_v\}_{v\in V_k}||_{\tilde{U}^k} = \int_{\mathbf{B}_k} \prod_{v\in V_k} f_v(\tau_v(x)).$$

Note that this does not include the complex conjugation operator. The importance of this functional is that it is preserved by edge actions:

Lemma 5.6. *If $\mathcal{H}$ is as before, $E \subset V_{k+1}$ is an edge (that is, a one-dimensional face), and $\sigma \in \mathrm{Aut}_k(\mathcal{H})$, then the action $l_{E,\sigma}$ preserves $\tilde{U}^k$ in the sense that*

$$\left|\left| l_{E,\sigma}\left(\bigoplus_{v\in V_{k+1}} f_v\right)\right|\right|_{\tilde{U}^k} = \left|\left|\bigoplus_{v\in V_{k+1}} f_v\right|\right|$$

for all collections $\{f_v\}_{v\in V_{k+1}}$ of functions in $L^\infty(\mathcal{H})$. □

The proof of the above lemma is somewhat tricky and relies on Lemma 5.3 in order to pass from $\mathbf{B}_k$ to a power of $\mathbf{A}$ as well as the characterization of the Fourier σ-algebras in Lemma 5.5.

In a rather different direction, we now turn to a discussion of what will be called *pre-cocycles* on the ultraproduct group $\mathbf{A}$. These objects are variants of the cocycles used by Furstenberg and Weiss in [10] in their analysis of certain sparse averages in ergodic theory. The intuition here is borrowed from the construction of group extensions in group theory, but placed firmly in the context of ergodic theory. The discussion will be pitched at a somewhat higher level of generality than is strictly necessary, as eventually we will just take G to be a unitary group over the complex numbers. One should note at this stage that this is the first entrance of a nonabelian group into the theory; later work (for example, [31]) has made it clear that this noncommutativity is an essential part of the algebraic structure of higher order Fourier analysis.

For now, though, let G be a second countable compact topological group with its associated Borel σ-algebra $\mathcal{G}$ and normalized Haar measure ν. As G will in general not be abelian, we will denote the group operation by multiplication. A function

$$f : \mathbf{A} \to G$$

is called a pre-cocycle of order k if f is measurable in $\mathcal{F}_k$ and the function

$$F_t(a) = f(a+t)f(a)^{-1}$$

is measurable in $\mathcal{F}_{k-1}$ for every fixed $t \in \mathbf{A}$. This seemingly arbitrary form should be reminiscent of the promised generators α of the pairwise orthogonal $L^\infty(\mathcal{F}_{k-1})$ modules in the statement of Theorem 5.1. In the context of the circle group (the unit circle of $\mathcal{C}$), the inverse operation corresponds to complex conjugation, so if we construct pre-cocycles into the unit circle they will give us exactly what we want. Studying objects of this form in more generality will allow us to prove the existence of such generators.

Two pre-cocycles f_1 and f_2 of order k will be called *equivalent* if there exists a function h measurable in $\mathcal{F}_{k-1}$ and an element $g \in G$ such that

$$f_1 = f_3 \cdot f_2 \cdot g.$$

Measurability is of course now defined with respect to the Borel σ-algebra $\mathcal{G}$ on G rather than the Borel σ-algebra on $\mathbb{R}$ or $\mathbb{C}$.

A pre-cocycle gives rise to two actions, one of $\mathbf{A}$ on $\mathbf{A} \times G$ and one of G on $\mathbf{A} \times G$. The former is given by

$$a \cdot (b, h) = (b + a, f(b + a)f(b)^{-1}h)$$

and the latter is given by

$$g \cdot (b, h) = (b, hg).$$

It is a routine verification that these two actions commute with each other, and by the invariance of the measures μ and ν both actions preserve the product measure $\mu \times \nu$. By the definition of a cocycle, the product σ-algebra $\mathcal{F}_{k-1} \times \mathcal{G}$ generated by rectangles is invariant under the action of $\mathbf{A}$.

To use pre-cocycles, we will have to use some very technical results in ergodic theory. If $\mathcal{I}$ is the set of $\mathbf{A}$-invariant sets in $\mathcal{F}_{k-1} \times \mathcal{G}$, it can be checked that

(i) The set $\mathcal{I}$ is in fact a σ-algebra.
(ii) The action of G, as defined above, leaves $\mathcal{I}$ invariant.
(iii) Furthermore, G acts ergodically on $\mathcal{I}$, which means that the only subsets of $\mathcal{I}$ fixed by the action are the sets in the trivial σ-algebra.

A general lemma that can be found in, for example, [11] then implies that there is a closed subgroup $H \subseteq G$ such that the action of G on $\mathcal{I}$ is isomorphic to the action of G on the left coset space of H in G. This subgroup is called the Mackey group corresponding to the pre-cocycle f; it is defined only up to conjugacy class. If $\mathcal{I}$ is trivial, then we say f is *minimal*. By another ergodic-theoretic lemma whose proof can also be found in [11], every pre-cocycle f is equivalent to a minimal pre-cocycle $f' : \mathbf{A} \to H$ whose image is the Mackey group of f.

The argument then proceeds to connect these pre-cocycles back with structures that have already been defined. Using an averaging operator, Lemma 4.16, the invariance of various structures with respect to the defined actions, and a density argument, it is now possible to prove the following:

Proposition 5.7. *If $f : \boldsymbol{A} \to G$ is a minimal pre-cocyle, then there is a σ-algebra $\mathcal{H}$ on $\boldsymbol{A}$ such that $\mathcal{F}_{k-1} \subseteq \mathcal{H} \subseteq \mathcal{F}_k$ and a measure-preserving equivalence $\phi : \boldsymbol{A} \times G \to \boldsymbol{A}$ between $\mathcal{F}_{k-1} \times \mathcal{G}$ and $\mathcal{H}$ that commutes with the above-defined action of $\boldsymbol{A}$. The equivalence ϕ also behaves nicely when restricted to the trivial σ-algebra on G, giving an equivalence between $\mathcal{F}_{k-1} \times \{\emptyset, G\}$ and $\mathcal{F}_{k-1}$.* □

As a corollary, if $\mathcal{H}$ is the σ-algebra constructed in the proposition, then G induces a faithful action on $\mathcal{H}$, meaning that two distinct elements of G induce different permutations of $\mathcal{H}$. Denote this action by

$$p : G \to \mathrm{Aut}_k(\mathcal{H}).$$

We are now ready to put together the ergodic theory into one simple and useful result. Let $f : \mathbf{A} \to G$ be a minimal pre-cocycle of order k, let $\mathcal{H}$ be the σ-algebra guaranteed by ergodic theory in the form of Proposition 5.7, and let g_1 and g_2 be arbitrary elements of G. Pick a vertex $w \in V_{k+1}$ and let E_1 and E_2 be two edges of V_{k+1} that meet at w. We make the clever definition

$$c_w = \left[l_{E_1,p(g_1)}, l_{E_2,p(g_2)}\right],$$

where the brackets indicate the commutator of the actions. Of course, c_w acts on the space

$$\bigoplus_{v \in V_{k+1}} L^\infty(\mathcal{H}),$$

and it is easy to calculate that the action is given by

$$c_w = l_{w,p([g_1,g_2])}.$$

From Lemma 5.6, the form $\tilde{U}^k$ is preserved by c_w, so certainly even if we pick various other vertices $w_1, w_2, \ldots \in V_{k+1}$ and compose them, the composition of the operators still preserves $\tilde{U}^k$. We can now use a trick to deduce that $p([g_1, g_2])$ itself must act trivially: for a function $h \in L^\infty(\mathcal{H})$, define

$$g = h - p([g_1, g_2])h.$$

Then by definition

$$||g||_{U^{k+1}}^{2^{k+1}} = \int_{\mathbf{B}_k} \prod_{v \in V_{k+1}} C^{|v|}[f(\tau_v(x)) - p([g_1, g_2])f(\tau_v(x))] \, d\mu(x).$$

Expanding this product and using that all of the c_w act trivially, we get perfect cancelation; that is, $||g||_{U^{k+1}}$. Since $\mathcal{F}_k$, by Theorem 4.34 g must be orthogonal to itself, hence zero, hence $p([g_1, g_2])$ acts trivially. Because p is a faithful action, $[g_1, g_2]$ is itself the identity element; that is, g_1 commutes with g_2. Since we picked g_1 and g_2 arbitrarily, we arrive at the following surprising conclusion:

Proposition 5.8. *Let $f : \boldsymbol{A} \to G$ be a minimal pre-cocycle of order k. Then G is abelian.* □

In the language of ergodic theory, this result is the statement that in this context isometric extensions are abelian.

To proceed with the proof of Theorem 5.1, we introduce the "higher order group algebras," a higher order analogue of a convolution algebra. First let

$$M_k = L^2(\mathbf{A} \times \mathbf{A}, \mathcal{F}_k \times \mathcal{F}_k, \mu \times \mu)$$

be the space of Hilbert-Schmidt operators, which comes equipped with a multiplication operation $\circ$ in the usual way as follows:

$$(K_1 \circ K_2)(x, y) = \int_{\mathbf{A}} K_1(x, z)K_2(z, y) \, d\mu(z).$$

Now let $C_k \subset M_k$ be the kernels that have the property such that, as a function of t, $K(x + t, y + t)$ is in $L^2(\mathcal{F}_{k-1})$ for every $x, y \in \mathbf{A}$. It is fairly easy to see that both C_k and M_k are C^*-algebras; that they are both closed under involution is trivial and it follows from general theory that M_k closed with respect to the usual operator norm. That C_k is as well follows from yet another use of Lemma 4.16.

If C is an operator in $C_k \cap L^\infty(\mathbf{A} \times \mathbf{A})$, then C^*C is self-adjoint. The image $\operatorname{im}(C^*C)$ of a self-adjoint operator is a separable Hilbert subspace of $L^2(\mathbf{A})$. But

$$\operatorname{im}(C^*C) = \operatorname{im}(C),$$

so in fact $\mathrm{im}(C)$ is a separable Hilbert subspace of $L^2(\mathbf{A})$. By standard theory, we have a decomposition

$$\mathrm{im}(C) = \bigoplus_{i=1}^{\infty} V_i,$$

where the V_i are finite-dimensional eigenspaces corresponding to different eigenvectors, and each $V_i \subset L^\infty(\mathbf{A})$.

The following is the last major proposition, using the abelian extension result to conclude the following:

Proposition 5.9. *As defined above, each V_i is contained in a finite-rank module over $L^\infty(\mathcal{F}_{k-1})$, and this finite-rank module can be decomposed into rank one modules, each of which is generated by a function $\alpha : \boldsymbol{A} \to \mathbb{C}$ with $|\alpha| = 1$ and $\alpha(x)\overline{\alpha(x+t)} \in \mathcal{F}_{k-1}$.*

Proof (sketch). Let P_i be the Hilbert space projection to V_i, which are all elements of C_k. Let $f_1, \ldots, f_d \in L^\infty(\mathbf{A})$ be an orthonormal basis for V_i, and let $f : \mathbf{A} \to \mathcal{C}^d$ be defined as

$$f(x) = (f_1(x), \ldots, f_d(x)).$$

Let $\mathbf{E} \subseteq \mathbf{A}$ be the set of $x \in \mathbf{A}$ on which there exist elements $t_1, \ldots, t_d \in \mathbf{A}$ such that the matrix

$$(f(x+t_1), f(x+t_2), \ldots, f(x+t_d))$$

has full rank d. We can use orthonormality together with the fact that the P_i are elements of C_k to show that $\mathbf{E}$ is a set of positive measure that is contained in $\mathcal{F}_{k-1}$.

We can then employ the Gram-Schmidt process on $\mathbf{E}$ for the columns of the above matrix. We get functions $o_1, o_2, \ldots, o_d : \mathbf{E} \to \mathbb{C}^d$ such that $\{o_i(x)\}_{i=1}^d$ is an orthonormal basis for every $x \in \mathbf{E}$ and we can express

$$o_i(x) = \sum_{j=1}^{d} \lambda_{i,j}(x) f(x+t_j)$$

for certain coefficients $\lambda_{i,j}$. It is clear that these coefficients are measurable in $\mathcal{F}_{k-1}$ as well.

To extend the o_i to the whole of $\mathbf{A}$, find countably many subsets $\mathbf{E}_1, \mathbf{E}_2, \ldots$ of $\mathbf{E}$, all measurable in $\mathcal{F}_{k-1}$, such that $\mathbf{A}$ is a disjoint union of translates of these sets. That this is possible follows from a result in ergodic theory known as Rohlin's lemma (see, for instance, [11]), together with the fact that the system of $\mathbf{A}$ acting on itself with σ-algebra $\mathcal{F}_{k-1}$ is trivially ergodic. Then we can translate the basis $\{o_i\}_{i=1}^\infty$ accordingly to extend the basis to all of $\mathbf{A}$. Using the above equation, any translate of f can be expressed in this basis with coefficients measurable in $\mathcal{F}_{k-1}$. Therefore V_i is contained in an $L^\infty(\mathcal{A})$-module of rank d.

To show that this module can be decomposed into rank-one modules, we use the abelian extension result. If $O(x)$ is the unitary matrix formed by the $o_i(x)$, the association $x \mapsto O(x)$ gives a pre-cocycle of degree k from $\mathbf{A}$ to the unitary group $U(d)$. This is essentially by construction; since the projections are in C_k we have that $\langle f(x), f(x+t)\rangle$ is measurable in $\mathcal{F}_{k-1}$ for each fixed t, which is the requirement for the map to be a pre-cocycle. Since every pre-cocycle is equivalent to a minimal pre-cocycle, $O(x)$ is equivalent to an $O'(x)$ going to some subgroup H of $U(d)$ that, by Proposition 5.8, must be abelian. As H is abelian, by basic representation theory it decomposes into one-dimensional irreducible representations, so we can find pre-cocycles α taking values in $\mathbb{C}$ of absolute value 1. In the context of the unit circle, inverse is equivalent to complex conjugation, so that α is a pre-cocycle immediately implies that α has the desired form. ☐

The proof of Theorem 5.1 can now proceed. We will prove that every invariant $\mathcal{H}$ such that $\mathcal{F}_{k-1} \subset \mathcal{H} \subset \mathcal{F}_k$ and $\mathcal{H}$ is relative separable over $\mathcal{F}_{k-1}$ is generated by rank-one modules of the form given by Proposition 5.9. To do this, we will construct for each $x \in \mathbf{A}$ an infinite (that is, $\mathbb{N} \times \mathbb{N}$) matrix $M_{x,y}$ such that the function $x \mapsto M_{x,y}$ for a certain fixed y generates a σ algebra containing $\mathcal{H}$ and such that each entry of $M_{x,y}$ is in C_k. This implies by Proposition 5.9 that $x \mapsto M_{x,y}$ is measurable in a σ-algebra generated by rank-one modules of the desired form, so in particular $\mathcal{H}$ is. Because $\mathcal{F}_k$ is the union of all invariant σ-algebras that are relative separable over $\mathcal{F}_{k-1}$, this is sufficient to prove Theorem 5.1.

To construct such a matrix, let $f_1, f_2, \ldots$ be the relative orthonormal basis of $L^2(\mathcal{H})$ over $L^2(\mathcal{F}_{k-1})$ guaranteed by Lemma 4.27, meaning that

$$E(f_i f_j | \mathcal{F}_{k-1}) = \delta_{ij}.$$

For any $f \in L^2(\mathcal{H})$, therefore, we can uniquely write

$$f(x) = \sum_{j=1}^{\infty} \lambda_j(x) f_j(x),$$

where the $\lambda_j = E(f f_j | \mathcal{F}_{k-1})$ are measurable in $\mathcal{F}_{k-1}$. Therefore in particular, we have uniquely defined functions $\lambda_{i,j}(t,x)$ such that

$$f_i(xt) = \sum_{j=1}^{\infty} \lambda_{i,j}(t,x) f_j(x).$$

Let $M_{x,y}$ be the infinite matrix whose (i,j)th entry is $\lambda_{i,j}(y-x, x)$. From the definition, it is straightforward to verify that

$$M_{x,y} M_{y,z} = M_{x,z} \qquad \text{for almost all } (x,y,z),$$
$$M_{x,y} = M_{y,x}^{-1} = M_{x,y}^{*} \qquad \text{for almost all } (x,y).$$

Therefore there is some y such that the above equations are verified for almost all x and z; fix this y. Using these two equations, we can conclude that since $M_{x,y} M_{x+t,y}$ is measurable in $\mathcal{F}_{k-1}$ for every fixed t, $M_{x,y}$ is measurable in $\mathcal{F}_k$; therefore each entry of $M_{x,y}$ is in C_k. By construction, the σ-algebra generated by $x \mapsto M_{x,y}$ contains all of $\mathcal{H}$, and we are done.

5.2. **Prospects for extensions to infinite spaces.** There are clear reasons why one would want to define higher order Fourier-analytic structures on infinite measure spaces. In ordinary Fourier analysis, for instance, the theory of the Fourier transform on $\mathbb{R}$ with the Lebesgue measure is fundamental and widely useful. If we are considering locally compact groups with Haar measure, the measure space is infinite if and only if the group is not compact. Of course, $\mathbb{R}$ falls into this category, as do many important number-theoretic groups, such as the adele groups of number fields. Ordinary Fourier analysis on these adele groups has famously yielded fantastic results, starting with Tate's proof of the functional equation [32], and it is not unreasonable to want to extend the finer control given by higher order Fourier analysis to this setting.

Unfortunately, there are several difficulties which make this prospect remote, although perhaps still possible. First, if we are given spaces A_i with infinite measures μ_i, the measure μ on the ultraproduct $\mathbf{A}$ is not σ-finite. For if $(\mathbf{A}, \mathcal{A}, \mu)$ were σ-finite, then it is the union of countably many elements of $\mathcal{A}$ of finite measure. Up to measure zero, therefore, it is the union of countably many elements of $\mathcal{S}$ of finite measure. But by countable compactness, we can rewrite this countable union as a finite union, which implies that $\mu(\mathbf{A})$ is finite.

The failure of σ-finiteness has immediate analytic consequences. For example, Fubini's theorem does not in general hold for such spaces (see Section 252 in [9] for the most general form of the theorem). This point is worth emphasizing because it is erroneously claimed in several sources, including for example in Section 3.9 of [26], that a Fubini-type theorem can be recovered for complete measure spaces even in the absence of σ-finiteness. In fact, we have the following counterexample, which can be found in [9]:

Proposition 5.10. *Let $(X, \mathcal{A}, \mu)$ be the interval $[0, 1]$ with Lebesgue measure and let $(Y, \mathcal{B}, \nu)$ be the interval $[0, 1]$ with counting measure (i.e., every set is measurable, with the measure given by the number of elements in the set if finite and infinity otherwise). Then there exists a function f such that*

$$\int_{X\times Y} |f(x, y)| \, d(\mu \times \nu)(x, y) < \infty$$

but

$$\int_X \int_Y f(x, y) \, d\mu(x) \, d\nu(y) \neq \int_Y \int_X f(x, y) \, d\nu(y) \, d\mu(x).$$

Proof. Let

$$W = \{(x, x) : x \in [0, 1]\} \subset X \times Y,$$

and set $f = \chi_W$, the characteristic function of the diagonal. We can write

$$W = \bigcap_{n\geq 1} \bigcup_{k=0}^{n} \left[\frac{k}{n+1}, \frac{k+1}{n+1}\right] \times \left[\frac{k}{n+1}, \frac{k+1}{n+1}\right],$$

so W is in the σ-algebra generated by rectangles in the product space and is therefore measurable; in particular, f is therefore measurable as well.

Use part (c) of the Theorem in 251I of [9], which tells us that our product measure $\mu \times \nu$ agrees with the "complete locally determined" product measure on finite measurable sets, and the Corollary in 252F of [9], which states that if the horizontal section W_y is μ-negligible for ν-almost every $y \in Y$ then the complete locally determined product measure of W is zero. Since the horizontal section W_y is always a single point, which is μ-negligible, the complete locally determined product measure of W is zero, and hence $(\mu \times \nu)(W) = 0$. Therefore

$$\int_{X\times Y} |f(x, y)| \, d(\mu \times \nu)(x, y) = 0.$$

But a quick calculation gives

$$\int_X \int_Y f(x, y) \, d\mu(x) \, d\nu(y) = \int_X \mu(W_y) \, d\nu(y) = 0$$

and

$$\int_Y \int_X f(x, y) \, d\nu(y) \, d\mu(x) = \int_Y \nu(W_x) \, d\nu(y) = 1.$$

□

Certainly, both measure spaces in the above proposition are complete, so this gives us a concrete counterexample and a justification for taking great care with Fubini's theorem in such settings. Fortunately, Keisler's Fubini theorem in the ultraproduct relies only on the ordinary Fubini theorem on the original groups A_i, so there is hope that the theorem could still hold in some form more generally.

A second difficulty arises with the measure space $(\mathbf{A}, \mathcal{A}, \mu)$ itself in proving that the construction is unique. In the finite case, this was a corollary of the Carathéodory extension theorem, which guarantees uniqueness in the σ-finite case. Uniqueness fails in general because the proof relies in an essential way on the Radon-Nikodym theorem, which is true only for σ-finite spaces. Fortunately, this difficulty has been resolved in a paper of Henson [16], where it is proved that the extension $\mathcal{A}(\mathbf{A})$ is in fact unique.

In the infinite setting, one might naturally ask what remains true in the integration theory on the ultraproduct. A detailed study has not been made, but it is easy to construct counterexamples to some of the propositions that are true in the finite case. Additionally, according to [4], it is worth noting that in the infinite case an additional condition must be specified for S-integrability to have any reasonable properties; namely, for a lifting f_i to be S-integrable if $\mu(\mathbf{A}) = \infty$, we must require also that for all $\mathbf{E} = \lim_{i\to\omega} E_i \subset A_i$ in $\mathcal{S}(\mathbf{A})$ such that $\mathrm{st}(\lim_{i\to\omega} f_i) = 0$ on $\mathbf{E}$, we have

$$\mathrm{st}\left(\lim_{i\to\omega} \int_{E_i} f_i\right) = 0.$$

It is easy to check that this condition is automatically satisfied if μ is finite.

The additional difficulties arising from Gowers norms in the infinite case have already been discussed. Finally, it should be pointed out that the proof of existence of conditional expectations relies on the Radon-Nikodym theorem, so conditional expectations are not well-defined in this infinite setting.

In order to construct a satisfactory theory of higher order Fourier analysis on infinite spaces, therefore, the following procedure should be followed. As much of the integration theory on the ultraproduct as possible should be recovered. Everywhere conditional probability is used, projections onto L^2 Hilbert spaces will have to be used instead, which means that in general L^1 functions will not have well-defined projections. Then one could attempt to recover the measure-theoretic results in the new context, tweaking definitions as necessary. It may be useful to first study a test case with a rich analytic structure, such as the group $\mathbb{R}$, especially given the difficulties in manipulating the Gowers norms in completely general contexts.

5.3. **Extension to finitely additive measures.** The especially astute reader will have noticed that, in the construction of the σ-algebra $\mathcal{A}$ on the ultraproduct group $\mathbf{A}$, it was not strictly necessary to start with σ-additive measures on all of the original groups A_i. In fact, it suffices to take *finitely* additive measures μ_i on each A_i, and the construction proceeds entirely as before. The underlying reason is that σ-additivity of μ on $\mathbf{A}$, rather than arising as a result of the σ-additivity of the μ_i, comes "for free" as a result of countable compactness (and therefore as a result of the saturation theorem). Once the additivity of μ has been checked, which requires only the additivity of the μ_i (or, properly, the additivity of ω-many of the μ_i, although this freedom is not particularly helpful or relevant), σ-additivity is automatic.

A (discrete) group is called *amenable* if there is a left-invariant finitely additive probability measure that can be defined on it. Therefore if a sequence of abelian groups are all amenable, then they together with their invariant measures can be used to construct a higher order Fourier-analytic structure. Fortunately, we have the following:

Proposition 5.11. *All abelian groups are amenable.*

Proof (sketch). First one shows that a group is amenable if and only if its finitely generated subgroups are (that is, locally amenable groups are amenable), and that the direct product of finitely many amenable groups is amenable. Then by the structure theorem for finitely generated abelian groups, it suffices to show that cyclic groups and $\mathbb{Z}$ are both amenable. All finite groups are clearly amenable because one can simply take the normalized counting measure.

To show that $\mathbb{Z}$ is amenable, we have to construct a translation-invariant finitely additive probability measure on it. Though the usual proof uses the Hahn-Banach theorem, there is also a proof (found, for example, in [18]) in the spirit of this paper using a nonprincipal ultrafilter, as follows. Fix a nonprincipal ultrafilter ω on $\mathbb{N}$. For each subset $E \subset \mathbb{Z}$ and $i \in \mathbb{N}$, define

$$\phi_i(E) = \frac{|E \cap [-i, i]|}{2i + 1},$$

the partial density of E in $\mathbb{Z}$. Take any real $p \in [0, 1]$ and define the sets

$$N_p^+(E) = \{i \in \mathbb{N} : p > \phi_i(E)\},$$
$$N_p^-(E) = \{i \in \mathbb{N} : p \leq \phi_i(E)\}.$$

Because these sets are complementary, one of them belongs to ω. Furthermore, if $N_p^+(E) \in \omega$ for some p, then certainly $N_q^+(E) \in \omega$ for all $q > p$, and likewise if $N_p^-(E) \in \omega$ for some p, then $N_q^-(E) \in \omega$ for all $q < p$. Therefore the following subsets of $[0, 1]$ form a Dedekind cut in $[0, 1]$:

$$P^+(E) = \{p \in [0, 1] : N_p^+(E) \in \omega\},$$
$$P^-(E) = \{p \in [0, 1] : N_p^-(E) \in \omega\}.$$

Therefore either $P^+(E)$ contains a least element or $P^-(E)$ contains a greatest element; call this element $\phi(A)$. We claim that this function is our translation-invariant finitely additive probability measure. It is certainly finitely additive and we have $\phi(\mathbf{Z}) = 1$; translation invariance follows because for any $E \in \mathbb{Z}$, $n \in \mathbb{Z}$, and real $\epsilon > 0$, we have that

$$|\phi_i(E) - \phi_i(E + n)| \leq \epsilon$$

for all sufficiently large i. □

The example of $\mathbb{Z}$ itself is perhaps the most interesting; however, this extension to finitely additive probability measure via amenability is possibly of limited utility because the method does not work for "ordinary" measures on infinite spaces, like the Haar measure of a locally compact group that is not compact.

Appendix A. Proof of Łoś' theorem and the saturation theorem

Proof of Łoś' theorem. We employ an induction on complexity of formulas. We will be slightly pedantic in recognition of the fact that this form of argument is not well-known outside of mathematical logic. If $\phi(v^1, \dots, v^n)$ is atomic then the theorem is true for $\phi(v^1, \dots, v^n)$ trivially, by the definition of the interpretation of constants, function symbols, and relations in the ultraproduct. It suffices to prove the following three induction steps: if $\phi(v^1, \dots, v^n)$ satisfies Łoś theorem, then $\neg\phi(v^1, \dots, v^n)$ does (negation); if $\phi(v^1, \dots, v^n)$ and $\psi(v^1, \dots, v^n)$ do, then $\phi(v^1, \dots, v^n) \wedge \psi(v^1, \dots, v^n)$ does (conjunction); and if $\phi(v^1, \dots, v^n, x)$ does, then $\exists x \phi(v^1, \dots, v^n, x)$ does (existential quantification).

For negation, we have the following chain of equivalences, assuming that $a^j = \lim_{i \to \omega} a_i^j$ for $1 \leq j \leq n$ and that Łoś' theorem is true for $\phi(v^1, \dots, v^n)$:

$$\begin{aligned} &\mathcal{M} \models \neg\phi(a^1, \dots, a^n) \\ \iff &\text{not } \mathcal{M} \models \phi(a^1, \dots, a^n) \\ \iff &\{i \in \mathbb{N} : \mathcal{M}_i \models \phi(a_i^1, \dots, a_i^n)\} \notin \omega \\ \iff &\{i \in \mathbb{N} : \text{not } \mathcal{M}_i \models \phi(a_i^1, \dots, a_i^n)\} \in \omega \\ \iff &\{i \in \mathbb{N} : \mathcal{M}_i \models \neg\phi(a_i^1, \dots, a_i^n)\} \in \omega. \end{aligned}$$

To go from the third to the fourth line, we used the ultrafilter property that exactly one of an index set and its complement are in ω.

For conjunction, assuming Łoś' theorem is true for $\phi(v^1, \dots, v^n)$ and $\psi(v^1, \dots, v^n)$ separately, we have the following chain of equivalences:

$$\begin{aligned} &\mathcal{M} \models \phi(a^1, \dots, a^n) \wedge \psi(a^1, \dots, a^n) \\ \iff &\mathcal{M} \models \phi(a^1, \dots, a^n) \text{ and } \mathcal{M} \models \psi(a^1, \dots, a^n) \\ \iff &\{i \in \mathbb{N} : \mathcal{M}_i \models \phi(a_i^1, \dots, a_i^n)\} \in \omega \text{ and } \{i \in \mathbb{N} : \mathcal{M}_i \models \psi(a_i^1, \dots, a_i^n)\} \in \omega \\ \iff &\{i \in \mathbb{N} : \mathcal{M}_i \models \phi(a_i^1, \dots, a_i^n)\} \cap \{ i \in \mathbb{N} : \mathcal{M}_i \models \psi(a_i^1, \dots, a_i^n)\} \in \omega \\ \iff &\{i \in \mathbb{N} : \mathcal{M}_i \models \phi(a_i^1, \dots, a_i^n) \text{ and } \mathcal{M}_i \models \psi(a_i^1, \dots, a_i^n)\} \in \omega \\ \iff &\{i \in \mathbb{N} : \mathcal{M}_i \models \phi(a_i^1, \dots, a_i^n) \wedge \psi(a_i^1, \dots, a_i^n)\} \in \omega. \end{aligned}$$

Here the key stage is between the third and fourth line, where we use the fact that $A \cup B \in \omega$ if and only if $A \in \omega$ and $B \in \omega$ (this is a property of every filter, not just ultrafilters).

For existential quantification, assuming Łoś' theorem is true for $\phi(v^1, \dots, v^n, x)$, we have the following chain of equivalences:

$$\begin{aligned} &\mathcal{M} \models \exists x \phi(a^1, \dots, a^n, x) \\ \iff &\text{there exists } b = \lim_{i \to \omega} b_i \in M \text{ such that } \mathcal{M} \models \phi(a^1, \dots, a^n, b) \\ \iff &\text{there exists } b = \lim_{i \to \omega} b_i \in M \text{ such that } \{i \in \mathbb{N} : \mathcal{M}_i \models \phi(a_i^1, \dots, a_i^n, b_i)\} \in \omega. \end{aligned}$$

The last statement certainly implies that $\{i \in \mathbb{N} : \mathcal{M}_i \models \exists x \phi(a_i^1, \dots, a_i^n, x)\} \in \omega$. In the other direction, if this holds, we can pick b_i such that $\mathcal{M}_i \models \phi(a_i^1, \dots, a_i^n, x)$ for each $i \in \omega$, and pick b_i for all other i arbitrarily, so the two statements are equivalent. This completes the induction and the proof. □

We need one easy model-theoretic equivalence to prove the saturation theorem. This allows us to concentrate on 1-types instead of all n-types.

Proposition A.1. *Let κ be an infinite cardinal and $\mathcal{M}$ a . Then the following are equivalent:*

(i) $\mathcal{M}$ is κ-saturated.

(ii) If $A \subset M$ with $|A| < \kappa$ and p is a (possibly incomplete) n-type over A, then p is realized in $\mathcal{M}$.

(iii) If $A \subset M$ with $|A| < \kappa$ and $p \in S_1^{\mathcal{M}}(A)$, then p is realized in $\mathcal{M}$.

Proof. We prove that (i) implies (ii) implies (iii) implies (i).

Statement (i) implies (ii) because for every imcomplete type p over A, there is a complete type p' over A such that $p \subseteq p'$. The fact that p' is realized immediately implies that p is as well.

That (ii) implies (iii) is immediate.

The last implication is by induction on n. The base case $n = 1$ is just (iii). Let $p \in S_n^{\mathcal{M}}(A)$, and let $q \in S_{n-1}^{\mathcal{M}}(\emptyset)$ be the type

$$q = \{\phi(v_1, \dots, v_{n-1}) : \phi \in p\}.$$

By induction, q is realized by some $(n-1)$-tuple $(a_1, \dots, a_{n-1})$ in M^{n-1}. Let $r \in S_1^{\mathcal{M}}(A \cup \{a_1, \dots, a_{n-1}\})$ be the type

$$r = \{\psi(a_1, \dots, a_{n-1}, w) : \psi(v_1, \dots, v_n) \in p\}.$$

By the base case, we can realize r by some $b \in M$. Then by construction the n-tuple $(a_1, \dots, a_{n-1}, b)$ realizes p, so $\mathcal{M}$ is saturated. □

Proof of the saturation theorem. Let $A \subset M$ be countable. For this proof, if $a \in A$, choose a sequence $a_i \in M_i$ such that $a = \lim_{i \to \omega} a_i$.

By the above proposition, it suffices to show that every 1-type is satisfied in $\mathcal{M}$. Because the language $\mathcal{L}$ is countable, the total number of formulas in the language is countable, so certainly every 1-type is countable. Therefore take and enumerate a 1-type $p = \{\phi_n(v) : n \in \mathbb{N}\}$, a set of $\mathcal{L}_A$ formulas such that $p \cup \mathrm{Th}_A(\mathcal{M})$ is satisfiable. By taking conjunctions, it suffices to assume without loss of generality that $\phi_{n+1}(v) \implies \phi_n(v)$ for all n. Let $\phi_n(v)$ be $\theta_i(v, a^{n,1}, \dots, a^{n,m_n})$, where each θ_i is an $\mathcal{L}$-formula (we are just pulling out all of the constants not in the original language).

Let

$$D_n = \{i \in \mathbb{N} : \mathcal{M}_i \models \exists v \theta_i(v, a_i^{n,1}, \dots, a_i^{n,m_n})\}.$$

Since $p \cup \mathrm{Th}_A(\mathcal{M})$ is satisfiable, certainly so is the existence statement $\exists v \phi_n(v) \cup \mathrm{Th}_A(\mathcal{M})$. This implies that $\exists v \phi_n(v) \in \mathrm{Th}_A(\mathcal{M})$, because the full theory of a model is complete. Therefore

$$\mathcal{M} \models \exists v \phi_n(v) \implies \mathcal{M} \models \exists v \theta_n(v, a^{n,1}, \dots, a^{n,m_n}),$$

which implies by Łoś' theorem that $D_n \in \omega$.

Define sets $J_n = \{i \in \mathbb{N} : n \leq i\}$. Each must be a member of the ultrafilter, since each is a cofinite set: otherwise, a finite set would be in the ultrafilter, contrary to the assumption that ω is nonprincipal. Therefore if we define

$$X_n = J_n \cap D_n,$$

then $X_n \in \omega$ as well. Because $\phi_{n+1}(v) \implies \phi_n(v)$, the D_n are a descending chain, and the J_n trivially are as well. The intersection of all of the J_n is the empty set. Therefore the X_i

also form a descending chain whose intersection is the empty set. Thus for each i, there is a single greatest number n_i such that $i \in X_{n_i}$ (if no such element exists, we just set $n_i = 0$). If $n_i = 0$, define b_i to to be an arbitrary element of M_i. Otherwise, let b_i be such that

$$\mathcal{M}_i \models \phi_{n_i}(b_i, a_i^{n,1}, \ldots, a_i^{n,m_n}).$$

Such an element clearly exists because the model makes true the corresponding existence statement. Then by construction, whenever $n > 0$ and $i \in X_n$, by the definition of n_i we have $n \leq n_i$, so using that $\phi_{m+1}(v) \implies \phi_m(v)$ several times we get

$$\mathcal{M}_i \models \phi_n(b_i, a_i^{n,1}, \ldots, a_i^{n,m_n}).$$

Except for $n = 0$, this holds whenever $i \in X_n$; that is, whenever $n \leq i$ and $i \in D_n$.

By Łoś' theorem, since $X_n \in \omega$, for each n we can conclude that

$$\mathcal{M} \models \phi_n(b),$$

where $b = \lim_{i \to \omega} b_i$. Note that the possible problems at $n = 0$ do not affect anything, since the ultraproduct is insensitive to any particular coordinate. Therefore we have found a b that realizes all of p, so we are done. □

References

[1] P. Candela, *Uniformity norms and nilsequences (an introduction)*, notes accessible at `http://www.dpmms.cam.ac.uk/~pc308/UNN-Notes.pdf`.

[2] C. C. Chang and H. J. Keisler, *Model Theory*, Studies in Logic and the Foundations of Mathematics, vol. 73, North-Holland, Amsterdam, 1977.

[3] K. Ciesielski, H. Fejzić, and C. Freiling, *Measure zero sets with non-measurable sum*, Real Anal. Exchange, 27:2, 783-794, 2001.

[4] N. J. Cutland, *Loeb Measures in Practice: Recent Advances*, Lecture Notes in Mathematics, vol. 1751, Springer-Verlag, Berlin, 2001.

[5] N. J. Cutland, *Nonstandard measure theory and its applications*, Bull. London Math. Soc., 15, 529-589, 1983.

[6] N. J. Cutland, V. Nneves, F. Oliveira, and J. Sousa-Pinto, eds., *Developments in Nonstandard Mathematics*, Pitman Research Notes in Mathematics Series, vol. 336, Longman, Harlow, 1995.

[7] G. Elek and B. Szegedy, *Limits of hypergraphs, removal and regularity lemmas: a non-standard approach*, preprint, accessible at `http://arxiv.org/abs/0705.2179`.

[8] G. Elek and B. Szegedy, *A measure-theoretic approach to the theory of dense hypergraphs*, preprint, accessible at `http://arxiv.org/abs/0810.4062`.

[9] D. H. Fremlin, *Measure Theory: Volume 2, 2nd edition*, Torres Fremlin, 2010.

[10] H. Furstenberg and B. Weiss, *A mean ergodic theorem for* $\frac{1}{N}\sum n = 1^N f(T^n x) g(T^{n^2} x)$, in Convergence in Ergodic Theory and Probability, Ohio State Univ. Math. Res. Inst. Publ., vol. 5, Walter de Gruyter & Co., Berlin, 1996.

[11] E. Glasner, *Ergodic Theory via Joinings*, Mathematical Surveys and Monographs, vol. 101, Amer. Math. Soc., 2003.

[12] W. T. Gowers, *Fourier analysis and Szemerédi's theorem*, in Proceedings of the International Congress of Mathematicians, vol. I, Berlin, 1998.

[13] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal., 11, 465-588, 2001.

[14] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. Math., 167:2, 481-547, 2008.

[15] P. R. Halmos, *Measure Theory*, Graduate Texts in Mathematics, vol. 18, Springer-Verlag, New York, 1974.

[16] C. W. Henson, *Unbouded Loeb measures*, Proc. Amer. Math. Soc., 74, 143-150, 1979.

[17] B. Host and B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. Math., 161:2, 397-488, 2005.

[18] V. G. Kanoveĭ, *Borel Equivalence Relations: Structure and Classification*, University Lectures Series, vol. 44, Amer. Math. Soc., 2008.

[19] H. J. Keisler, *An Infinitesimal Approach to Stochastic Analysis*, Mem. Amer. Math. Soc., 1984.

[20] T. L. Lindström, *Hyperfinite stochastic integration I: The nonstandard theory*, Math. Scand., 46, 265-292, 1980.

[21] P. A. Loeb, *Conversion from nonstandard to standard measure spaces and applications in probability theory*, Trans. Amer. Math. Soc., 211, 113-122, 1975.

[22] D. Marker, *Model Theory: An Introduction*, Graduate Texts in Mathematics, vol. 217, Springer-Verlag, New York, 2000.

[23] D. Ramakrishnan and R. J. Valenza, *Fourier Analysis on Number Fields*, Graduate Texts in Mathematics, vol. 186, Springer-Verlag, New York, 1999.

[24] A. Robinson, *Non-standard analysis*, Princeton University Press, 1996.

[25] E. M. Stein and R. Shakarchi, *Real analysis: measure theory, integration, & Hilbert spaces*, Princeton University Press, 2005.

[26] C. Swartz, *Measure, integration, and function spaces*, World Scientific, 1994.

[27] B. Szegedy, *Higher order Fourier analysis as an algebraic theory I*, preprint, accessible at `http://arxiv.org/abs/1001.4282`.

[28] B. Szegedy, *Higher order Fourier analysis as an algebraic theory II*, preprint, accessible at `http://arxiv.org/abs/0911.1157`.

[29] B. Szegedy, *Higher order Fourier analysis as an algebraic theory III*, preprint, accessible at `http://arxiv.org/abs/0903.0897`.

[30] B. Szegedy, *Gowers norms, regularization and limits of functions on abelian groups*, preprint, accessible at `http://arxiv.org/abs/1010.6211`.

[31] B. Szegedy, *On higher order Fourier analysis*, preprint, accessible at `http://arxiv.org/abs/1203.2260`.

[32] J. T. Tate, *Fourier analysis in number fields and Hecke's zeta-functions*, Ph.D. thesis, Princeton, 1950.

[33] T. Tao, *Higher order Fourier analysis*, preprint, accessible at `http://terrytao.wordpress.com/books/higher-order-fourier-analysis/`.

[34] T. Tao, *Ultraproducts as a bridge between hard analysis and soft analysis*, notes accessible at `http://terrytao.wordpress.com/2011/10/15/254a-notes-6-ultraproducts-as-a-bridge-between-hard-analysis-and-soft-analysis/`.