A quantitative ergodic theory proof of Szemerédi's theorem

Terence Tao
Department of Mathematics, UCLA, Los Angeles CA 90095-1555
tao@math.ucla.edu
http://www.math.ucla.edu/∼tao

Submitted: May 14, 2004; Accepted: Oct 30, 2006; Published: Nov 6, 2006
Mathematics Subject Classification: 11B25, 37A45

Abstract

A famous theorem of Szemerédi asserts that given any density $0 < \delta \leq 1$ and any integer $k \geq 3$, any set of integers with density δ will contain infinitely many proper arithmetic progressions of length k. For general k there are essentially four known proofs of this fact; Szemerédi's original combinatorial proof using the Szemerédi regularity lemma and van der Waerden's theorem, Furstenberg's proof using ergodic theory, Gowers' proof using Fourier analysis and the inverse theory of additive combinatorics, and the more recent proofs of Gowers and Rödl-Skokan using a hypergraph regularity lemma. Of these four, the ergodic theory proof is arguably the shortest, but also the least elementary, requiring passage (via the Furstenberg correspondence principle) to an infinitary measure preserving system, and then decomposing a general ergodic system relative to a tower of compact extensions. Here we present a quantitative, self-contained version of this ergodic theory proof, and which is "elementary" in the sense that it does not require the axiom of choice, the use of infinite sets or measures, or the use of the Fourier transform or inverse theorems from additive combinatorics. It also gives explicit (but extremely poor) quantitative bounds.

1 Introduction

A famous theorem of van der Waerden [44] in 1927 states the following.

Theorem 1.1 (Van der Waerden's theorem). *[44] For any integers $k, m \geq 1$ there exists an integer $N = N_{vdW}(k, m) \geq 1$ such that every colouring $\mathbf{c} : \{1, \ldots, N\} \to \{1, \ldots, m\}$ of $\{1, \ldots, N\}$ into m colours contains at least one monochromatic arithmetic progression of length k (i.e. a progression in $\{1, \ldots, N\}$ of cardinality k on which $\mathbf{c}$ is constant).*

See for instance [22] for the standard "colour focusing" proof; another proof can be found in [36]. This theorem was then generalized substantially in 1975 by Szemerédi [39] (building upon earlier work in [33], [38]), answering a question of Erdős and Turán [8], as follows:

Theorem 1.2 (Szemerédi's theorem). *For any integer $k \geq 1$ and real number $0 < \delta \leq 1$,*
there exists an integer $N_{SZ}(k, \delta) \geq 1$ such that for every $N \geq N_{SZ}(k, \delta)$, every set $A \subset \{1, \ldots, N\}$ of cardinality $|A| \geq \delta N$ contains at least one arithmetic progression of length k.

It is easy to deduce Van der Waerden's theorem from Szemerédi's theorem (with $N_{\mathrm{vdW}}(k, m) := N_{\mathrm{SZ}}(k, \frac{1}{m})$) by means of the pigeonhole principle. The converse implication is however substantially less trivial.

There are many proofs already known for Szemerédi's theorem, which we discuss below; the main purpose of this paper is present yet another such proof. This may seem somewhat redundant, but we will explain our motivation for providing another proof later in this introduction.

Remarkably, while Szemerédi's theorem appears to be solely concerned with arithmetic combinatorics, it has spurred much further research in other areas such as graph theory, ergodic theory, Fourier analysis, and number theory; for instance it was a key ingredient in the recent result [23] that the primes contain arbitrarily long arithmetic progressions. Despite the variety of proofs now available for this theorem, however, it is still regarded as a very difficult result, except when k is small. The cases $k = 1, 2$ are trivial, and the case $k = 3$ is by now relatively well understood (see [33], [11], [35], [37], [6], [25], [7] for a variety of proofs). The case $k = 4$ also has a number of fairly straightforward proofs (see [38], [34], [19], [9]), although already the arguments here are more sophisticated than for the $k = 3$ case. However for the case of higher k, only four types of proofs are currently known, all of which are rather deep. The original proof of Szemerédi [39] is highly combinatorial, relying on van der Waerden's theorem (Theorem 1.1) and the famous Szemerédi regularity lemma (which itself has found many other applications, see [27] for a survey); it does provide an upper bound on $N_{\mathrm{SZ}}(k, \delta)$ but it is rather poor (of Ackermann type), due mainly to the reliance on the van der Waerden theorem and the regularity lemma, both of which have notoriously bad dependence of the constants. Shortly afterwards, Furstenberg [10] (see also [15], [11]) introduced what appeared to be a completely different argument, transferring the problem into one of recurrence in ergodic theory, and solving that problem by a number of ergodic theory techniques, notably the introduction of a Furstenberg tower of compact extensions (which plays a role analogous to that of the regularity lemma). This ergodic theory argument is the shortest and most flexible of all the known proofs, and has been the most successful at leading to further generalizations of Szemerédi's theorem (see for instance [3], [5], [12], [13], [14]). On the other hand, the infinitary nature of the argument means that it does not obviously provide any effective bounds for the quantity $N_{\mathrm{SZ}}(k, \delta)$. The third proof is more recent, and is due to Gowers [20] (extending earlier arguments in [33], [19] for small k). It is based on combinatorics, Fourier analysis, and inverse arithmetic combinatorics (in particular multilinear versions of Freiman's theorem and the Balog-Szemerédi theorem). It gives far better bounds on $N_{\mathrm{SZ}}(k, \delta)$ (essentially of double exponential growth in δ rather than Ackermann or iterated tower growth), but also requires far more analytic machinery and quantitative estimates. Finally, very recent arguments of Gowers [21] and Rödl, Skokan, Nagle, Tengan, Tokushige, and Schacht [29], [30], [31], [28], relying primarily on a hypergraph version of the Szemerédi regularity lemma, have been discovered; these arguments are somewhat similar in spirit to Szemerédi's original proof (as well as the proofs in [35], [37] in the $k = 3$ case and [9] in the $k = 4$ case) but is conceptually somewhat more straightforward (once one accepts the need to work with hypergraphs instead of graphs, which does unfortunately introduce a number of additional technicalities). Also these arguments can handle certain higher dimensional extensions of Szemerédi's theorem first obtained by ergodic theory methods in [12].

As the above discussion shows, the known proofs of Szemerédi's theorem are extremely diverse. However, they do share a number of common themes, principal among which is the establishment of a *dichotomy between randomness and structure*. Indeed, in an extremely abstract and heuristic sense, one can describe all the known proofs of Szemerédi's theorem collectively as follows. Start with the set A (or some other object which is a proxy for A, e.g. a graph, a hypergraph, or a measure-preserving system). For the object under consideration, define some concept of *randomness* (e.g. ε-regularity, uniformity, small Fourier coefficients, or weak mixing), and some concept of *structure* (e.g. a nested sequence of arithmetically structured sets such as progressions or Bohr sets, or a partition of a vertex set into a controlled number of pieces, a collection of large Fourier coefficients, a sequence of almost periodic functions, a tower of compact extensions of the trivial factors, or a $k-2$-step nilfactor). Obtain some sort of *structure theorem* that splits the object into a structured component, plus an error which is random relative to that structured component. To prove Szemerédi's theorem (or a variant thereof), one then needs to obtain some sort of *generalized von Neumann theorem* to eliminate the random error, and then some sort of *structured recurrence theorem* for the structured component.

Obviously there is a great deal of flexibility in executing the above abstract scheme, and this explains the large number of variations between the known proofs of Szemerédi type theorems. Also, each of the known proofs finds some parts of the above scheme more difficult than others. For instance, Furstenberg's ergodic theory argument requires some non-elementary machinery to set up the appropriate proxy for A, namely a measure-preserving probability system, and the structured recurrence theorem (which is in this case a recurrence theorem for a tower of compact extensions) is also somewhat technical. In the Fourier-analytic arguments of Roth and Gowers, the structured component is simply a nested sequence of long arithmetic progressions, which makes the relevant recurrence theorem a triviality; instead, almost all the difficulty resides in the structure theorem, or more precisely in enforcing the assertion that lack of uniformity implies a density increment on a smaller progression. The more recent hypergraph arguments of Gowers and Rödl-Skokan-Nagel-Schacht are more balanced, with no particular step being exceptionally more difficult than any other, although the fact that hypergraphs are involved does induce a certain level of notational and technical complexity throughout.

Finally, Szemerédi's original argument contains significant portions (notably the use of the Szemerédi regularity lemma, and the use of density increments) which fit very nicely into the above scheme, but also contains some additional combinatorial arguments to connect the various steps of the proof together.

In this paper we present a new proof of Szemerédi's theorem (Theorem 1.2) which implements the above scheme in a reasonably elementary and straightforward manner. This new proof can best be described as a "finitary" or "quantitative" version of the ergodic theory proofs of Furstenberg [10], [15], in which one stays entirely in the realm of finite sets (as opposed to passing to an infinite limit in the ergodic theory setting). As such, the axiom of choice is not used, and an explicit bound for $N_{\mathrm{SZ}}(k,\delta)$ is in principle possible[1] (although the bound is extremely poor, perhaps even worse than Ackermann growth, and certainly worse than the bounds obtained by Gowers [20]). We also borrow some tricks and concepts from other proofs; in particular from the proof of the Szemerédi regularity lemma we borrow the L^2 incrementation trick in order to obtain a structure theorem with effective bounds, while from the arguments of Gowers [20] we borrow the Gowers uniformity norms U^{k-1} to quantify the concept of randomness. One of our main innovations is to complement these norms with the (partially dual) *uniform almost periodicity norms* UAP^{k-2} to quantify the concept of an *uniformly almost periodic function of order* $k-2$. This concept will be defined rigorously later, but suffice to say for now that a model example of a uniformly almost periodic function of order $k-2$ is a finite polynomial-trigonometric sum $f : \mathbf{Z}_N \to \mathbf{C}$ of the form[2]

$$F(x) := \frac{1}{J}\sum_{j=1}^{J} c_j e(P_j(x)/N) \text{ for all } x \in \mathbf{Z}_N, \tag{1}$$

where $\mathbf{Z}_N := \mathbf{Z}/N\mathbf{Z}$ is the cyclic group of order N, $J \geq 1$ is an integer, the c_j are complex numbers bounded in magnitude by 1, $e(x) := e^{2\pi i x}$, and the P_j are polynomials of degree at most $k-2$ and with coefficients in $\mathbf{Z}_N$. The uniform almost periodicity norms serve to quantify how closely a function behaves like (1), and enjoy a number of

---

[1]It may also be possible in principle to extract some bound for $N_{\mathrm{SZ}}(k,\delta)$ directly from the original Furstenberg argument via proof theory, using such tools as Herbrand's theorem; see for instance [17] where a similar idea is applied to the Furstenberg-Weiss proof of van der Waerden's theorem to extract Ackermann-type bounds from what is apparently a nonquantitative argument. However, to the author's knowledge this program has not been carried out previously in the literature for the ergodic theory proof of Szemerédi proof. Also we incorporate some other arguments in order to simplify the proof and highlight some new concepts (such as a new Banach algebra of uniformly almost periodic functions).

[2]Actually, these functions are a somewhat special class of uniformly almost periodic functions of order $k-2$, which one might dub the *quasiperiodic functions of order* $k-2$. The relationship between the two seems very closely related to the distinction in ergodic theory between $k-2$-step nilsystems and systems which contain polynomial eigenfunctions of order $k-2$; see [16], [26] for further discussion of this issue. It is also closely related to the rather vaguely defined issue of distinguishing "almost polynomial" or "almost multilinear" functions from "genuinely polynomial" or "genuinely multilinear" functions, a theme which recurs in the work of Gowers [19], [20], and also in the theorems of Freiman and Balog-Szemerédi from inverse additive combinatorics which were used in Gowers' work. It seems of interest to quantify and pursue these issues further.

pleasant properties, most notably that they form a Banach algebra; indeed one can think of these norms as a higher order variant of the classical Wiener algebra of functions with absolutely convergent Fourier series.

The argument is essentially self-contained, aside from some basic facts such as the Weierstrass approximation theorem; the only main external ingredient needed is van der Waerden's theorem (to obtain the recurrence theorem for uniformly almost periodic functions), which is standard. As such, we do not require any familiarity with any of the other proofs of Szemerédi's theorem, although we will of course discuss the relationship between this proof and the other proofs extensively in our remarks. In particular we do not use the Fourier transform, or theorems from inverse arithmetic combinatorics such as Freiman's theorem or the Balog-Szemerédi theorem, and we do not explicitly use the Szemerédi regularity lemma either for graphs or hypergraphs (although the proof of that lemma has some parallels with certain parts of our argument here). Also, while we do use the language of ergodic, measure, and probability theory, in particular using the concept of conditional expectation with respect to a factor, we do so entirely in the context of finite sets such as $\mathbf{Z}_N$; as such, a factor (or σ-algebra) is nothing more than a finite partition of $\mathbf{Z}_N$ into "atoms", and conditional expectation is merely the act of averaging a function on each atom[3]. As such, we do not need such results from measure theory as the construction of product measure (or conditional product measure, via Rohlin's lemma [32]), which plays an important part of the ergodic theory proof, notably in obtaining the structure and recurrence theorems. Also, we do not use the compactness of Hilbert-Schmidt or Volterra integral operators directly (which is another key ingredient in Furstenberg's structure theorem), although we will still need a quantitative finite-dimensional version of this fact (see Lemmas 9.3, 10.2 below). Because of this, our argument could technically be called "elementary". However we will need a certain amount of structural notation (of a somewhat combinatorial nature) in order to compensate for the lack of an existing body of notation such as is provided by the language of ergodic theory.

In writing this paper we encountered a certain trade-off between keeping the paper brief, and keeping the paper well-motivated. We have opted primarily for the latter; if one chose to strip away all the motivation and redundant arguments from this paper one could in fact present a fairly brief proof of Theorem 1.2 (roughly 20 pages in length); see [42]. We also had a similar trade-off between keeping the arguments simple, and attempting to optimize the growth of constants for $N_{\mathrm{SZ}}(k, \delta)$ (which by the arguments here could be as bad as double-Ackermann or even triple-Ackermann growth); since it seems clear that the arguments here have no chance whatsoever to be competitive with the bounds obtained by Gowers' Fourier-analytic proof [20] we have opted strongly in favour of the former.

Remark 1.3. Because our argument uses similar ingredients to the ergodic theory arguments, but in a quantitative finitary setting, it seems likely that one could modify these arguments relatively easily to obtain quantitative finitary versions of other ergodic theory recurrence results in the literature, such as those in [12], [13], [14], [3], [5]. In many of these cases, the ordinary van der Waerden theorem would have to be replaced by a more

[3]Readers familiar with the Szemerédi regularity lemma may see parallels here with the proof of that lemma. Indeed one can phrase the proof of this lemma in terms of conditional expectation; see [41].

general result, but fortunately such generalizations are known to exist (see e.g. [46] for further discussion). In principle, the quantitative ergodic approach could in fact have a greater reach than the traditional ergodic approach to these problems; for instance, the recent establishment in [23] that the primes contained arbitrarily long arithmetic progressions relied heavily on this quantitative ergodic point of view, and does not seem at this point to have a proof by traditional ergodic methods (or indeed by any of the other methods available for proving Szemerédi's theorem, although the recent hypergraph approach of Gowers [21] and of Rödl-Skokan-Nagle-Schacht [28], [29], [30] seems to have a decent chance of being "relativized" to pseudorandom sets such as the "almost primes"; see [43]). Indeed, some of the work used to develop this paper became incorporated into [23], and conversely some of the progress developed in [23] was needed to conclude this paper.

Remark 1.4. It is certainly possible to avoid using van der Waerden's theorem explicitly in our arguments, for instance by incorporating arguments similar to those used in the proof of this theorem into the main argument[4]. A decreased reliance on van der Waerden's theorem would almost certainly lead to better bounds for $N_{\mathrm{SZ}}(k, \delta)$, for instance the Fourier-analytic arguments of Gowers [19], [20] avoids this theorem completely and obtains bounds for $N_{\mathrm{SZ}}(k, \delta)$ which are far better than that obtained by any other argument, including ours. However this would introduce additional arguments into our proof which more properly belong to the Ramsey-theoretic circle of ideas surrounding van der Waerden's theorem, and so we have elected to proceed by the simpler and "purer" route of using van der Waerden's theorem directly. Also, as remarked above, the argument as presented here seems more able to extend to other recurrence problems.

Remark 1.5. Our proof of Szemerédi's theorem here is similar in spirit to the proof of the transference principle developed in [23] by Ben Green and the author which allowed one to deduce a Szemerédi theorem relative to a pseudorandom measure from the usual formulation of Szemerédi's theorem; this transference principle also follows the same basic scheme used to prove Szemerédi's theorem (with Szemerédi's theorem itself taking on the role of the structured recurrence theorem). Indeed, the two arguments were developed concurrently (and both were inspired, not only by each other, but by all four of the existing proofs of Szemerédi's theorem in the literature, as well as arguments from the much better understood $k = 3, 4$ cases); it may also be able to combine the two to give a more direct proof of Szemerédi's theorem relative to a pseudorandom measure. There are two main differences however between our arguments here and those in [23]. Firstly, in the arguments here no pseudorandom measure is present. Secondly, the role of structure in [23] was played by the *anti-uniform functions*, or more precisely a tower of factors constructed out of basic anti-uniform functions. Our approach uses the same concept,

[4] This is to some extent done for instance in Furstenberg's original proof [10], [15]. A key component of that proof was to show that the multiple recurrence property was preserved under compact extensions. Although it is not made explicit in those papers, the argument proceeds by "colouring" elements of the extension on each fiber, and using "colour focusing" arguments closely related to those used to prove van der Waerden's theorem. The relevance of van der Waerden's theorem and its generalizations in the ergodic theory approach is made more explicit in later papers, see e.g. [16], [3], [5], and also the discussion in [46]

but goes further by analyzing the basic anti-uniform functions more carefully, and in fact concluding that such functions are uniformly almost periodic[5] of a certain order $k-2$.

Acknowledgements. This work would not have been possible without the benefit of many discussions with Hillel Furstenberg, Ben Green, Timothy Gowers, Bryna Kra, and Roman Sasyk for for explaining both the techniques and the intuition behind the various proofs of Szemerédi's theorem and related results in the literature, and for drawing the author's attention to various simplifications in these arguments. Many of the ideas here were also developed during the author's collaboration with Ben Green, and we are particularly indebted to him for his suggestion of using conditional expectations and an energy increment argument to prove quantitative Szemerédi-type theorems. We also thank Van Vu for much encouragement throughout this project, Mathias Schacht for some help with the references, and Alex Kontorovich for many helpful corrections. The author also thanks Australian National University and Edinburgh University for their hospitality where this work was conducted. The author is a Clay Prize Fellow and is supported by a grant from the Packard Foundation.

2 The finite cyclic group setting

We now begin our new proof of Theorem 1.2. Following the abstract scheme outlined in the introduction, we should begin by specifying what objects we shall use as proxies for the set A. The answer shall be that we shall use non-negative bounded functions $f : \mathbf{Z}_N \to \mathbf{R}^+$ on a cyclic group $\mathbf{Z}_N := \mathbf{Z}/N\mathbf{Z}$. In this section we set out some basic notation for such functions, and reduce Theorem 1.2 to proving a certain quantitative recurrence property for these functions.

Remark 2.1. The above choice of object of study fits well with the Fourier-based proofs of Szemerédi's theorem in [33], [34], [19], [20], at least for the initial stages of the argument. However in those arguments one eventually passes from $\mathbf{Z}_N$ to a smaller cyclic group $\mathbf{Z}_{N'}$ for which one has located a density increment, iterating this process until randomness has been obtained (or the density becomes so high that finding arithmetic progressions becomes very easy). In contrast, we shall keep N fixed and use the group $\mathbf{Z}_N$ throughout the argument; it will be a certain family of factors which changes instead. This parallels the ergodic theory argument [10], [15], [11], but also certain variants of the Fourier argument such as [6], [7]. It also fits well with the philosophy of proof of the Szemerédi regularity lemma.

We now set up some notation. We fix a large prime number N, and fix $\mathbf{Z}_N := \mathbf{Z}/N\mathbf{Z}$ to be the cyclic group of order N. We will assume that N is extremely large; basically, it will be larger than any quantity depending on any of the other parameters which

[5] In [23] the only facts required concerning these basic anti-uniform functions were that they were bounded, and that pseudorandom measures were uniformly distributed with respect to any factor generated by such functions. This was basically because the argument in [23] invoked Szemerédi's theorem as a "black box" to deal with this anti-uniform component, whereas clearly this is not an option for our current argument.

appear in the proof. We will write $O(X)$ for a quantity bounded in magnitude by CX where C is independent of N; if C depends on some other parameters (e.g. k and δ), we shall subscript the $O(X)$ notation accordingly (e.g. $O_{k,\delta}(X)$) to indicate the dependence. Generally speaking we will order these subscripts so that the extremely large or extremely small parameters are at the right. We also write $X \ll Y$ or $Y \gg X$ as synonymous with $X = O(Y)$, again denoting additional dependencies in the implied constant C by subscripts (e.g. $X \ll_{k,\delta} Y$ means that $|X| \leq C(k,\delta)Y$ for some $C(k,\delta)$ depending only on k, δ).

Definition 2.2. If $f : X \to \mathbf{C}$ is a function[6], and A is a finite non-empty subset of X, we define the *expectation of f conditioning on A*[7]

$$\mathbf{E}_A f = \mathbf{E}_{x\in A} f(x) := \frac{1}{|A|}\sum_{x\in A} f(x)$$

where $|A|$ of course denotes the cardinality of A. If in particular f is an indicator function $f = \mathbf{1}_\Omega$ for some $\Omega \subseteq X$, thus $f(x) = 1$ when $x \in \Omega$ and $f(x) = 0$ otherwise, we write

$$\mathbf{P}_A(\Omega) := \mathbf{E}_A \mathbf{1}_\Omega = |\Omega|/|A|.$$

Similarly, if $P(x)$ is an event depending on x, we write

$$\mathbf{P}_A(P) := \mathbf{E}_A \mathbf{1}_P = \frac{1}{|A|}\{x \in A : P(x) \text{ is true}\},$$

where $\mathbf{1}_{P(x)} = 1$ when $P(x)$ is true and $\mathbf{1}_{P(x)} := 0$ otherwise.

We also adopt the following ergodic theory notation: if $f : \mathbf{Z}_N \to \mathbf{R}$ is a function, we define the integral

$$\int_{\mathbf{Z}_N} f = \mathbf{E}_{\mathbf{Z}_N} f = \frac{1}{N}\sum_{x=1}^{N} f(x)$$

and the shifts $T^n f : \mathbf{Z}_N \to R$ for any $n \in \mathbf{Z}_N$ or $n \in \mathbf{Z}$ by

$$T^n f(x) := f(x-n),$$

[6]Strictly speaking, we could give the entire proof of Theorem 1.2 using only real-valued functions rather than complex-valued, as is done in the ergodic theory proofs, thus making the proof slightly more elementary and also allowing for some minor simplifications in the notation and arguments. However, allowing the functions to be complex valued allows us to draw more parallels with Fourier analysis, and in particular to discuss such interesting examples of functions as (1).

[7]We have deliberately chosen this notation to coincide with the usual notations of probability $P(\Omega)$ and expectation $\mathbf{E}(f)$ for random variables to emphasize the probabilistic nature of many of our arguments, and indeed we will also combine this notation with the probabilistic one (and take advantage of the fact that both forms of expectation commute with each other). Note that one can think of $\mathbf{E}_{x\in A} f(x) = \mathbf{E}_A f$ as the conditional expectation of $f(x)$, where x is a random variable with the uniform distribution on X, conditioning on the event $x \in A$.

and similarly define $T^n\Omega$ for any $\Omega \subset \mathbf{Z}_N$ by $T^n\Omega := \Omega + n$, thus $T^n \mathbf{1}_\Omega = \mathbf{1}_{T^n\Omega}$. Clearly these maps are algebra homomorphisms (thus $T^n(fg) = (T^n f)(T^n g)$ and $T^n(f+g) = T^n f + T^n g$), preserve constant functions, and are integral-preserving (thus $\int_{\mathbf{Z}_N} T^n f = \int_{\mathbf{Z}_N} f$). They also form a group, thus $T^{n+m} = T^n T^m$ and T^0 is the identity, and are unitary with respect to the usual inner product $\langle f, g \rangle := \int_{\mathbf{Z}_N} f\overline{g}$. We shall also rely frequently[8] on the Banach algebra norm

$$\|f\|_{L^\infty} := \sup_{x \in \mathbf{Z}_N} |f(x)|$$

and the Hilbert space structure

$$\langle f, g \rangle := \int_{\mathbf{Z}_N} f\overline{g}; \quad \|f\|_{L^2} := \langle f, f \rangle^{1/2} = (\int_{\mathbf{Z}_N} |f|^2)^{1/2};$$

later on we shall also introduce a number of other useful norms, in particular the *Gowers uniformity norms* U^{k-1} and the *uniform almost periodicity norms* UAP^{k-2}.

To prove Theorem 1.2, it will suffice to prove the following quantitative recurrence version of that theorem.

Definition 2.3. A function $f : \mathbf{Z}_N \to \mathbf{C}$ is said to be *bounded* if we have $\|f\|_{L^\infty} \leq 1$.

Theorem 2.4 (Quantitative recurrence form of Szemerédi's theorem). *For any integer $k \geq 1$, any large prime integer $N \geq 1$, any $0 < \delta \leq 1$, and any non-negative bounded function $f : \mathbf{Z}_N \to \mathbf{R}^+$ with*

$$\int_{\mathbf{Z}_N} f \geq \delta \tag{2}$$

we have

$$\mathbf{E}_{r \in \mathbf{Z}_N} \int_{\mathbf{Z}_N} \prod_{j=0}^{k-1} T^{jr} f \gg_{k,\delta} 1. \tag{3}$$

Remark 2.5. This is the form of Szemerédi's theorem required in [23]. This result was then generalized in [23] (introducing a small error $o_{k,\delta}(1)$) by replacing the hypothesis that f was bounded by the more general hypothesis that f was pointwise dominated by a *pseudorandom measure*. This generalization was crucial to obtain arbitrarily long progressions in the primes. We will not seek such generalizations here, although we do remark that the arguments in [23] closely parallel to the ones here.

We now show how the above theorem implies Theorem 1.2.

[8] Of course, since the space of functions on $\mathbf{Z}_N$ is finite-dimensional, all norms are equivalent up to factors depending on N. However in line with our philosophy that we only wish to consider quantities which are bounded uniformly in N, we think of these norms as being genuinely distinct.

Proof of Theorem 1.2 assuming Theorem 2.4. Fix k, δ. Let $N \geq 1$ be large, and suppose that $A \subset \{1, \ldots, N\}$ has cardinality $|A| \geq \delta N$. By Bertrand's postulate, we can find a large prime number N' between kN and $2kN$. We embed $\{1, \ldots, N\}$ in $\mathbf{Z}_{N'}$ in the usual manner, and let A' be the image of A under this embedding. Then we have $\int_{\mathbf{Z}_N} \mathbf{1}_{A'} \geq \delta/2k$, and hence by (3)

$$\mathbf{E}_{r \in \mathbf{Z}_{N'}} \int_{\mathbf{Z}_{N'}} \prod_{j=0}^{k-1} T^{jr} \mathbf{1}_{A'} \gg_{k,\delta} 1,$$

or equivalently

$$|\{(x, r) \in \mathbf{Z}_{N'} : x, x-r, \ldots, x-(k-1)r \in A'\}| \gg_{k,\delta} (N')^2.$$

Since $N' \geq kN$ and $A' \subset \{1, \ldots, N\}$, we see that $1 \leq x \leq N$ and $-N \leq r \leq N$ in the above set. Also we may remove the $r = 0$ component of this set since this contributes at most N to the above sum. If N is large enough, the right-hand side is still positive, and this implies that A contains a progression $x, x-r, \ldots, x-(k-1)r$, as desired. $\square$

Remark 2.6. One can easily reverse this implication and deduce Theorem 2.4 from Theorem 1.2; the relevant argument was first worked out by Varnavides [45]. In the ergodic theory proofs, Szemerédi's theorem is also stated in a form similar to (3), but with $\mathbf{Z}_N$ replaced by an arbitrary measure-preserving system (and r averaged over some interval $\{1, \ldots, N\}$ going to infinity), and the left-hand side was then shown to have positive limit inferior, rather than being bounded from below by some explicit constant. However these changes are minor, and again it is easy to pass from one statement to the other, at least with the aid of the axiom of choice (see [15], [4] for some further discussion on this issue).

It remains to deduce Theorem 2.4. This task shall occupy the remainder of the paper.

3 Overview of proof

We shall begin by presenting the high-level proof of Theorem 2.4, implementing the abstract scheme outlined in the introduction.

One of the first tasks is to define measures of randomness and structure in the function f. We shall do this by means of two families of norms[9]: the *Gowers uniformity norms*

$$\|f\|_{U^0} \leq \|f\|_{U^1} \leq \ldots \leq \|f\|_{U^{k-1}} \leq \ldots \leq \|f\|_{L^\infty}$$

introduced in [20] (and studied further in [26], [23]) and a new family of norms, the *uniform almost periodicity norms*

$$\|f\|_{UAP^0} \geq \|f\|_{UAP^1} \geq \ldots \geq \|f\|_{UAP^{k-2}} \geq \ldots \geq \|f\|_{L^\infty}$$

[9] Strictly speaking, the U^0 and U^1 norms are not actually norms, and the UAP^0 norm can be infinite when f is non-constant. However, these issues will be irrelevant for our proof, and in the most interesting case $k \geq 3$ there are no such degeneracies.

which turn out to be somewhat dual to the Gowers uniformity norms. We shall mainly rely on the U^{k-1} and UAP^{k-2} norms; the other norms in the family are required only for mathematical induction purposes. We shall define the Gowers uniformity and uniform almost periodicity norms rigorously in Sections 4 and 5 respectively. For now, we shall simply give a very informal (and only partially accurate) heuristic: a function bounded in UAP^{k-2} will typically look something like the polynomially quasiperiodic function (1) where all the polynomials have degree at most $k-2$, whereas a function small in U^{k-1} is something like a function which is "orthogonal" to all such quasiperiodic functions (1).

Next, we state the three main (and independent) sub-theorems which we shall use to deduce Theorem 2.4. The first sub-theorem, which is rather standard (and the easiest of the three to prove), asserts that Gowers-uniform functions (i.e. functions with small U^{k-1} norm) are negligible for the purposes of computing (3); it will be proven in Section 4.

Theorem 3.1 (Generalized von Neumann theorem). *[20] Let $k \geq 2$, and let $\lambda_0, \ldots, \lambda_{k-1}$ be distinct elements of $\mathbf{Z}_N$. Then for any bounded functions $f_0, \ldots, f_{k-1} : \mathbf{Z}_N \to \mathbf{C}$ we have*

$$\left| \mathbf{E}_{r \in \mathbf{Z}_N} \int_{\mathbf{Z}_N} \prod_{j=0}^{k-1} T^{\lambda_j r} f_j \right| \leq \min_{1 \leq j \leq k} \|f_j\|_{U^{k-1}}.$$

Remark 3.2. As indicated, this part of the argument is based on the arguments of Gowers [20]; however it is purely combinatorial, relying on the Cauchy-Schwarz inequality rather than on Fourier analytic techniques (which occupy other parts of the argument in [20]). Variants of this theorem go back at least as far as Furstenberg [10]; see also [23], [26] for some variants of this theorem. We remark that the linear shifts $\lambda_j r$ can be replaced by more general objects such as polynomial shifts, after replacing the U^{k-1} norm by a higher Gowers uniformity norm; this is implicit for instance in [3].

The second sub-theorem is a special case of the main theorem, and addresses the complementary situation to Theorem 3.1, where f is now uniformly almost periodic instead of Gowers-uniform; it will be proven in Section 10.

Theorem 3.3 (Almost periodic functions are recurrent). *Let $d \geq 0$ and $k \geq 1$ be integers, and let $f_{U^\perp}, f_{UAP}$ be non-negative bounded functions such that we have the estimates*

$$\|f_{U^\perp} - f_{UAP}\|_{L^2} \leq \frac{\delta^2}{1024k} \tag{4}$$

$$\int_{\mathbf{Z}_N} f_{U^\perp} \geq \delta \tag{5}$$

$$\|f_{UAP}\|_{UAP^d} < M \tag{6}$$

for some $0 < \delta, M < \infty$. Then we have

$$\mathbf{E}_{0 \leq r \leq N_1} \int_{\mathbf{Z}_N} \prod_{j=0}^{k-1} T^{\mu j r} f_{U^\perp}(x) \gg_{d,k,\delta,M} 1 \tag{7}$$

for all $\mu \in \mathbf{Z}_N$ and $N_1 \geq 1$.

Remark 3.4. This argument is a quantitative version of certain ergodic theory arguments by Furstenberg and later authors, and is the only place where the van der Waerden theorem (Theorem 1.1) is required. It is by far the hardest component of the argument. In principle, the argument gives explicit bounds for the implied constant in (7) but they rely (repeatedly) on Theorem 1.1 and are thus quite weak. As mentioned earlier, we need this theorem only when $d = k - 2$, but allowing d to be arbitrary is convenient for the purposes of proving this theorem by induction. It is important that the quantity $\frac{\delta^2}{1024k}$ used in the right-hand side of (4) does not depend on M. This significantly complicates the task of proving this theorem when M is large, of course, since the error between $f_{U^\perp}$ and f_{UAP} may seem to dominate whatever gain one can obtain from (6). Nevertheless, one can cope with such large errors by means of the machinery of factors and conditional expectation. This ability to tolerate reasonably large L^2 errors in this recurrence result is also crucially exploited in the "Zorn's lemma" step in the ergodic theory arguments, in which one shows that the limit of a chain of extensions with the recurrence property is also recurrent. The parameters μ, N_1 are technical and are needed to facilitate the inductive argument used to prove this Theorem; ultimately we shall take $\mu := 1$ and $N_1 := N - 1$.

Finally, we need a structure theorem, proven in Section 8, that splits an arbitrary function into a Gowers-uniform component and an uniformly almost periodic component (plus an error).

Theorem 3.5 (Structure theorem). *Let $k \geq 3$, and let f be a non-negative bounded function obeying* (2) *for some $\delta > 0$. Let $F : \mathbf{R}^+ \to \mathbf{R}^+$ be an arbitrary function (which may depend on k and δ). Then we can find a positive number $M = O_{k,\delta,F}(1)$, a bounded function f_U, and non-negative bounded functions $f_{U^\perp}$, f_{UAP} such that we have the splitting*

$$f = f_U + f_{U^\perp}$$

and the estimates (4), (5), (6) *with $d := k - 2$, as well as the uniformity estimate*

$$\|f_U\|_{U^{k-1}} \leq \frac{1}{F(M)}. \tag{8}$$

Remark 3.6. The subscripts U and $U^\perp$ stand for *Gowers uniform* and *Gowers anti-uniform* respectively. Thus this theorem asserts that while a general function f need not have any uniformity properties whatsoever, it can be decomposed into pieces which are either uniform in the sense of Gowers, or are instead uniformly almost periodic, or are simply small in L^2. This theorem is something of a hybrid between the Furstenberg structure theorem [15] and the Szemerédi regularity lemma [40]. A similar structure theorem was a key component to [23]. The fact that the error tolerance in (4) does not go to zero as $M \to \infty$ is crucial in order to obtain this insensitivity to the choice of right-hand side of (8).

Remark 3.7. Each of the above three theorems have strong parallels in the genuinely ergodic theory setting. For instance, the analogues of the U^d norms in that setting were worked out by Host and Kra [26], where the analogue of Theorem 3.1 was also (essentially) proven. The structure theorem seems to correspond to the existence of a universal characteristic factor for Szemerédi-type recurrence properties (see e.g. [26], [47] for a discussion), although unlike the papers in [26], [47] we do *not* attempt to characterise this factor in terms of nilflows here. The recurrence theorem is very similar in spirit to $k-2$ iterations of the basic fact, established in [15], that recurrence properties are preserved under compact extensions (although our proof is not based on that argument, but instead on later colouring arguments such as the one in [3]). One can also extend the definition of the Banach algebra UAP^d defined below to the ergodic theory setting. It seems of interest to pursue these connections further, and in particular to rigorously pin down the relationship between almost periodicity of order $k-2$ and $k-2$-step nilsystems.

Assuming these three theorems, we can now quickly conclude Theorem 2.4.

Proof of Theorem 2.4. Let f, k, δ be as in Theorem 2.4. We may take $k \geq 3$ since the cases $k = 1, 2$ are trivial. Let $F : \mathbf{R}^+ \to \mathbf{R}^+$ be a growth function to be chosen later. Let M, f_U, $f_{U^\perp}$, f_{UAP} be as in Theorem 3.5. We can then split the left-hand side of (3) as the sum of 2^k terms of the form $\mathbf{E}_{r\in\mathbf{Z}_N} \int_{\mathbf{Z}_N} \prod_{j=0}^{k-1} T^{jr} f_j$, where each of the functions $f_0, \ldots, f_{k-1}$ are equal to either f_U or $f_{U^\perp}$. The term in which all the f_j are equal to $f_{U^\perp}$ is $\gg_{k,\delta,M} 1$ by Theorem 3.3 (taking $\mu := 1$ and $N_1 := N-1$). The other $2^k - 1$ terms have magnitude at most $\|f_U\|_{U^{k-1}} \leq 1/F(M)$ thanks to Theorem 3.1. Adding all this together, and taking $F(M)$ to be sufficiently rapidly growing function of M (and also depending on k, δ) we see that

$$\mathbf{E}_{r\in\mathbf{Z}_N} \int_{\mathbf{Z}_N} \prod_{j=0}^{k-1} T^{jr} f \gg_{k,\delta,M} 1.$$

Since $M = O_{k,\delta,F}(1) = O_{k,\delta}(1)$ the claim (3) follows. $\square$

It remains to define the U^{k-1} and UAP^{k-2} norms properly, and prove Theorems 3.1, 3.3, 3.5. This shall occupy the remainder of the paper.

4 Uniformity norms, and the generalized von Neumann theorem

In this section we define the Gowers uniformity norms U^d properly, and then prove Theorem 3.1. The motivation for these norms comes from the *van der Corput lemma*, which is very simple in the context of the cyclic group $\mathbf{Z}_N$:

Lemma 4.1 (Van der Corput Lemma). *For any function $f \in \mathbf{Z}_N \to \mathbf{C}$, we have*

$$|\int_{\mathbf{Z}_N} f|^2 = \mathbf{E}_{h\in\mathbf{Z}_N} \int_{\mathbf{Z}_N} \overline{f} T^h f.$$

Proof. Expanding both sides the identity becomes

$$\mathbf{E}_{x,y\in\mathbf{Z}_N} \overline{f}(x) f(y) = \mathbf{E}_{x,h\in\mathbf{Z}_N} \overline{f(x)} f(x+h)$$

and the claim follows by the substitution $y = x + h$. $\square$

Motivated by this lemma, we define

Definition 4.2 (Gowers uniformity norms). [20] Let $f : \mathbf{Z}_N \to \mathbf{C}$ be a function. We define the d^{th} Gowers uniformity norm $\|f\|_{U^d}$ recursively by

$$\|f\|_{U^0} := \int_{\mathbf{Z}_N} f \tag{9}$$

and

$$\|f\|_{U^d} := (\mathbf{E}_{h \in \mathbf{Z}_N} \|\overline{f} T^h f\|_{U^{d-1}}^{2^{d-1}})^{1/2^d} \tag{10}$$

for all $d \geq 1$.

Example 4.3. From Lemma 4.1, (9), (10) we obtain the explicit formula

$$\|f\|_{U^1} = |\int_{\mathbf{Z}_N} f|. \tag{11}$$

In particular, the U^1 norm (and hence all higher norms) are always non-negative. The U^2 norm can also be interpreted as the l^4 norm of the Fourier coefficients of f via the identity

$$\|f\|_{U^2} = (\sum_{\xi \in \mathbf{Z}_N} |\mathbf{E}_{x \in \mathbf{Z}_N} f(x) e(-x\xi/N)|^4)^{1/4}, \tag{12}$$

though we will not need this fact here. The higher U^d norms do not seem have any particularly useful Fourier-type representations, however by expanding (10) out recursively one can write the U^d norm as a sum of f over d-dimensional cubes (see [20], [23], [26] for further discussion of this).

Remark 4.4. The U^0 and U^1 norms are not, strictly speaking, norms; the latter is merely a semi-norm, and the former is not a norm at all. However, the higher norms U^d, $d \geq 2$ are indeed norms (they are homogeneous, non-degenerate, and obey the triangle inequality), and are also related to a certain 2^d-linear inner product; see [20], [23], or [26] for a proof of these facts (which we will not need here), with the $d = 2$ case following directly from inspection of (12). Also one can show the inequality $\|f\|_{U^d} \leq \|f\|_{U^{d+1}}$ for any $d \geq 0$. Thus for $k \geq 2$, we have a rather interesting nested sequence of Banach spaces U^{k-1} of functions $f : \mathbf{Z}_N \to \mathbf{C}$, equipped with the U^{k-1} norm; these Banach spaces and their duals $(U^{k-1})^*$ were explored to a limited extent in [23], and we shall continue their study later in this paper. Functions which are small in U^2 norm are termed *linearly uniform* or *Gowers-uniform of order 1*, and thus have small Fourier coefficients by (12); functions small in U^3 norm are *quadratically uniform* or *Gowers-uniform of order 2*, and so forth. The terminology here is partly explained by the next example; again, see [20], [23], or [26] for further discussion.

Example 4.5. By induction[10] we see that $\|f\|_{U^d} \leq \|f\|_{L^\infty}$ for all d; in particular we have $\|f\|_{U^d} \leq 1$ when f is bounded. We now present an example (which is, in fact, the only

[10] Actually, more is true: the U^d norms of f increase monotonically and converge to $\|f\|_{L^\infty}$ as $d \to \infty$, although the convergence can be quite slow and depends on N. We will not prove this fact here.

example up to scalar multiplication) in which equality holds. Let $P : \mathbf{Z}_N \to \mathbf{Z}_N$ be a polynomial with coefficients in $\mathbf{Z}_N$, and let $f(x) := e(P(x)/N)$. Then one can show that $\|f\|_{U^d} = 1$ when $d \geq \deg(P)$, and $\|f\|_{U^d} = o_{\deg P}(1)$ when $d < \deg(P)$; the former fact can be proven by induction and the trivial observation that for each fixed h, the polynomial $P(x+h) - P(x)$ has degree at most $\deg(P) - 1$, while the latter fact also follows from induction, the above observation, and Lemma 4.1; we omit the details. In fact one can improve the $o_{\deg P}(1)$ bound to $O_{\deg P}(N^{-1/2^{d+1}})$, by using the famous Weil estimates. By using the triangle inequality for U^d (see e.g. [20], [23]) one can also deduce similar statements for the polynomially quasiperiodic functions (1).

One can easily verify by induction that the U^d norms are invariant under shifts, thus $\|T^n f\|_{U^d} = \|f\|_{U^d}$, and also invariant under dilations, thus if $\lambda \in \mathbf{Z}_N \backslash 0$ and $f_\lambda(x) := f(x/\lambda)$ then $\|f_\lambda\|_{U^d} = \|f\|_{U^d}$.

We can now prove the generalized von Neumann theorem.

Proof of Theorem 3.1. We induct on k. When $k = 2$ we use the fact that $(x, r) \mapsto (x + \lambda_1 r, x + \lambda_2 r)$ is a bijection from $\mathbf{Z}_N^2$ to $\mathbf{Z}_N^2$ (recalling that N is prime) to conclude that

$$\mathbf{E}_{r \in \mathbf{Z}_N} \int_{\mathbf{Z}_N} \prod_{j=0}^{1} T^{\lambda_j r} f_j = \left(\int_{\mathbf{Z}_N} f_0 \right) \times \left(\int_{\mathbf{Z}_N} f_1 \right)$$

and the claim then follows easily from (11) and the boundedness of f_0, f_1. Now suppose that $k > 2$ and the claim has already been proven for $k - 1$. By permuting the λ_j if necessary we may assume that the minimum of the $\|f_j\|_{U^{k-1}}$ is attained when $j = 0$. By making the scaling $r \mapsto \lambda_0 r$ if necessary we may assume that $\lambda_0 = 1$. By applying the expectation-preserving map $T^{-\lambda_{k-1} r}$ (i.e. by subtracting λ_{k-1} from each of the λ_j) we may assume that λ_{k-1} is zero. The claim can now be written as

$$\left| \int_{\mathbf{Z}_N} f_{k-1} \mathbf{E}_{r \in \mathbf{Z}_N} (\prod_{j=0}^{k-2} T^{\lambda_j r} f_j) \right| \leq \|f_0\|_{U^{k-1}}.$$

By the Cauchy-Schwarz inequality and the boundedness of f_{k-1}, it suffices to prove that

$$\int_{\mathbf{Z}_N} |\mathbf{E}_{r \in \mathbf{Z}_N} \prod_{j=0}^{k-2} T^{\lambda_j r} f_j|^2 \leq \|f_0\|_{U^{k-1}}^2.$$

But from Lemma 4.1 (and the fact that $h \mapsto \lambda_j h$ is a bijection on $\mathbf{Z}_N$) we have

$$\begin{aligned} \int_{\mathbf{Z}_N} |\mathbf{E}_{r \in \mathbf{Z}_N} \prod_{j=0}^{k-2} T^{\lambda_j r} f_j|^2 &= \mathbf{E}_{h, r \in \mathbf{Z}_N} \int_{\mathbf{Z}_N} \overline{\left(\prod_{j=0}^{k-2} T^{\lambda_j r} f_j \right)} \left(\prod_{j=0}^{k-2} T^{\lambda_j (r+h)} f_j \right) \\ &= \mathbf{E}_{h \in \mathbf{Z}_N} \mathbf{E}_{r \in \mathbf{Z}_N} \int_{\mathbf{Z}_N} \prod_{j=0}^{k-2} T^{\lambda_j r} \left(\overline{f_j} T^{\lambda_j h} f_j \right). \end{aligned}$$

On the other hand, from the induction hypothesis and the reduction to the case $\lambda_0 = 1$ we have

$$\left|\mathbf{E}_{r\in\mathbf{Z}_N}\int_{\mathbf{Z}_N}\prod_{j=0}^{k-2}T^{\lambda_j r}(\overline{f_j}T^{\lambda_j h}f_j)\right| \leq \|f_0T^hf_0\|_{U^{k-2}}$$

for all $h \in \mathbf{Z}_N$. Combining these two facts together, we obtain

$$\int_{\mathbf{Z}_N}|\mathbf{E}_{r\in\mathbf{Z}_N}\prod_{j=0}^{k-2}T^{\lambda_j r}f_j|^2 \leq \mathbf{E}_{h\in\mathbf{Z}_N}\|f_0T^hf_0\|_{U^{k-2}}$$

and the claim follows from (10) and Hölder's inequality. □

Remark 4.6. The notion of Gowers uniformity considered here, namely that the U^{k-1} norm is small, generalizes the concept of *pseudorandomness* or *linear uniformity* in the $k = 3$ case, which amounts to the assertion that all the Fourier coefficients of f (except possibly for the zero coefficient) are small; this is the notion used for instance in [33], [6], [25], [7]. This notion is essentially equivalent to the pair correlations of all the shifts T^nf to be small on the average. For higher k, this notion is insufficient to obtain theorem such as Theorem 3.1, see [19], [20] for further discussion. In Szemerédi's original arguments [38], [39], the appropriate concept of uniformity is provided by the notion of *ε-regularity*, which roughly corresponds to controlling all the U^d norms for $d = O_\varepsilon(1)$, while in the ergodic theory arguments of Furstenberg and later authors, the notion of uniformity used is that of *weak mixing*, which roughly corresponds to controlling the U^d norms for *all* d. Thus these notions of uniformity are significantly stronger than the one considered here, which fixes d at $k-1$. There is of course a cost to using such a strong notion of uniformity, and it is that one has to make the tower of structures extremely large in order to eventually attain such uniformity. In Szemerédi's regularity lemma, for instance, one is forced to lose constants which are of tower-exponential type in the regularity parameter ε; see [18]. In the ergodic theory arguments, the situation is even worse; the tower of invariant factors given by Furstenberg's structure theorem (the ergodic theory analogue of Szemerédi's regularity lemma) can be as tall as any countable ordinal, but no taller; see [2].

Remark 4.7. In the ergodic theory setting, one can also define analogues of the U^{k-1} norms, giving rise to the concept of invariant factors whose complement consists entirely of functions which are *Gowers-uniform of order* $k-2$; using this notion (which is much weaker than weak mixing) it is possible to obtain a version of Furstenberg's structure theorem using only a tower of height $k-2$ (in particular, a tower of finite height). Indeed, it was the author's discovery of this fact which led eventually to the quantitative proof presented here; we have since learnt that this fact is essentially implicit in the work of Host and Kra [26] and Ziegler [47].

5 Almost periodic functions

Having defined the Gowers uniformity norms U^{k-1} used for the generalized Von Neumann theorem, we now turn to defining the dual concept of the uniform almost periodicity norms

UAP^{k-2} which we will need for both the recurrence theorem and structure theorems. Roughly speaking, if a function F has a bounded UAP^{k-2} norm, then it should resemble a function of the form (1), which we shall loosely refer to as a *quasiperiodic function of order* $k-2$. To quantify this we make the following observation: if F is of the form (1), then

$$T^n F = \mathbf{E}_{j \in J} c_{n,j} g_j, \tag{13}$$

where g_j is the bounded function $g_j(x) := e(P_j(x)/N)$, and $c_{n,j}(x)$ is the $c_{n,j}(x) = e((P_j(x+n) - P_j(x))/N)$. The point here is that the dependence on n on the right-hand side only arises through the functions $c_{n,j}$, and those functions are of the form (1) but with degree $k-3$ instead of $k-2$. Thus, the shifts of an quasiperiodic functions of order $k-2$ can be written as linear combinations of fixed bounded functions g_j, where the coefficients $c_{n,j}$ are not constant, but are instead quasiperiodic functions of one lower order[11]. We can pursue the same idea to define the UAP^d norms recursively as follows.

Definition 5.1 (Banach algebras). A space A of functions on $\mathbf{Z}_N$, equipped with a norm $\|\|_A : A \to \mathbf{R}^+$, is said to be a *Banach algebra* if A is a vector space, $\|\|_A$ is a norm (i.e. it is homogeneous, non-degenerate, and verifies the triangle inequality) which is invariant under conjugation $f \mapsto \overline{f}$, and A is closed under pointwise product with $\|fg\|_A \leq \|f\|_A \|g\|_A$ for all $f, g \in A$. We also assume that

$$\|F\|_{L^\infty} \leq \|F\|_A \tag{14}$$

for all $F \in A$ (actually this property can be deduced from the pointwise product property and the finite-dimensionality of A). We adopt the convention that $\|f\|_A = \infty$ if $f \notin A$. We say that A is *shift-invariant* if $\|f\|_A = \|T^n f\|_A$ for all n, and *scale-invariant* if $\|f_\lambda\|_A = \|f\|_A$ for all $\lambda \in \mathbf{Z}_N \backslash \{0\}$.

Definition 5.2 (Uniform almost periodicity norms). If A is a shift-invariant Banach algebra of functions on $\mathbf{Z}_N$, we define the space $UAP[A]$ to be the space of all functions F for which the orbit $\{T^n F : n \in \mathbf{Z}\}$ has a representation of the form

$$T^n F = M\mathbf{E}(c_{n,h} g_h) \text{ for all } n \in \mathbf{Z}_N \tag{15}$$

where $M \geq 0$, H is a finite non-empty set, $g = (g_h)_{h \in H}$ is a collection of bounded functions, $c = (c_{n,h})_{n \in \mathbf{Z}_N, h \in H}$ is a collection of functions in A with $\|c_{n,h}\|_A \leq 1$, and h is a random variable taking values in H. We define the norm $\|F\|_{UAP[A]}$ to be the infimum of M over all possible representations of this form.

Remark 5.3. Note that we are not imposing any size constraints on H; in fact one could allow H to be infinite, so that h becomes a continuous random variable rather than a discrete one, without actually affecting this definition. It turns out however that we will not need any information about H, or more generally about the probability distribution of the random variable h. The key point is that the Volterra operator $(c_h)_{h \in H} \mapsto \mathbf{E}(c_h g_h)$ will be a "compact" operator uniformly in choice of h and $(g_h)_{h \in H}$.

[11] This observation was motivated by the use of *relatively almost periodic functions* in the ergodic theory arguments of Furstenberg [10], [15], [11] and later authors.

We first observe that the construction $A \mapsto UAP[A]$ maps shift-invariant Banach algebras to shift-invariant Banach algebras:

Proposition 5.4. *If A is a shift-invariant Banach algebra, then so is $UAP[A]$. Furthermore $UAP[A]$ contains A, and $\|f\|_{UAP[A]} \leq \|f\|_A$ for all $f \in A$. Finally, if A is scale-invariant then so is $UAP[A]$.*

Remark 5.5. This is a quantitative analogue of the well known fact in ergodic theory that the almost periodic functions form a shift-invariant algebra.

Proof. It is easy to see that $UAP[A]$ is shift-invariant, conjugation-invariant, closed under scalar multiplication, preserves scale-invariance, and that the $UAP[A]$ norm is non-negative and homogeneous. From (14) and (15) we see that $\|f\|_{L^\infty} \leq \|f\|_{UAP[A]}$ for all $f \in UAP[A]$, from which we deduce that the $UAP[A]$ norm is non-degenerate. Also we easily verify that $UAP[A]$ contains A with $\|f\|_{UAP[A]} \leq \|f\|_A$ for all $f \in A$. Next, we show that $UAP[A]$ is closed under addition and that the UAP^d norm enjoys the triangle inequality. By homogeneity and nondegeneracy it suffices to show that the unit ball is convex, i.e. if $F, F' \in UAP[A]$ are such that $\|F\|_{UAP[A]}, \|F'\|_{UAP[A]} < 1$ then $(1-\theta)F + \theta F' \in UAP[A]$ with $\|(1-\theta)F + \theta F'\|_{UAP[A]} < 1$ for all $0 \leq \theta \leq 1$. By Definition 5.2 we can find non-empty finite sets H, H', bounded functions $(g_h)_{h \in H}$ and $(g'_{h'})_{h' \in H'}$, and functions $(c_{n,h})_{n \in \mathbf{Z}_N, h \in H}$ and $(c'_{n,h'})_{n \in \mathbf{Z}_N, h' \in H'}$ in A and random variables h, h' taking values in H and H' respectively such that we have the representations

$$T^n F = \mathbf{E}(c_{n,h} g_h); \quad T^n F' = \mathbf{E}(c'_{n,h'} g'_{h'}) \text{ for all } n \in \mathbf{Z}_N \tag{16}$$

and the estimates

$$\|c_{n,h}\|_A, \|c'_{n,h'}\|_A \leq 1 \text{ for all } n \in \mathbf{Z}_N, h \in H, h' \in H'.$$

Also, by relabeling H' if necessary we may assume that H and H' are disjoint. In such a case we can concatenate $(c_{n,h})_{n \in \mathbf{Z}_N, h \in H}$ and $(c'_{n,h'})_{n \in \mathbf{Z}_N, h' \in H'}$ to a single function $(\tilde{c}_{n,\tilde{h}})_{n \in \mathbf{Z}_N, \tilde{h} \in H \cup H'}$ and similarly concatenate the g functions to $(\tilde{g}_{n,\tilde{h}})_{\tilde{h} \in H \cup H'}$. If one then defines the random variable $\tilde{h}$ to equal h with probability $(1-\theta)$ and h' with probability θ (or more precisely, the probability distribution of $\tilde{h}$ is $1-\theta$ times that of h plus θ times that of h') then one sees from linearity of expectation that

$$T^n(F + F') = \mathbf{E}(\tilde{c}_{n,\tilde{h}} \tilde{g}_{\tilde{h}}) \text{ for all } n \in \mathbf{Z}_N$$

and the claim follows.

Next, we establish the algebra property. By homogeneity and nondegeneracy again it suffices to show that the unit ball is closed under multiplication. To see this, start with (16). Without loss of generality we may assume that the random variables h, h' are independent (because it is only their individual distributions which matter for (16), not their joint distribution). But in that case we have

$$T^n(FF') = \mathbf{E}(c_{n,h,h'} g_{h,h'}) \text{ for all } n \in \mathbf{Z}_N$$

where $c_{n,h,h'} := c_{n,h} c_{n,h'}$ and $g_{h,h'} := g_h g_{h'}$. Since the product of two bounded functions is a bounded function, the claim follows from the algebra property of A. $\square$

Thanks to this proposition, we can define the UAP^d norms recursively for $d \geq 0$, by setting UAP^0 to be the trivial Banach algebra of all constant functions (equipped with the L^∞ norm), and then setting $UAP^d := UAP[UAP^{d-1}]$ for all $d \geq 1$. Thus UAP^d is a shift-invariant, scale-invariant Banach algebra for all $d \geq 0$.

In the case $d = 3$, it was observed by Ben Green (private communication) the UAP^1 norm was in fact equivalent to the Wiener norm. We present the elegant proof here.

Proposition 5.6. *For any $f : \mathbf{Z}_N \to \mathbf{C}$ we have*

$$\|F\|_{UAP^1} = \sum_{\xi \in \mathbf{Z}_N} |\mathbf{E}_{x \in \mathbf{Z}_N} F(x) e(-x\xi/N)|.$$

Proof. One can prove by direct computation that each character $e(x\xi/N)$ has a UAP^1 norm of at most 1. From the Fourier inversion formula and the triangle inequality for UAP^1 we thus obtain the upper bound

$$\|F\|_{UAP^1} \leq \sum_{\xi \in \mathbf{Z}_N} |\mathbf{E}_{x \in \mathbf{Z}_N} F(x) e(-x\xi/N)|.$$

Now we prove the lower bound; it suffices to show that if $\|F\|_{UAP^1} < 1$ then

$$\sum_{\xi \in \mathbf{Z}_N} |\mathbf{E}_{x \in \mathbf{Z}_N} F(x) e(-x\xi/N)| \leq 1.$$

By Definition 5.2, we have a representation

$$T^n F(x) = \mathbf{E}(c_{n,h} g_h(x))$$

for all $n \in \mathbf{Z}_N$, where the $c_{n,h}$ are constants bounded in magnitude by 1 and the g_h are bounded functions. Since

$$\mathbf{E}_{x \in \mathbf{Z}_N} F(x) e(-x\xi/N) = \mathbf{E}_{n,x \in \mathbf{Z}_N} T^n F(x) e(-x\xi/N) e(n\xi/N)$$

we conclude that

$$\sum_{\xi \in \mathbf{Z}_N} |\mathbf{E}_{x \in \mathbf{Z}_N} F(x) e(-x\xi/N)| \leq \mathbf{E} \sum_{\xi \in \mathbf{Z}_N} |\mathbf{E}_{n,x \in \mathbf{Z}_N} c_{n,h} g_h(x) e(-x\xi/N) e(n\xi/N)|.$$

Separating the x and n averages and applying Cauchy-Schwarz, we can bound the right-hand side by

$$\mathbf{E}(\sum_{\xi \in \mathbf{Z}_N} |\mathbf{E}_{n \in \mathbf{Z}_N} c_{n,h} e(n\xi/N)|^2)^{1/2} (\sum_{\xi \in \mathbf{Z}_N} |\mathbf{E}_{x \in \mathbf{Z}_N} g_h(x) e(-x\xi/N)|^2)^{1/2}.$$

Applying Plancherel's theorem and the bounds on $c_{n,h}$ and g_h we obtain a bound of 1 as desired. $\square$

Example 5.7. If F is of the form (1), then one can verify by (13) and induction that $F \in UAP^{k-2}$ with $\|F\|_{UAP^{k-2}} \leq 1$ (here h is an element of $\{1, \ldots, J\}$ chosen uniformly at random). In particular, as observed above, UAP^1 contains the Wiener algebra of functions with absolutely convergent Fourier series. In our finitary setting of $\mathbf{Z}_N$, this implies that *every* function lies in UAP^1 and hence in all higher UAP^d norms, though the norm may grow with N and thus be very large. Note that the property of being in the Wiener algebra is substantially stronger than being *almost periodic*, which is roughly equivalent to asking that the Fourier coefficients are summable in $l^{2-\varepsilon}$ for some $\varepsilon > 0$ rather than being summable in l^1. For further comparison, the property of being bounded in $(U^2)^*$, as discussed in [23], is stronger than being almost periodic but weaker than being bounded in UAP^1; it is equivalent to asking for the Fourier coefficients to be summable in $l^{4/3}$. See [23] for further discussion.

Example 5.8. There are more subtle examples of almost periodic functions than the quasiperiodic ones. One example is the function

$$f(x) := e(\lfloor \frac{ax}{N} \rfloor \frac{b}{N})\psi(\frac{ax}{N} \bmod 1)\psi(\frac{x}{N} \bmod 1)$$

for some fixed $1 \leq a, b \leq N$, where x is thought of as an integer from 1 to N, $\lfloor x \rfloor$ denotes the integer part of x and $\psi(x)$ is a smooth cutoff to the interval $[0.4, 0.6]$ (say).. This function has an UAP^1 norm of $O(1)$ uniformly in a, b, N, but the required representation of the form (15) is not particularly obvious (for instance one can set g_h to be various translations and modulations of f, and then $T^n f$ can be decomposed as an absolutely summable combination of the g_h using smooth partitions of unity and Fourier series). In this case, one can eventually work out that f also has an absolutely convergent Fourier series; however things are even less clear for the function

$$f(x) := e(\lfloor ax/N \rfloor \frac{bx}{N})\psi(\frac{ax}{N} \bmod 1)\psi(\frac{x}{N} \bmod 1), \tag{17}$$

which has an UAP^2 norm of $O(1)$ but seems to have no particular resemblance with any quadratic phase function. These "generalized quadratic phase functions" are related to 2-step nilsystems, which are known to not always admit quadratic eigenfunctions; see e.g. [16] for further discussion. Intriguingly, hints of this "generalized quadratic" structure also emerge in the work of Gowers [19]. For further discussion see [24].

The structure theorem, Theorem 3.5, can be viewed as some sort of duality relationship between UAP^{k-2} and U^{k-1}. We now provide two demonstrations of this duality. The first such demonstration is rather simple, but is not actually used in the proof of Szemerédi's theorem; the second demonstration will be to some extent a converse of the first and is one of the key components used to prove Theorem 3.5.

Proposition 5.9 (Uniformity is orthogonal to almost periodicity). *Let $k \geq 2$. For any functions f, F with $F \in UAP^{k-2}$, we have*

$$|\langle f, F \rangle| \leq \|f\|_{U^{k-1}} \|F\|_{UAP^{k-2}}.$$

Proof. We induct on k. When $k = 2$ the claim follows from (11) and the fact that F is necessarily constant. Now suppose that $k \geq 3$ and the claim has already been proven for $k-1$. By homogeneity it suffices to show that if $\|f\|_{U^{k-1}}, \|F\|_{UAP^{k-2}} < 1$ then $|\langle f, F\rangle| < 1$.

By Definition 5.2 we can find H, g, c, h with the representation (15), with g_h bounded and $\|c_{n,h}\|_{AP^{k-3}} \leq 1$ for all $n \in \mathbf{Z}_N$, $h \in H$. Next, we use (15) the unitary nature of T^n to write

$$\langle f, F\rangle = \langle T^n f, T^n F\rangle = \int_{\mathbf{Z}_N} T^n f \mathbf{E}(\overline{c_{n,h} g_h});$$

averaging over n and rearranging we thus have

$$\langle f, F\rangle = \mathbf{E} \int_{\mathbf{Z}_N} \mathbf{E}_{n \in \mathbf{Z}_N} (T^n f \overline{c_{n,h}}) \overline{g_h}.$$

By the Cauchy-Schwarz inequality and the boundedness of the g_h, we thus have

$$|\langle f, F\rangle| \leq \left(\mathbf{E} \int_{\mathbf{Z}_N} |\mathbf{E}_{n \in \mathbf{Z}_N} (T^n f \overline{c_{n,h}})|^2 \right)^{1/2}.$$

But from Lemma 4.1 we have the pointwise identity

$$|\mathbf{E}_{n \in \mathbf{Z}_N} (T^n f \overline{c_{n,h}})|^2 = \mathbf{E}_{n,r \in \mathbf{Z}_N} (T^n(\overline{f} T^r f) c_{n,h} \overline{c_{n+r,h}})$$

whence

$$|\langle f, F\rangle| \leq \left(\mathbf{E}\mathbf{E}_{n,r \in \mathbf{Z}_N} \langle \overline{f} T^r f, T^{-n}(\overline{c_{n,h}} c_{n+r,h})\rangle \right)^{1/2}.$$

Since UAP^{k-3} is a shift-invariant Banach algebra we have $\|T^{-n}(\overline{c_{n,h}} c_{n+r,h})\|_{UAP^{k-3}} \leq 1$. By the inductive hypothesis we thus have

$$|\langle \overline{f} T^r f, T^{-n}(\overline{c_{n,h}} c_{n+r,h})\rangle| \leq \|\overline{f} T^r f\|_{U^{k-2}},$$

whence

$$|\langle f, F\rangle| \leq \mathbf{E}(\mathbf{E}_{n,r \in \mathbf{Z}_N} \|\overline{f} T^r f\|_{U^{k-2}})^{1/2}.$$

The outer expectation can be discarded since the quantity inside the expectation is deterministic. We may similarly discard the redundant n average. Using Cauchy-Schwarz and (10), we thus obtain

$$|\langle f, F\rangle| \leq \|f\|_{U^{k-1}} < 1$$

as desired. □

Remark 5.10. One can use this Proposition to give an alternate proof of Theorem 3.1, based on the observation (easily verified by induction) that if $f_1, \ldots, f_{k-1}$ are bounded functions and $\lambda_1, \ldots, \lambda_{k-1}$ are disjoint non-zero elements of $\mathbf{Z}_N$, then $\mathbf{E}_{r \in \mathbf{Z}_N} \prod_{j=1}^{k-1} T^{\lambda_j r} f_j$ lies in UAP^{k-2} with norm at most 1.

Remark 5.11. In the notation of [23], this shows that the UAP^{k-2} norm is larger than or equal to the $(U^{k-1})^*$ norm. However, the UAP^{k-2} norm is strictly stronger. For instance, as observed in [23] the $(U^2)^*$ norm is the $l^{4/3}$ norm of the Fourier coefficients, which by Proposition 5.6 is a strictly weaker norm than the UAP^1 norm.

We now give a partial converse to Proposition 5.9, which is the key to Theorem 3.5.

Lemma 5.12 (Lack of Gowers-uniformity implies correlation with a UAP function). *[23] Let f be a bounded function such that $\|f\|_{U^{k-1}} \geq \varepsilon$ for some $k \geq 3$ and $\varepsilon > 0$. Then there exists a bounded function $F \in UAP^{k-2}$ with $\|F\|_{UAP^{k-2}} \leq 1$ such that $|\langle f, F\rangle| \geq \varepsilon^{2^{k-1}}$.*

Proof. We need the concept of a *dual function* from [23]; the ergodic theory analogue of such functions have also been recently studied in [26], [1]. For any function $f : \mathbf{Z}_N \to \mathbf{C}$ and any $d \geq 0$, we define the *dual function of order d* of f, denoted $\mathcal{D}_d(f)$, by the recursive formula

$$\mathcal{D}_0(f) := 1 \tag{18}$$

(i.e. $\mathcal{D}_0(f)$ is just the constant function 1) and

$$\mathcal{D}_d(f) := \mathbf{E}_{h \in \mathbf{Z}_N} \overline{\mathcal{D}_{d-1}(\overline{f} T^h f)} T^h f \tag{19}$$

for all $d \geq 1$.

We now claim the identity

$$\langle f, \mathcal{D}_d(f)\rangle = \|f\|_{U^d}^{2^d}$$

for all $d \geq 0$. When $d = 0$ the claim follows from (18) and (9). Now suppose inductively that $d \geq 1$ and the claim has already been proven for $d - 1$. By (19) (and the definition of the inner product) we have

$$\langle f, \mathcal{D}_d(f)\rangle = \mathbf{E}_{h \in \mathbf{Z}_N} \langle \mathcal{D}_{d-1}(\overline{f} T^h f), \overline{f} T^h f\rangle,$$

and the claim now follows from the inductive hypothesis and (10).

We thus set $F := \mathcal{D}_{k-1}(f)$. It is clear from induction that F is bounded; it remains to show that F has UAP^{k-2} norm less than 1. Indeed, we make the more general claim that

$$\|\mathcal{D}_d(f)\|_{UAP^{d-1}} \leq 1 \text{ for all bounded } f \text{ and all } d \geq 1.$$

When $d = 1$ this is clear since $\mathcal{D}_d(f)$ is just the constant function $\mathbf{E}(f)$ in this case. Now suppose inductively that $d \geq 2$ and the claim has already been proven for $d - 1$. Applying T^n to both sides of (19) and making the change of variables $h \leftarrow n + h$ we obtain

$$T^n \mathcal{D}_d(f) := \mathbf{E}_{h \in \mathbf{Z}_N} c_{n,h} g_h, \text{ where } c_{n,h} := T^n \overline{\mathcal{D}_{d-1}(\overline{f} T^{h-n} f)} \text{ and } g_h := T^h f.$$

The functions g_h are clearly bounded. Since UAP^{d-2} is a shift-invariant Banach algebra, we see from inductive hypothesis that the functions $c_{n,h}$ lie in UAP^{d-2} with norm at most 1, and the claim follows (thinking of h as an element of $\mathbf{Z}_N$ chosen uniformly at random). ◻

6 Factors of almost periodic functions

To prove both the recurrence theorem (Theorem 3.3) or the structure theorem (Theorem 3.5) it is convenient not just to work with almost periodic functions, but also with certain *factors* (or σ-algebras) generated by them (in analogy with the ergodic theory arguments). Thus in this section we develop the theory of such factors.

We begin by recalling the definition of a factor (not necessarily associated with an almost periodic function). Here we shall identify factors with the σ-algebras that they generate. This latter concept is of course equivalent in the finitary setting of $\mathbf{Z}_N$ to the more familiar notion of a *partition* of $\mathbf{Z}_N$, but we will retain the language of probability theory in order to maintain the analogy with the ergodic theory arguments, and in order to benefit from such useful concepts as conditional expectation, orthogonality, measurability, energy, and so forth. See [41] for a further discussion of the connection between the Szemerédi regularity lemma, partitions, and conditional expectation with respect to factor.

Definition 6.1 (Factors). A *factor* $\mathcal{B}$ in $\mathbf{Z}_N$ is any collection of subsets of $\mathbf{Z}_N$ which contains the empty set $\emptyset$ and the full set $\mathbf{Z}_N$, and is closed under complementation, unions and intersections. We define the *atoms* of a factor to be the minimal non-empty elements of $\mathcal{B}$ (with respect to set inclusion); it is clear that the atoms in $\mathcal{B}$ form a partition of $\mathbf{Z}_N$, and $\mathcal{B}$ consists precisely of arbitrary unions of its atoms (including the empty union $\emptyset$); thus there is a one-to-one correspondence between factors and partitions of $\mathbf{Z}_N$. A function $f : \mathbf{Z}_N \to \mathbf{C}$ is said to be *measurable* with respect to a factor $\mathcal{B}$ if all the level sets of f lie in $\mathcal{B}$, or equivalently if f is constant on each of the atoms of $\mathcal{B}$.

We define $L^2(\mathcal{B}) \subseteq L^2(\mathbf{Z}_N)$ to be the closed subspace of the Hilbert space $L^2(\mathbf{Z}_N)$ consisting of $\mathcal{B}$-measurable functions. We can then define the conditional expectation operator $f \mapsto \mathbf{E}(f|\mathcal{B})$ to be the orthogonal projection of $L^2(\mathbf{Z}_N)$ to $L^2(\mathcal{B})$. An equivalent definition of conditional expectation is

$$\mathbf{E}(f|\mathcal{B})(x) := \mathbf{E}_{\mathcal{B}(x)} f$$

for all $x \in \mathbf{Z}_N$, where $\mathcal{B}(x)$ is the unique atom in $\mathcal{B}$ which contains x. It is clear that conditional expectation is a linear self-adjoint orthogonal projection on $L^2(\mathbf{Z}_N)$, preserves non-negativity, expectation, and constant functions. In particular it maps bounded functions to bounded functions. If $\mathbf{E}(f|\mathcal{B})$ is zero we say that f is *orthogonal to* $\mathcal{B}$.

If $\mathcal{B}, \mathcal{B}'$ are two factors, we use $\mathcal{B} \vee \mathcal{B}'$ to denote the factor generated by $\mathcal{B}$ and $\mathcal{B}'$ (i.e. the factor whose atoms are the intersections of atoms in $\mathcal{B}$ with atoms in $\mathcal{B}'$).

Observe that when $\mathcal{B}$ is the trivial factor $\{\emptyset, \mathbf{Z}_N\}$ then the conditional expectation $\mathbf{E}(f|\mathcal{B})$ is just the constant function equal to the integral $\int_{\mathbf{Z}_N} f$. Every factor induces a unique orthogonal decomposition $f = \mathbf{E}(f|\mathcal{B}) + (f - \mathbf{E}(f|\mathcal{B}))$ of a function f into the component $\mathbf{E}(f|\mathcal{B})$ measurable with respect to $\mathcal{B}$, and the component $f - \mathbf{E}(f|\mathcal{B})$ orthogonal to $\mathcal{B}$. More generally, if $\mathcal{B}$ is a subalgebra of $\mathcal{B}'$ (thus $\mathcal{B}'$ is *finer* than $\mathcal{B}$, or $\mathcal{B}$ is *coarser* than $\mathcal{B}'$) then we can orthogonally decompose the finer expectation $\mathbf{E}(f|\mathcal{B}') =$

$\mathbf{E}(f|\mathcal{B}) + (\mathbf{E}(f|\mathcal{B}') - \mathbf{E}(f|\mathcal{B}))$ into the coarser expectation, and a component measurable in $\mathcal{B}'$ but orthogonal to $\mathcal{B}$.

We now show that each almost periodic function generates a well-behaved factor at every scale ε.

Proposition 6.2 (UAP functions generate a compact factor). *Let $d \geq 0$, let $G \in UAP^d$ be such that $\|G\|_{UAP^d} \leq M$ for some $M > 0$, and let $\varepsilon > 0$. Then there exists a factor $\mathcal{B}_\varepsilon(G) = \mathcal{B}_\varepsilon(G, d)$ consisting of at most $O_{M,\varepsilon}(1)$ atoms, such that we have the following two properties:*

- *(G lies in its own factor) We have the approximation property*

$$\|G - \mathbf{E}(G|\mathcal{B}_\varepsilon(G) \vee \mathcal{B})\|_{L^\infty} \ll \varepsilon \tag{20}$$

  for any factor $\mathcal{B}$.

- *(Approximation by almost periodic functions) For any bounded non-negative function f which is measurable in $\mathcal{B}_\varepsilon(G)$, and any $\delta > 0$, there exists a bounded non-negative function $f_{UAP} \in UAP^d$ such that*

$$\|f - f_{UAP}\|_{L^2} \leq \delta \tag{21}$$

  and

$$\|f_{UAP}\|_{UAP^d} \ll_{M,\varepsilon,\delta} 1. \tag{22}$$

Remark 6.3. As the proof shows, the above Proposition in fact holds if UAP^d is replaced by any other Banach algebra.

Proof. We shall prove this by constructing $\mathcal{B}_\varepsilon(G)$ using randomized level sets (or "generalized Bohr sets" of G), using some ideas from [23]. Let $S := \{z \in \mathbf{C} : -1/2 \leq \Re(z), \Im(z) < 1/2\}$ be the standard unit square in the complex plane, and let $\mathbf{Z}[i] := \{a + bi : a, b \in \mathbf{Z}\}$ denote the Gaussian integers. Let $\alpha \in S$ be a complex number chosen uniformly at random from S. We can then define the factor $\mathcal{B}_{\varepsilon,\alpha}(G)$ to be the algebra whose atoms are the sets $\{G^{-1}(\varepsilon(S + \zeta + \alpha)) : \zeta \in \mathbf{Z}[i]\}$. It will suffice to show that with positive probability this algebra $\mathcal{B}_{\varepsilon,\alpha}(G)$ is a candidate for $\mathcal{B}_\varepsilon(G)$ (with the bounds in (20), (28) uniform in α).

The bound (20) is clear since G is constrained to lie in a square of diameter $O(\varepsilon)$ on each atom of $\mathcal{B}_\varepsilon$ (and hence on each atom of any finer factor). Since $\|G\|_{L^\infty} \leq \|G\|_{UAP^d} \leq M$, G takes values in a ball of radius $O(M)$, and thus the number of atoms in $\mathcal{B}_\varepsilon$ is indeed $O_{M,\varepsilon}(1)$ as claimed. Now we turn to the approximation property. It will suffice to prove that for each $\delta > 0$ and every $\eta > 0$, the approximation property is true (with the bound in (28) allowed to depend on η) with probability at least $1 - \eta$, since one can then set $\delta := 2^{-n}$ and $\eta := \delta/2$ for $n = 1, 2, \ldots$ (for instance) and conclude the claim is true for all δ with positive probability.

Now fix δ, η. Since every bounded non-negative functions f is a convex combination of indicator functions of the form $\mathbf{1}_\Omega$ where $\Omega \in \mathcal{B}_{\varepsilon,\alpha}(G)$, and the number of such functions is $O_{M,\varepsilon}(1)$ (since $\mathcal{B}_{\varepsilon,\alpha}(G)$ only $O_{M,\varepsilon}(1)$ atoms), it suffices (after shrinking η appropriately) to prove the claim for a single indicator function $f := \mathbf{1}_\Omega$.

Fix Ω; we can then write

$$f = \mathbf{1}_\Omega = \mathbf{1}_W(\varepsilon^{-1}G - \alpha)$$

where $W \subset \mathbf{C}$ is some union of $O_X(1)$ translates of the unit cube S.

Let $0 < \sigma \ll 1$ be a small number (it will eventually be much smaller than δ, η, or ε) to be chosen later. Let ∂W_σ be the σ-neighbourhood of the boundary ∂W of W. By Urysohn's lemma combined with the Weierstrass approximation theorem (and the fact that $G = O(M)$) we can write

$$f = P(\varepsilon^{-1}G - \alpha, \overline{\varepsilon^{-1}G - \alpha}) + O(\sigma) + O(\mathbf{1}_{\partial W_\sigma}(\varepsilon^{-1}G - \alpha)) \tag{23}$$

for some polynomial $P = P_{W,M,\varepsilon,\sigma}$ of two complex variables. Denote the first term on the right-hand side of (23) by f_{UAP}, then from UAP^d hypothesis on G and the Banach algebra nature of UAP^d we have

$$\|f_{UAP}\|_{UAP^d} = O_{M,P}(1) = O_{M,\varepsilon,\sigma}(1), \tag{24}$$

which will give (22) once σ is selected properly at the end of the argument.

Now consider the third term on the right-hand side of (23). Observe that

$$\begin{aligned} \|\mathbf{1}_{\partial W_\sigma}(\varepsilon G - \alpha)\|_{L^2}^2 &= \mathbf{P}_{x \in \mathbf{Z}_N}(\varepsilon^{-1}G(x) - \alpha \in \partial W_\sigma) \\ &\leq \mathbf{P}_{x \in \mathbf{Z}_N}(\varepsilon^{-1}G(x) - \alpha \in \partial S_\sigma + \zeta \text{ for some } \zeta \in \mathbf{Z}[i]) \end{aligned}$$

where ∂S_σ is the σ-neighbourhood of the boundary ∂S of the unit square. Observe that as α varies over S, the event

$$\varepsilon^{-1}G(x) - \alpha \in \partial S_\sigma + \zeta \text{ for some } \zeta \in \mathbf{Z}[i]$$

has probability $O(\sigma)$ regardless of what $\varepsilon^{-1}G(x)$ is. We thus have

$$\mathbf{E}(\|\mathbf{1}_{\partial W_\sigma}(\varepsilon G - \alpha)\|_{L^2}^2) \ll \sigma.$$

By Markov's inequality, we thus see that the expression inside the expectation is $O_\eta(\sigma)$ with probability at least $1 - \eta$. Inserting this into (23) we obtain

$$\|f - f_{UAP}\|_{L^2} \ll_\delta \sigma + \sigma^{1/2}$$

and the claim (21) follows by setting σ sufficiently small depending on δ (and then (22) will hold by (24)). $\square$

Henceforth we shall fix an assignment of a factor $\mathcal{B}_\varepsilon(G) = \mathcal{B}_\varepsilon(G, d)$ with the above properties for each almost periodic function $G \in UAP^d$ and each $\varepsilon > 0$. Note that while we did use a randomization argument here, it is possible to make such an assignment constructive, for instance by well-ordering all the factors of $\mathbf{Z}_N$ in some constructive way and then choosing the minimal algebra which obeys the above properties (with the exact choice of bounds in (28), etc. held fixed). Thus we do not require the axiom of choice at this step (or indeed at any step in this argument). For similar reasons we may ensure that this procedure is shift-invariant, in the sense that

$$T^n \mathcal{B}_\varepsilon(G) = \mathcal{B}_\varepsilon(T^n G) \text{ for all } n \in \mathbf{Z}_N, \tag{25}$$

where $T^n\mathcal{B} := \{T^n\Omega : \Omega \in \mathcal{B}\}$ is the factor $\mathcal{B}$ shifted by n. This shift invariance amounts to making sure the same α is chosen for all the shifts T^nG of a fixed function G, which is easy enough to ensure since the constraints needed for α are independent of the choice of n.

The above proposition pertained to a factor generated by a single almost periodic function, but we can easily extend it to algebras generated by multiple functions as follows.

Definition 6.4 (Compact factors). Let $d \geq 0$ and $X \geq 0$. A factor $\mathcal{B}$ is said to be *compact of order d and complexity at most X* if it has the form

$$\mathcal{B} = \mathcal{B}_{\varepsilon_1}(G_1) \vee \ldots \vee \mathcal{B}_{\varepsilon_K}(G_K) \tag{26}$$

for some $0 \leq K \leq X$, some $\varepsilon_1, \ldots, \varepsilon_K \geq \frac{1}{X+1}$, and some $G_1, \ldots, G_K \in UAP^{d-1}$ with norm $\|G_j\|_{UAP^d} \leq X$ for all $1 \leq j \leq K$. We define the *d-complexity* (or simply *complexity*) of a factor to be the minimal X for which one has the above representation, or ∞ if no such representation exists. In particular, the trivial factor $\mathcal{B} = \{\emptyset, \mathbf{Z}_N\}$ is compact of order d with complexity 0.

Remark 6.5. The terminology is motivated here by ergodic theory, see e.g. [15]; a compact factor of order d here corresponds in the ergodic setting, roughly speaking, to a tower of height d of compact extensions of the trivial algebra. The complexity X is a rather artificial quantity which we use as a proxy for keeping all the quantities used to define $\mathcal{B}$ under control.

The key property we need concerning these factors is that the measurable functions of factors which are compact of order d are well approximated by almost periodic functions of order d:

Proposition 6.6 (UAP functions are dense in compact factors). *Let $d \geq 0$, $X \geq 0$, and let $\mathcal{B}$ be a factor which is compact of order d and complexity at most X. Let f be a bounded non-negative function which is measurable with respect to $\mathcal{B}$, and let $\delta > 0$. Then we can find a bounded non-negative $f_{UAP} \in UAP^d$ such that*

$$\|f - f_{UAP}\|_{L^2} \leq \delta \tag{27}$$

and

$$\|f_{UAP}\|_{UAP^d} \ll_{d,\delta,X} 1. \tag{28}$$

Proof. We first verify the claim when f is the indicator function $\mathbf{1}_A$ of an atom A of $\mathcal{B}$. From Definition 6.4 we can expand $\mathcal{B}$ in the form (26), and hence we can write $A = A_1 \cap \ldots \cap A_K$ where each A_j is an atom in $\mathcal{B}_{\varepsilon_j}(G_j)$. From Proposition 6.2 and the bounds on ε_j, G_j, K coming from Definition 6.4, we can find bounded non-negative functions $f_{UAP,j} \in UAP^d$ for all $1 \leq j \leq K$ such that

$$\|\mathbf{1}_{A_j} - f_{UAP,j}\|_{L^2} \leq \delta/K$$

and

$$\|f_{UAP,j}\|_{UAP^d} = O_{d,\delta/K,\varepsilon_j,X}(1) = O_{d,\delta,X}(1).$$

Since $\mathbf{1}_{A_j}$ and $f_{UAP,j}$ are both bounded and non-negative we have the elementary pointwise inequality

$$|\prod_{j=1}^K \mathbf{1}_{A_j} - \prod_{j=1}^K f_{UAP,j}| \leq \sum_{j=1}^K |\mathbf{1}_{A_j} - f_{UAP,j}|$$

and hence if we set $f_{UAP} := \prod_{j=1}^K f_{UAP,j}$ then (27) follows from the triangle inequality, and (28) follows the Banach algebra nature of UAP^d. Since f_{UAP} is clearly bounded and non-negative, the claim follows.

Now suppose f is an arbitrary bounded non-negative function measurable with respect to $\mathcal{B}$. Then we can write $f = \sum_A c_A \mathbf{1}_A$ where A ranges over the atoms of $\mathcal{B}$ and $0 \leq c_A \leq 1$ are constants. Let $\sigma = \sigma(d, \delta, X) > 0$ be a small number to be chosen later, then by the preceding discussion we can find bounded non-negative $f_{UAP,A} \in UAP^d$ for all A such that

$$\|\mathbf{1}_A - f_{UAP,A}\|_{L^2} \leq \sigma$$

and

$$\|f_{UAP,A}\|_{UAP^d} \ll_{d,X,\sigma} 1.$$

If we then set $\tilde{f}_{UAP} := \sum_A c_A f_{UAP,A}$ and observe from Proposition 6.2 that $\mathcal{B}$ contains at most $O_{d,X}(1)$ atoms, we thus have

$$\|f - \tilde{f}_{UAP}\|_{L^2} \ll_{d,X} \sigma$$

and

$$\|\tilde{f}_{UAP}\|_{UAP^d} \ll_{d,X,\sigma} 1.$$

We are however not done yet, because while f_{UAP} is non-negative, it is not bounded by 1; instead we have a bound of the form $0 \leq f_{UAP}(x) \ll_{d,X} 1$. To fix this we need a real-valued polynomial $P(x) = P_{d,\delta,X}(x)$ such that

$$|P(x) - \max(x, 1)| \leq \delta/2 \text{ and } 0 \leq P(x) \leq 1 \text{ for all } x \text{ in the range of } f_{UAP};$$

such a polynomial exists by the Weierstrass approximation theorem. If we then set $f_{UAP} := P(\tilde{f}_{UAP})$, then f_{UAP} is bounded and non-negative, and we have

$$\|f_{UAP} - \max(\tilde{f}_{UAP}, 1)\|_{L^2} \leq \delta/2$$

and (since UAP^d is a Banach algebra)

$$\|f_{UAP}\|_{UAP^d} \ll_{d,\delta,X,\sigma} 1.$$

On the other hand, since f is bounded above by 1, we have

$$\|f - \max(\tilde{f}_{UAP}, 1)\|_{L^2} \leq \|f - \tilde{f}_{UAP}\|_{L^2} \ll_{d,X} \sigma,$$

and the claims then follow from the triangle inequality if σ is chosen sufficiently small depending on d, δ, X. □

7 The energy incrementation argument

The proof of the recurrence theorem (Theorem 3.3) and the structure theorem (Theorem 3.5) relies not only on factors of almost periodic functions, which we constructed in the previous section, but also on the notion of the *energy* of a factor with respect to a collection of functions, and of the recursive *energy incrementation argument* which we will need to prove both the recurrence theorem and the structure theorem. This energy incrementation argument, which was inspired by the proof of the Szemerédi regularity lemma (see e.g. [40]), is perhaps one of the most important aspects of our proof of Szemerédi's theorem, but unfortunately is also the one which causes the Ackermann-type (or worse) blowup of bounds. It is the counterpart of the more well-known *density incrementation argument* which appears in several proofs of Szemerédi's theorem (starting with Roth's original argument [33], but see also [19], [20], [25], [7], [38], [34], [39]). In that strategy one passes from the original set $\{1, \ldots, N\}$ to a decreasing sequence of similarly structured subsets (e.g. arithmetic progressions or Bohr sets) while forcing the density δ of the set A to increase as one progresses along the sequence; eventually one finds enough "randomness" to obtain an arithmetic progression. The hope is to show this algorithm terminates successfully by using the trivial fact that the density is always bounded above by 1. To do this, it is important that the density increment depend only on δ, and not on other parameters such as N or the complexity of the structured subset. This rather stringent requirement on the density increment is one cause of technical complexity and length in several of the arguments mentioned above.

In our situation, the role of "structured subset" will be played by a factor generated by almost periodic functions, and the role of density played by the energy of that factor. This energy will automatically increase as the factor gets finer, and is also automatically bounded. Once again, however, the energy increment may be very small, depending for instance on the complexity of the factor, and this algorithm may once again fail to terminate. This problem also appears in the ergodic theory setting, in the context of an infinite tower of factor extensions; to resolve this one must show that the supremum of any tower of extensions with the recurrence property also has the recurrence property. This appears difficult since the factors in this tower may become arbitrarily complex, and the lower bound obtained by the recurrence property may go to zero as one approaches the supremum of the tower. Nevertheless, one can conclude the argument, basically by observing that any measurable function in the supremum of the tower can be approximated in L^2 norm (say) by a measurable function in some finite component of this tower, and a simple argument then allows one to deduce recurrence for the former function from recurrence from the latter function regardless of how small the recurrence bound is for the latter; see e.g. [15] for an example of this. This may serve to explain why we have the error tolerance (4) in the recurrence theorem.

We begin by defining the *energy* of a factor (relative to some fixed collection of functions); this can be thought of as somewhat analogous to the more standard notion of the *entropy* of an algebra in both information theory and ergodic theory, but the energy will be adapted to a specific fixed collection of functions $f_1, \ldots, f_m$, whereas the entropy is in some sense concerned with all possible functions at once.

Definition 7.1 (Energy). Given a m-tuple $f = (f_1, \ldots, f_m)$ of functions $f_j : \mathbf{Z}_N \to \mathbf{C}$ of functions and a factor $\mathcal{B}$, we define the *energy* $\mathcal{E}_f(\mathcal{B})$ to be the quantity

$$\mathcal{E}_f(\mathcal{B}) := \sum_{j=1}^{m} \|\mathbf{E}(f_j|\mathcal{B})\|_{L^2}^2. \tag{29}$$

In practice m will be very small (either 1 or 3, in fact). Observe that we have the trivial bounds

$$0 \leq \mathcal{E}_f(\mathcal{B}) \leq \sum_{j=1}^{m} \|f_j\|_{L^2}^2. \tag{30}$$

Also from Pythagoras's theorem and the orthogonality considerations discussed above we see that if $\mathcal{B}'$ is finer than $\mathcal{B}$, then

$$\sum_{j=1}^{m} \|\mathbf{E}(f_j|\mathcal{B}') - \mathbf{E}(f_j|\mathcal{B})\|_{L^2}^2 = \mathcal{E}_f(\mathcal{B}') - \mathcal{E}_f(\mathcal{B}). \tag{31}$$

In particular, the energy of $\mathcal{B}'$ is larger than or equal to $\mathcal{B}$.

We now describe, in abstract terms, the idea of the energy increment strategy. Suppose one is trying to prove a statement $P(M)$ involving some large parameter $M > 0$ which one hopes to keep under control; for instance, one may be trying to bound some fixed expression E from above by M or from below by $1/M$. To begin with, this statement does not depend on any factors. But now we introduce a factor $\mathcal{B}$, which we initialize to be the trivial algebra $\mathcal{B} = \{\emptyset, \mathbf{Z}_N\}$, and try to prove $P(M)$ using an argument which is in some sense "relative to $\mathcal{B}$" (in particular, the bound M may depend on some measure of how "complex" $\mathcal{B}$ is). Either this argument works, or it encounters some obstruction. The idea is then to show that the obstruction forces the existence of a new factor $\mathcal{B}'$ which is finer than $\mathcal{B}$ (and typically more complex than $\mathcal{B}$) and has slightly more energy. One then replaces $\mathcal{B}$ by $\mathcal{B}'$ and then repeats the above strategy, hoping to use the trivial bound (30) to show that the argument must eventually work relative to *some* factor.

The difficulty with this strategy is that the energy increment obtained by this method typically depends on the complexity of the factor $\mathcal{B}$, which tends to grow rather quickly.

As such it is possible for this method to get bogged down at some intermediate energy in the range (30) and not terminate in any controlled amount of time. To get around this, it turns out that instead of just using a pair $\mathcal{B} \subset \mathcal{B}'$ of factors, it is better to use a *triplet* $\mathcal{B} \subset \mathcal{B}' \subset \mathcal{B}''$ of factors, with the energy gap between $\mathcal{B}$ and $\mathcal{B}'$ allowed to be moderately large (bounded by a quantity that does not depend on the complexity of any of these algebras). The idea is then to try to prove $P(M)$ relative to the pair $(\mathcal{B}, \mathcal{B}')$, but using bounds which depend only on the complexity of $\mathcal{B}$ and not on $\mathcal{B}'$. If the argument encounters an obstruction, then one can replace $\mathcal{B}'$ by a more complex $\mathcal{B}''$, with an energy increment again depending only on the complexity of $\mathcal{B}$; thus this energy increment will not go to zero as $\mathcal{B}'$ becomes more complex. There is now a second obstruction when the energy gap $\mathcal{B}'$ and $\mathcal{B}$ becomes too large, but then one replaces $\mathcal{B}$ by $\mathcal{B}'$; this can only occur a finite number of times because we do not allow the bounds for this energy gap to depend on the complexity and thus the energy increment here is bounded from below by a fixed constant.

To make this argument more precise we encapsulate it in the following abstract lemma (which has a certain resemblance to Zorn's lemma, and can be in fact thought of as a "quantitative" version of that lemma; it also resembles the proof of the Szemerédi regularity lemma). We are indebted to Ben Green for suggesting the use of this type of energy incrementation argument, which is for instance used in our joint paper [23] to establish arbitrarily long arithmetic progressions in the primes.

Lemma 7.2 (Abstract energy incrementation argument). *Suppose there is a property $P(M)$ which can depend on some parameter $M > 0$. Let $d \geq 0$, and let $f = (f_1, \ldots, f_m)$ be a collection of m bounded functions.*

Suppose also that we have an $\tau > 0$ for which the following dichotomy holds: for any $X, X' > 0$, and given any factor $\mathcal{B}$ which is compact of order d with complexity at most X, and any factor $\mathcal{B}'$ which is finer than $\mathcal{B}$ and also compact of order d with complexity at most X', then if the energy gap condition

$$\mathcal{E}_f(\mathcal{B}') - \mathcal{E}_f(\mathcal{B}) \leq \tau^2, \tag{32}$$

holds, then either $P(M)$ is true for some $M = O_{d,\tau,X,X'}(1)$, or we can find a factor $\mathcal{B}''$ finer than $\mathcal{B}'$ which is compact of order d with complexity at most $O_{d,\tau,X,X'}(1)$ such that we have the energy increment property

$$\mathcal{E}_f(\mathcal{B}'') - \mathcal{E}_f(\mathcal{B}') \gg_{d,\tau,X} 1 \tag{33}$$

(note that the implied constant does not *depend on X').*

Then $P(M)$ is true for some $M = O_{m,d,\tau}(1)$.

Remark 7.3. The point of this lemma is that it reduces the task of proving some property $P(M)$ to the easier task of proving a *dichotomy*; either P can be proven, or we can increment the energy of a certain factor while keeping the complexity under control. It is crucial that τ does not depend on the complexities X, X', and that the energy increment in (33) depends only on the lower complexity X and not the higher complexity X', otherwise this lemma fails. Note that no quantitative knowledge on the growth of this complexity on of the energy increment bound is necessary, although of course the explicit form of the final bound $O_{m,d,\tau}(1)$ on M will depend quite heavily on those growth rates. This argument proceeds by a double iteration and thus typically produces bounds which are of Ackermann type or worse, but in principle they are computable.

Proof. The proof proceeds by running the following double iteration algorithm, constructing a pair of factors $\mathcal{B}$ and $\mathcal{B}'$, both compact of order d and with $\mathcal{B}'$ finer than $\mathcal{B}$, as follows.

Step 1 Initialize $\mathcal{B}$ to the trivial algebra $\mathcal{B} := \{\emptyset, \mathbf{Z}_N\}$.

Step 2 Initialize $\mathcal{B}'$ to equal $\mathcal{B}$ (thus trivially verifying the energy gap condition (32)). Let X denote the complexity of $\mathcal{B}$.

Step 3 Let X' denote the complexity of $\mathcal{B}'$. If $P(M)$ is true for some $M = O_{d,\tau,X,X'}(1)$ then we halt the algorithm. Otherwise, we must by hypothesis be able to locate a factor $\mathcal{B}''$ which is compact of order d with complexity at most $O_{d,\tau,X,X'}(1)$ with the energy increment property (33), and we continue on to Step 4.

Step 4 If $\mathcal{E}_f(\mathcal{B}'') - \mathcal{E}_f(\mathcal{B}) \leq \tau^2$, then we replace $\mathcal{B}'$ by $\mathcal{B}''$ (thus preserving (32)) and return to Step 3. Otherwise, we replace $\mathcal{B}$ by $\mathcal{B}''$ and return to Step 2.

Observe that for each fixed $\mathcal{B}$ of complexity X, the algorithm can only iterate for at most $O_{d,\tau,X}(1)$ times before changing $\mathcal{B}$. This is because every time $\mathcal{B}'$ is changed, the energy $\mathcal{E}_f(\mathcal{B}')$ increases by at least $\gg_{d,\tau,X} 1$, but if the energy ever exceeds $\mathcal{E}_f(\mathcal{B}) + \tau^2$ then we must change $\mathcal{B}$. Note that it is crucial here that the energy increment in (33) not depend on the complexity X' of $\mathcal{B}'$, which may be growing quite rapidly during this iteration process. In particular, if $\mathcal{B}$ finally does change, its complexity will increase from X to at most $O_{d,\tau,X}(1)$. Next, observe that $\mathcal{B}$ can only be changed at most $O_{m,\tau}(1)$ times, because each time we change $\mathcal{B}$, the energy $\mathcal{E}_f(\mathcal{B})$ increases by at least τ^2, but the energy is always non-negative and is bounded by m. Combining these two observations we see that the entire algorithm must halt in $O_{m,d,\tau}(1)$ steps and all factors constructed by the algorithm have complexity at most $O_{m,d,\tau}(1)$. The claim follows. $\square$

8 Proof of the structure theorem

We now prove the structure theorem, Theorem 3.5. Naively, the idea would be to take $\mathcal{B}$ to be the factor formed by *all* the UAP^{d-1} functions, and then take $f_U = f - \mathbf{E}(f|\mathcal{B})$ and $f_{U^\perp} = \mathbf{E}(f|\mathcal{B})$ to conclude the result (the uniformity of f_U arising from Proposition 5.12, and Proposition 6.6 being used to locate f_{UAP}); this would be the exact analogue of how one would proceed in the genuinely ergodic setting when the underlying space is infinite and one does not care about quantitative control on the complexity of the factor. Unfortunately this approach does not work because there are far too many UAP^{d-1} functions available, and the complexity of $\mathcal{B}$ would explode with N (indeed, it is likely that $\mathcal{B}$ would simply be the total factor consisting of arbitrary subsets of $\mathbf{Z}_N$). Thus, in the quantitative setting, one must be substantially more "choosy" about which UAP^{d-1} functions to admit into the algebra $\mathcal{B}$ - they should only be the ones which have a good reason for being there, such as having a non-trivial correlation with the function f. It turns out that the best framework for doing this is given by the abstract energy incrementation argument given in the previous section, exploiting the fact that each function that one adds to the σ algebra increases the energy of that algebra, especially if there is a correlation with f.

In view of the energy incrementation argument, it suffices to prove the following dichotomy:

Lemma 8.1 (Structure theorem dichotomy). *Let $k \geq 3$, and let f be a non-negative bounded function obeying* (2) *for some $\delta > 0$. Let $F : \mathbf{R}^+ \to \mathbf{R}^+$ be a function. Let $\mathcal{B} \subset \mathcal{B}'$ be factors which are compact of order $k-2$ with complexity at most X, X' respectively, and and obey the energy gap condition* (32) *with $\tau := \frac{\delta^2}{5000k}$. Then at least one of the following must be true:*

- *(Success) We can find a positive number $M = O_{k,\delta,X}(1)$ a bounded function f_U, and non-negative bounded functions $f_{U^\perp}$, f_{UAP} such that we have the splitting $f = f_U + f_{U^\perp}$ and the estimates* (4)*,* (5)*,* (6) *with $d := k-2$, as well as the Gowers uniformity estimate* (8)*.*
- *(Energy increment) We can find a factor $\mathcal{B}''$ finer than $\mathcal{B}'$ which is compact of order $k-2$ and complexity $O_{k,\delta,X,X'}(1)$ such that*

$$\mathcal{E}_f(\mathcal{B}'') - \mathcal{E}_f(\mathcal{B}') \gg_{k,\delta,X,F} 1. \tag{34}$$

Indeed, Theorem 3.5 follows immediately by applying Lemma 8.1 to Lemma 7.2 (using $m = 1$ and using the bounded function f, and $\tau := \frac{\delta^2}{5000k}$).

Proof of Lemma 8.1. Fix $\mathcal{B}$, $\mathcal{B}'$. Since $\mathbf{E}(f|\mathcal{B})$ is non-negative and bounded, and $\mathcal{B}$ is compact of order $k-2$ with complexity $O(X)$, We may apply Proposition 6.6 to find a non-negative bounded function f_{UAP} such that

$$\|\mathbf{E}(f|\mathcal{B}) - f_{UAP}\|_{L^2} \leq \frac{\delta^2}{5000k} \tag{35}$$

and

$$\|f_{UAP}\|_{UAP^{k-2}} < M$$

for some $M = O_{k,\delta,X}(1)$, which we now fix. From (35) and Cauchy-Schwarz we observe that

$$\left|\int_{\mathbf{Z}_N} f - \int_{\mathbf{Z}_N} f_{UAP}\right| = \left|\int_{\mathbf{Z}_N} \mathbf{E}(f|\mathcal{B}) - f_{UAP}\right| \leq \frac{\delta^2}{5000k}$$

and in particular (by (2))

$$\int_{\mathbf{Z}_N} f_{UAP} \geq \delta/2.$$

Now split $f = f_U + f_{U^\perp}$, where $f_{U^\perp} := \mathbf{E}(f|\mathcal{B}')$ and $f_U := f - \mathbf{E}(f|\mathcal{B}')$. We have already proven the estimates (5), (6), while (4) follows from (32) (recall $\tau = \frac{\delta^2}{5000k}$), (31), (35), and the triangle inequality. If the estimate (8) held then we would now be done (in the "Success" half of the dichotomy), so suppose instead that

$$\|f_U\|_{U^{k-1}} \geq 1/F(M).$$

By Lemma 5.12, we can thus find a function $G \in UAP^{k-2}$ with $\|G\|_{UAP^{k-2}} \leq 1$ such that

$$|\langle f_U, G\rangle| \gg_{k,\delta,M,F} 1. \tag{36}$$

Now write $\mathcal{B}'' := \mathcal{B}' \vee \mathcal{B}_\varepsilon(G)$, where $\varepsilon = \varepsilon(k, \delta, M) > 0$ is to be chosen later. We thus split $f_U = (f - \mathbf{E}(f|\mathcal{B}'')) + (\mathbf{E}(f|\mathcal{B}'') - \mathbf{E}(f|\mathcal{B}'))$ and $G = (G - \mathbf{E}(G|\mathcal{B}'')) - (\mathbf{E}(G|\mathcal{B}'') - \mathbf{E}(G|\mathcal{B}')) + \mathbf{E}(G|\mathcal{B}')$. The first terms in both expansions are orthogonal to $\mathcal{B}''$ (and thus to $\mathcal{B}'$), while the second terms are measurable in $\mathcal{B}''$ and orthogonal to $\mathcal{B}'$, while the third term of G is measurable in $\mathcal{B}'$. Thus

$$\langle f_U, G\rangle = \langle f - \mathbf{E}(f|\mathcal{B}''), G - \mathbf{E}(G|\mathcal{B}'')\rangle + \langle \mathbf{E}(f|\mathcal{B}'') - \mathbf{E}(f|\mathcal{B}'), \mathbf{E}(G|\mathcal{B}'') - \mathbf{E}(G|\mathcal{B}')\rangle.$$

From (20) and the boundedness of f we have

$$|\langle f - \mathbf{E}(f|\mathcal{B}''), G - \mathbf{E}(G|\mathcal{B}'')\rangle| \ll \varepsilon.$$

Thus if we choose ε sufficiently small depending on k, δ, M, we see from (36) that

$$|\langle \mathbf{E}(f|\mathcal{B}'') - \mathbf{E}(f|\mathcal{B}'), \mathbf{E}(G|\mathcal{B}'') - \mathbf{E}(G|\mathcal{B}')\rangle| \gg_{k,\delta,M,F} 1.$$

Since G is bounded, we thus see from Cauchy-Schwarz that

$$\|\mathbf{E}(f|\mathcal{B}'') - \mathbf{E}(f|\mathcal{B}')\|_{L^2} \gg_{k,\delta,M,F} 1.$$

But then (34) follows from (31). Finally, the complexity bound on $\mathcal{B}''$ follows from Definition 6.4, the complexity bound on $\mathcal{B}'$, and the choice of ε and M. We are thus in the "Energy increment" half of the dichotomy, and the lemma follows. $\square$

The proof of the structure theorem is now complete.

Remark 8.2. It may be possible to prove this structure theorem more directly, without explicitly invoking factors, for instance by setting up a extremization problem such as that of minimizing the U^{k-1} norm of f_U subject to the constraints (4), (5), (6), the splitting $f = f_U + f_{U^\perp}$, and the bounded non-negativity of f_{UAP} and $f_{U^\perp}$. We were unable however to achieve this in a clean way, especially when it came to maintaining the boundedness and non-negativity conditions, whereas the conditional expectation method achieves this more painlessly.

9 Compactness on atoms, and an application of van der Waerden's theorem

To prove Szemerédi's theorem, the only thing that now remains is to prove the recurrence theorem for almost periodic functions, Theorem 3.3. In this section we present a key Proposition, which illustrates the applicability of van der Waerden's theorem (Theorem 1.1) to the problem of obtaining recurrence for a function f whose shifts $T^n f$ enjoy a representation such as (15). The key idea is that the functions on the right-hand side of (15) live in a sufficiently "compact" space of functions that they can be "finitely coloured", at which point van der Waerden's theorem can be used to establish recurrence[12]. As we show at the end of this section, we can quickly use this Proposition to deduce the $d = 1$ case (as well as the rather trivial $d = 0$ case) of Theorem 3.3 as a corollary.

Proposition 9.1 (Recurrence for conditionally UAP functions). *Let $\mathcal{B}$ be a factor, let $M > 0$, let H be a finite non-empty set, and let for each $n \in \mathbf{Z}_N$ and $h \in H$ let $c_{n,h}$ be a bounded $\mathcal{B}$-measurable function and let g_h be a bounded function. Let h be a random variable taking values in H, and define the functions F_n for all $n \in \mathbf{Z}_N$ by the formula*

$$F_n := M\mathbf{E}(c_{n,h} g_h) \tag{37}$$

(compare with (15)). Let $f_{U^\perp}$ be a bounded non-negative function, and for any $\delta > 0$, $n \in \mathbf{Z}_N$, and $k, k_* \in \mathbf{Z}_+$ let $E_n(k, \delta, k_*, \mathcal{B}) \in \mathcal{B}$ be the set

$$\begin{aligned} E_n(k, \delta, k_*, \mathcal{B}) := \{x \in \mathbf{Z}_N : \mathbf{E}(T^n f_{U^\perp} | \mathcal{B})(x) \geq \frac{\delta}{2} \text{ and} \\ \mathbf{E}(|T^n f_{U^\perp} - F_n||\mathcal{B})(x) \leq \frac{\delta}{8k}\}. \end{aligned} \tag{38}$$

Then for every $\delta > 0$ and $k \in \mathbf{Z}_+$ there exists $k_* = k_*(k, \delta, M)$ such that

$$\begin{aligned} &\mathbf{E}_{1 \leq r \leq N_0} \int_{\mathbf{Z}_N} \prod_{j=0}^{k-1} T^{\mu j r} f_{U^\perp} \\ &\qquad \gg_{k,\delta,M} \mathbf{E}_{1 \leq \lambda \leq N_0/k_*} \mathbf{P}_{\mathbf{Z}_N} \bigcap_{m=1}^{k_*} E_{\mu\lambda m}(k, \delta, k_*, \mathcal{B}) \end{aligned} \tag{39}$$

for all $\mu \in \mathbf{Z}_N$ and $N_0 \geq k_*$.

Remark 9.2. The point is that this theorem reduces the task of establishing lower bounds for recurrence expressions involving $f_{U^\perp}$, to that of establishing lower bounds for the

[12]This argument was inspired, not by the original ergodic theory arguments of Furstenberg, but of the later colouring-based arguments, for instance in [3]. It may be possible to adapt the older arguments in, say, [15] instead here, which have the advantage of using the same length k for the progression throughout the argument, instead of replacing k by a considerably larger k_* as is done here. This might ultimately lead to somewhat better final bounds, although it still seems that one would still get Ackermann-type dependence or worse on the constants.

recurrence behaviour of $\mathcal{B}$-measurable sets $E_{\mu\lambda m}(k,\delta,k_*,\mathcal{B})$. This is advantageous if $\mathcal{B}$ is "simpler" than the original function $f_{U^\perp}$; in practice, $f_{U^\perp}$ will be approximately an almost periodic function of order d, and $\mathcal{B}$ will be a compact σ algebra of order $d-1$, and thus functions measurable in $\mathcal{B}$ can be approximated by almost periodic functions of one lower order than d. This is the key to the proof of Theorem 3.3 we give in the next section, which proceeds by induction on d. On the other hand, the bounds on k_* given by our proof involve van der Waerden numbers, which will cause Ackermann type growth rates or worse in our final bound.

Proof of Proposition 9.1. To prove (39) it suffices to prove the "localized" version

$$\mathbf{E}_{1\leq a,s\leq k_*/k}\int_{\mathbf{Z}_N}\prod_{j=0}^{k-1}T^{\mu\lambda(a+js)}f_{U^\perp}\gg_{\delta,k,k_*}\mathbf{P}_{\mathbf{Z}_N}\bigcap_{m=1}^{k_*}E_{\mu\lambda m}(k,\delta,k_*,\mathcal{B}) \tag{40}$$

for each $1\leq\lambda\leq N_0/k_*$. Indeed, if (40) held then upon averaging in λ we obtain

$$\mathbf{E}_{1\leq\lambda\leq N_0/k_*;1\leq a,s\leq k_*/k}\int_{\mathbf{Z}_N}\prod_{j=0}^{k-1}T^{\mu\lambda(a+js)}f_{U^\perp}$$
$$\gg_{\delta,k,k_*}\mathbf{E}_{1\leq\lambda\leq N_0/k_*}\mathbf{P}_{\mathbf{Z}_N}\bigcap_{m=1}^{k_*}E_{\mu\lambda m}(k,\delta,k_*,\mathcal{B}).$$

The $T^{\mu\lambda a}$ can be factored out of the product and makes no difference to the expectation, thus it can be discarded. The a averaging then becomes redundant, and we obtain

$$\mathbf{E}_{1\leq\lambda\leq N_0/k_*;1\leq s\leq k_*/k}\int_{\mathbf{Z}_N}\prod_{j=0}^{k-1}T^{\mu\lambda js}f_{U^\perp}$$
$$\gg_{\delta,k,k_*}\mathbf{E}_{1\leq\lambda\leq N_0/k_*}\mathbf{P}_{\mathbf{Z}_N}\bigcap_{m=1}^{k_*}E_{\mu\lambda m}(k,\delta,k_*,\mathcal{B}).$$

The claim (39) then follows by observing that every $1\leq r\leq N_0$ has at most $O_{k_*,k}(1)$ representations of the form $r=\lambda(a+js)$ with $1\leq\lambda\leq N_0/k_*$ and $1\leq a,s\leq k_*/k$. (The dependence of k_* is ultimately irrelevant since k_* itself will ultimately depend on δ,k,M).

It remains to prove (40). Fix μ, λ. By absorbing μ into λ we may take $\mu=1$ (we will not use the upper or lower bound on λ). Since $\bigcap_{m=1}^{k_*}E_{\lambda m}(k,\delta,k_*,\mathcal{B})$ is measurable in $\mathcal{B}$, it is the union of atoms $A\in\mathcal{B}$. It will suffice to prove the pointwise estimate

$$\mathbf{E}_{1\leq a,s\leq k_*/k}\mathbf{E}_A\prod_{j=0}^{k-1}T^{\lambda(a+js)}f_{U^\perp}\gg_{\delta,k,k_*}1$$

for each such atom, as the claim then follows by multiplying this formula by $\mathbf{P}_{\mathbf{Z}_N}A$ and summing over all atoms in $E_\lambda(k,\delta,k_*,\mathcal{B})$.

Now fix the atom A. Since the number of pairs (a,s) is $O_{k,k_*}(1)$, it suffices to locate a *single* pair (a,s) with $1 \leq a, s \leq k_*/k$ such that

$$\mathbf{E}_A \prod_{j=0}^{k-1} T^{\lambda(a+js)} f_{U^\perp} \gg_{\delta,k} 1. \tag{41}$$

We now pass from the shifts $T^{\lambda(a+js)} f_{U^\perp}$ to the functions $F_{\lambda(a+js)}$. We claim that to prove (41) it would suffice to prove that

$$\|F_{\lambda(a+js)} - F_{\lambda a}\|_{L^2(A)} \leq \frac{\delta}{8k} \text{ for all } 0 \leq j \leq k-1, \tag{42}$$

where $L^2(A)$ is the Hilbert space given by the norm $\|F\|_{L^2(A)} := (\mathbf{E}_A |F|^2)^{1/2}$. To see this claim, observe from Cauchy-Schwarz that (42) implies

$$\mathbf{E}_A |F_{\lambda(a+js)} - F_{\lambda a}| \leq \frac{\delta}{8k} \text{ for all } 0 \leq j \leq k-1.$$

But by (38) and the choice of A, we also have

$$\mathbf{E}_A |F_{\lambda(a+js)} - T^{\lambda(a+js)} f_{U^\perp}| \leq \frac{\delta}{8k} \text{ for all } 0 \leq j \leq k-1.$$

From the triangle inequality we thus have

$$\mathbf{E}_A |T^{\lambda(a+js)} f - T^{\lambda a} f_{U^\perp}| \leq \frac{3\delta}{8k} \text{ for all } 0 \leq j \leq k-1.$$

This in particular implies

$$\mathbf{E}_A \mathbf{1}_{|T^{\lambda(a+js)} f_{U^\perp} - T^{\lambda a} f_{U^\perp}| > \frac{4}{5} T^{\lambda a} f} T^{\lambda a} f \leq \frac{15\delta}{32k} \text{ for all } 0 \leq j \leq k-1.$$

On the other hand, by (38) again we have

$$\mathbf{E}_A T^{\lambda a} f_{U^\perp} \geq \delta/2.$$

Thus

$$\mathbf{E}_A \mathbf{1}_{|T^{\lambda(a+js)} f_{U^\perp} - T^{\lambda a} f_{U^\perp}| \leq \frac{4}{5} T^{\lambda a} f \text{ for all } 0 \leq j \leq k-1} T^{\lambda a} f_{U^\perp} \geq \frac{\delta}{32k}.$$

By Hölder's inequality and the non-negativity of $f_{U^\perp}$ this implies that

$$\mathbf{E}_A \mathbf{1}_{|T^{\lambda(a+js)} f_{U^\perp} - T^{\lambda a} f_{U^\perp}| \leq \frac{4}{5} T^{\lambda a} f_{U^\perp} \text{ for all } 0 \leq j \leq k-1} (T^{\lambda a} f_{U^\perp})^k \geq (\frac{\delta}{32k})^k.$$

The claim (41) then follows from the elementary pointwise inequality

$$\prod_{j=0}^{k-1} T^{\lambda(a+js)} f_{U^\perp} \geq \frac{1}{5^k} \mathbf{1}_{|T^{\lambda(a+js)} f_{U^\perp} - T^{\lambda a} f_{U^\perp}| \leq \frac{4}{5} T^{\lambda a} f_{U^\perp} \text{ for all } 0 \leq j \leq k-1} (T^{\lambda a} f_{U^\perp})^k.$$

It remains to find a pair (a, s) obeying (42). Using (37) it suffices to find an (a, s) such that

$$\|\mathbf{E}(c_{\lambda(a+js),h} g_h) - \mathbf{E}(c_{\lambda a,h} g_h)\|_{L^2(A)} \leq \frac{\delta}{8Mk} \text{ for all } 0 \leq j \leq k-1. \tag{43}$$

Note that as the $c_{n,h}$ are measurable with respect to $\mathcal{B}$, they are constant on A, and so without loss of generality we can treat them just as bounded complex numbers (this is the whole point of working on individual atoms in the first place). The g_h are not constant, but we can think of them as bounded functions on A.

To proceed further we need the following compactness property of averages of the form $\mathbf{E}(c_h g_h)$ in $L^2(A)$.

Lemma 9.3 (Total boundedness property). *There exists integers $1 \leq m_1, \ldots, m_L \leq k_*$ for some $L \ll_{k,M,\delta} 1$ such that*

$$\inf_{1 \leq l \leq L} \|\mathbf{E}(c_{\lambda m,h} g_h) - \mathbf{E}(c_{\lambda m_l,h} g_h)\|_{L^2(A)} \leq \frac{\delta}{16Mk} \textit{ for all } 1 \leq m \leq k_*.$$

Remark 9.4. The key point here is that the bound on L does not depend on the size of H, A, or N. This is a quantitative analogue of the basic result (used in the ergodic theory proofs, see e.g. [15]) that a Volterra integral operator from one finite measure space to another is necessarily a compact operator in L^2, and thus the range of any bounded set can be covered by a finite number of δ-balls in L^2.

Proof. Let us write $f_m := \mathbf{E}(c_{m,h} g_h)|_A$. We construct an orthonormal system of functions $v_1, v_2, \ldots, v_J$ in $L^2(A)$ by performing the following algorithm, which can be viewed as a rudimentary version of the energy increment algorithm discussed in previous sections (with the role of factors replaced by the simpler notion of finite-dimensional subspaces of a Hilbert space).

Step 0 Initialize $J = 0$.

Step 1 Let $V \subset L^2(A)$ be the subspace spanned by the $v_1, \ldots, v_J$ (so initially this will be the trivial space $\{0\}$).

Step 2 If there exists a $1 \leq m \leq k_*$ such that $\mathrm{dist}_{L^2(A)}(f_m, V) \geq \delta/64Mk$, then by Hilbert space geometry we can find a unit vector v_{J+1} orthogonal to V (and thus to all the $v_1, \ldots, v_J$) such that $|\langle f_m, v_{J+1}\rangle_{L^2(A)}| \geq \delta/64Mk$. In such a case, we choose[13] such a v_{J+1}, increment J, and return to Step 1. Otherwise, we terminate the algorithm.

We claim that this algorithm terminates in $O_{k,M,\delta}(1)$ steps. Indeed, for each v_j generated by this algorithm, we see from construction that there exists an $m = m(j) \in \mathbf{Z}_N$ such that

$$|\mathbf{E}(c_{\lambda m,h} \langle g_h, v_j\rangle_{L^2(A)})| = |\langle f_m, v_j\rangle_{L^2(A)}| \geq \delta/64Mk.$$

[13]Note that since m ranges over a finite set, the axiom of choice is not needed here, since $\mathbf{Z}_N$ is clearly well-ordered. Because we are always in a finite (or at least finite dimensional) setting, similar considerations apply to other parts of the argument in which an arbitrary choice has to be made.

Here we have crucially taken advantage of the fact that $c_{\lambda m,h}$ is constant on A. Since $c_{\lambda m,h}$ is bounded, we thus see from the Cauchy-Schwarz inequality that

$$\mathbf{E}(|\langle g_h, v_j\rangle_{L^2(A)}|^2) \geq (\frac{\delta}{64Mk})^2.$$

Summing this in j, we obtain

$$\mathbf{E}(\sum_{j=1}^{J} |\langle g_h, v_j\rangle_{L^2(A)}|^2) \geq (\frac{\delta}{64Mk})^2 J.$$

But from the boundedness of the g_h, the orthonormality of the v_j, and Bessel's inequality, the left-hand side is at most 1. Thus $J \leq (\frac{64Mk}{\delta})^2 = O_{k,M,\delta}(1)$ as claimed.

Now observe from the construction of the algorithm that all the functions f_m will lie within $\delta/64Mk$ (in the $L^2(A)$ metric) of the J-dimensional space V. In particular, we see from the triangle inequality, the crude bound $\|f_m\|_{L^2(A)} \leq 1$ arising from our bounds on $c_{n,h}$ and g_h, and finite-dimensional geometry that there can be at most $O_{k,\delta,J}(1) = O_{k,M,\delta}(1)$ functions $f_{m_1}, \ldots, f_{m_L}$ which are all separated from each other by at least $\delta/16Mk$ in the $L^2(A)$ metric. The claim now follows by the usual greedy algorithm. □

Using this lemma, we can introduce a colouring function $\mathbf{c} : \{1, \ldots, k_*\} \to \{1, \ldots, L\}$ by

$$\mathbf{c}(m) := \inf\{1 \leq l \leq L : \|\mathbf{E}(c_{\lambda m,h} g_h) - \mathbf{E}(c_{\lambda m_l,h} g_h)\|_{L^2(A)} \leq \delta/16Mk\}.$$

By van der Waerden's theorem, if $k_* = k_*(k, L) = k_*(k, \delta, M)$ is chosen sufficiently large, then we can find $1 \leq a, s \leq k_*/k$ such that the progression $a, a+s, \ldots, a+(k-1)s$ is monochromatic. The claim (43) now follows from the triangle inequality. This concludes the proof of Proposition 9.1. □

As a quick corollary of this Proposition we can now prove the $d = 1$ case, at least, of Theorem 3.3.

Proof of Theorem 3.3 when $d = 1$. Let $f_{U^\perp}$, f_{UAP}, k, M, δ, ϵ be as in the Theorem. From (6) and Definition 5.2 we can find a finite non-empty set H, a collection of bounded constants $(c_{n,h})_{n\in\mathbf{Z}_N; h\in H}$, and bounded functions $(g_h)_{h\in H}$, and a random variable h taking values in H such that we have the representation (37), where $F_n := T^n f_{UAP}$. We thus apply Proposition 9.1 with $N_0 := N_1$ and $\mathcal{B}$ set equal to the trivial factor $\mathcal{B} = \{\emptyset, \mathbf{Z}_N\}$, since the $c_{n,h}$ are all almost periodic of order 0 and hence constant. But by (38), we see that $E_{\lambda m}(k, \delta, k_*, \mathcal{B})$ is either the empty set or all of $\mathbf{Z}_N$, with the latter occuring if

$$\int_{\mathbf{Z}_N} T^{\lambda m} f_{U^\perp} \geq \delta/2 \text{ and } \int_{\mathbf{Z}_N} |T^{\lambda m} f - F_{\lambda m}| \leq \frac{\delta}{8k}.$$

But the latter condition is automatic from (5), while the latter follows from (4), Cauchy-Schwarz, and the choice of ϵ; note that the shift $T^{\lambda m}$ has no effect on the integral $\int_{\mathbf{Z}_N}$. Thus $E_{\lambda m}(k, \delta, k_*, \mathcal{B}) = \mathbf{Z}_N$ for all λ, and the claim (7) follows from (39). □

Remark 9.5. As we shall see, the $d > 1$ case is somewhat more complicated, the problem being that one has to somehow "quotient out" the effect of a very large number of almost periodic functions of order $d - 1$ before the property of being almost periodic of order d emerges as a usable property. This appears to unfortunately be rather necessary, even when $d = 2$, at least with the arguments currently available; the author would consider this issue one of the least well understood components of the theory. Consider for instance a function f_{AP} of the form $f_{AP}(x) = \psi(x^2/N)$, where $\psi : \mathbf{R}/Z \to [0, 1]$ is a smooth bounded non-negative function which is periodic with period 1, which equals 1 on the interval $[-\delta, \delta]$, and vanishes outside $[-2\delta, 2\delta]$. This function can be shown to be almost periodic of order 2 with an UAP^2 norm of $O_\delta(1)$. Thus Theorem 3.3 should allow us to locate a large number of arithmetic progressions of length k in the support of ψ, for reasonably large values of k (e.g. $k = 5$). To actually establish even this special case, however, seems rather difficult, the simplest proof probably being the ergodic theory proof that lifts this problem up to establishing recurrence for the skew shift on the two-dimensional torus. Similarly for more complicated examples such as (17) (now the ergodic system is a two-step nilsystem, formed by quotienting the unipotent upper triangular 3×3 matrices by the subgroup of matrices with integer coefficients). In [20] this precise problem was encountered, and solved by using very directly the number-theoretic structure of x^2/N (and similarly polynomial objects), in particular a quantitative version of Weyl's theorem on the uniform distribution of polynomials. The problem of having to deal with generalized polynomials instead of polynomials was avoided by working on relatively short arithmetic progressions, in which one could approximate the former by the latter.

10 Recurrence for almost periodic functions

We now conclude the proof of Theorem 3.3, and thus of Theorem 1.2. We have already handled the $d = 1$ case. The case $d = 0$ can either be deduced from the $d = 1$ case, or can be worked out directly by an easy argument which we leave to the reader (the point being that f_{UAP} is now constant and $f_{U^\perp}$ and its shifts will have to be larger than, say, $\delta/2$ with very high probability, say at least $1 - 1/2k$). Thus it remains to handle the $d > 1$ cases. We may assume as an inductive hypothesis that d is fixed and the claim has already been proven for $d - 1$.

When $\mu = 0$ the claim follows easily from (5) and the boundedness of $f_{U^\perp}$, so we may take $\mu \neq 0$. But then we may rescale by μ and set $\mu = 1$.

We would like to apply Proposition 9.1 as we did in the $d = 1$ case. The difficulty now is that the functions $c_{n,h}$ generated by Definition 5.2 are no longer constant, but are themselves almost periodic of one lower order, $d - 1$. The strategy is then to locate a factor $\mathcal{B}$ generated by such functions (and hence compact of order $d - 1$) with respect to which the $c_{n,h}$ are close to being measurable (i.e. close to constant on most atoms). Proposition 9.1 then allows us to reduce the problem of establishing recurrence for $f_{U^\perp}$ to one of establishing a property very similar to recurrence for certain subsets of $\mathcal{B}$, which we can handle by combining the induction hypothesis with Proposition 6.6. As with the structure theorem, one would naively want to take $\mathcal{B}$ to be the factor generated by *all* the $c_{n,h}$ (and this is indeed what one does in the genuinely ergodic setting), but again we lose control of the complexity this way. Instead we must be much more selective with which $c_{n,h}$ we admit. Again, the easiest framework to implement this idea is given by the abstract energy increment lemma, Lemma 7.2. The point is that it may happen that the $c_{n,h}$ are refusing to be close to measurable on $\mathcal{B}$, or that other problems arise such as $\mathcal{B}$ failing to be sufficiently "shift-invariant" (this issue arose in the $d = 1$ case when one needed to eliminate the $T^{\lambda m}$ shift, although in that case the resolution to the problem was trivial). In that case, however, the simplest solution is to replace $\mathcal{B}$ by a larger factor $\mathcal{B}'$, to which one adds in all the obstructions (or at least a representative sample thereof) which one encountered in closing the argument, thus increasing the energy of $\mathcal{B}$.

We turn to the details. It will suffice to establish

Proposition 10.1 (Recurrence theorem dichotomy). *Let $d \geq 2$, and suppose that Theorem 3.3 has already been proven for $d-1$. Let $k \geq 1$ be integers, and let $M, \delta > 0$. All quantities in what follows can depend on d, δ, k, M (including the implicit bounds in $O()$ notation), and we omit future dependence on these parameters. let $f_{U^\perp}, f_{UAP}$ be non-negative bounded functions obeying the bounds* (4), (5), (6). *Write $f := (f_{U^\perp}, |f_{U^\perp} - f_{UAP}|, |f_{U^\perp} - f_{UAP}|^2)$. Let $\mathcal{B} \subset \mathcal{B}'$ be factors which are compact of order $k-2$ with complexity at most X, X' respectively, and such that* (32) *holds for some small $\tau > 0$ independent of X, X' to be chosen later. Then at least one of the following must be true:*

- *(Success) We have*

$$\mathbf{E}_{0 \leq r \leq N_1} \int_{\mathbf{Z}_N} \prod_{j=0}^{k-1} T^{jr} f_{U^\perp} \gg_{\tau,X} 1 \tag{44}$$

  for all $N_1 \geq 0$.

- *(Energy increment) We can find a factor $\mathcal{B}''$ finer than $\mathcal{B}'$ which is compact of order d and complexity $O_{\tau,X,X'}(1)$ such that*

$$\mathcal{E}_f(\mathcal{B}'') - \mathcal{E}_f(\mathcal{B}') \gg_{\tau,X} 1. \tag{45}$$

  Note that the implied constant in (45) *is independent of X'.*

Indeed, Theorem 3.3 will follow from this Lemma and Lemma 7.2 (setting $m = 3$).

It remains to prove Proposition 10.1. We will aim towards applying Proposition 9.1, by locating a large subset of $\mathcal{B}'$ where $f_{U^\perp}$ (and several of its shifts) are large on average, f_{UAP} is close to $f_{U^\perp}$ on average (as are various shifts of these functions), and the $c_{n,h}$ are close to constant, and then using the induction hypothesis to obtain lower bounds on the sets $E_{\lambda\mu}$ obtained this way. There may be some obstructions to implementing this strategy, but when they arise we will convert those obstructions to an energy increment, establishing (45) instead of (44).

Proof of Lemma 10.1. By (6) and Definition 5.2 we can find a finite non-empty set H, a collection of bounded functions $(c_{n,h})_{n \in \mathbf{Z}_N; h \in H}$ in UAP^{d-1} with $\|c_{n,h}\|_{UAP^{d-1}} \leq 1$, and bounded functions $(g_h)_{h \in H}$, and a random variable h taking values in H such that we have the representation

$$T^n f_{UAP} = M\mathbf{E}(c_{n,h} g_h) \tag{46}$$

for all $n \in \mathbf{Z}_N$. We cannot yet apply Proposition 9.1 since the $c_{n,h}$ are not necessarily measurable with respect to $\mathcal{B}'$; indeed there are too many of the $c_{n,h}$ to safely add all of them to $\mathcal{B}'$, which needs to have bounded complexity. Instead, we shall work using much smaller batches of $c_{n,h}$ and then average at the end.

We will need a large integer $N_0 = N_0(\tau, X) > 1$ to be chosen later[14]. If $N_1 \leq N_0$ then the claim (44) follows easily from (5) just by considering the $r = 0$ component of the left-hand side, so we will assume $N_1 > N_0$. We then observe that

$$\mathbf{E}_{0 \leq r \leq N_1} \int_{\mathbf{Z}_N} \prod_{j=0}^{k-1} T^{jr} f_{U^\perp} \gg_{N_0} \mathbf{E}_{1 \leq \mu \leq N_1/N_0} \mathbf{E}_{1 \leq r \leq N_0} \int_{\mathbf{Z}_N} \prod_{j=0}^{k-1} T^{\mu j r} f_{U^\perp} \tag{47}$$

because each $0 \leq r \leq N_0$ has at most $O_{N_0}(1)$ representations of the form $r = \mu r'$ where $1 \leq r' \leq N_0$ and $1 \leq \mu \leq N_1/N_0$.

Now fix a single $1 \leq \mu \leq N_1/N_0$, and consider the expression

$$\mathbf{E}_{1 \leq r \leq N_0} \int_{\mathbf{Z}_N} \prod_{j=0}^{k-1} T^{\mu j r} f_{U^\perp}. \tag{48}$$

Observe that the exponents $\mu j r$ now range in the relatively small set $\mu \cdot \{0, \ldots, (k-1)N_0\}$. This has localized the "n" index in (46) to a reasonably bounded set (one which is independent of N), but the "h" parameter is still ranging over a potentially unbounded set H. To resolve this we need the following variant of Lemma 9.3.

Lemma 10.2 (Finite-rank approximation). *Let $\mu \in \mathbf{Z}_N$. We can find $h_1, \ldots, h_{N_0^{100}} \in H$ (not necessarily distinct, and depending on μ) such that*

$$\|\mathbf{E}(c_{\mu m,h} g_h) - \mathbf{E}_{1 \leq j \leq D}(c_{\mu m,h_j} g_{h_j})\|_{L^2} \ll N_0^{-40} \tag{49}$$

for all $0 \leq m \leq (k-1)N_0$.

Proof. We use the second moment method. Set $D := N_0^{100}$, $G_h := c_{\mu m,h} g_h$, $F := \mathbf{E}(G_h)$, and let $h_1, \ldots, h_D$ be D independent samples of the random variable h. We will show that

$$\mathbf{P}(\|F - \mathbf{E}_{1 \leq j \leq D} G_{h_j}\|_{L^2} > \frac{(kN_0)^{1/2}}{N_0^{50}}) \leq \frac{1}{kN_0},$$

[14] There are a number of parameters involved here, which are at several different scales. In order to have some idea of what parameters should be large and what parameters should be small, we suggest using the hierarchy

$$d, k, \frac{1}{\delta}, M, k_* \ll \frac{1}{\tau} \ll X \ll N_0 \ll X' \ll N_1, N, |H|$$

which is a very typical arrangement of the parameters. The key points are that the energy gap τ does not depend on the large parameters $X, N_0, X', N_1, N, |H|$, that the energy increment in (34) does not depend on the very large parameters $X', N_1, N, |H|$, and the remaining bounds do not depend on the extremely large parameters $N_1, N, |H|$.

which implies the claim with probability at least $1 - \frac{(k-1)N_0+1}{kN_0} > 0$.

By Chebyshev's inequality, it will suffice to show that

$$\mathbf{E}(\|F - \mathbf{E}_{1\leq j\leq D} G_{h_j}\|_{L^2}^2) \leq 1/D.$$

The left-hand side can be expanded as

$$\mathbf{E}(\int_{\mathbf{Z}_N} |F(x)|^2 - 2\Re \mathbf{E}_{1\leq j\leq D}(\overline{F} G_{h_j}) + |\mathbf{E}_{1\leq j\leq D} G_{h_j}|^2)$$

which we expand and rearrange further as

$$\|F\|_{L^2}^2 - 2\Re \mathbf{E}_{1\leq j\leq D} \int_{\mathbf{Z}_N} \overline{F}\mathbf{E}(G_{h_j}) + \mathbf{E}_{1\leq j,j'\leq D} \int_{\mathbf{Z}_N} \mathbf{E}(\overline{G_{h_{j'}}} G_{h_j}). \tag{50}$$

Since $h_j, h_{j'}$ were chosen with the same distribution as h, and are independent when $j \neq j'$, we have the pointwise identities

$$\mathbf{E}(G_{h_j}) = \mathbf{E}(G_h) = F$$

and

$$\mathbf{E}(\overline{G_{h_{j'}}} G_{h_j}) = \mathbf{E}(\overline{G_{h'}})\mathbf{E}(G_h) = |F|^2 \text{ when } j \neq j'.$$

We thus can rewrite (50) as

$$\|F\|_{L^2}^2 - 2\|F\|_{L^2}^2 + \|F\|_{L^2}^2 + \mathbf{E}_{1\leq j,j'\leq D}\delta_{j,j'} \int_{\mathbf{Z}_N} (\overline{G_{h_{j'}}} G_{h_j} - |F|^2)$$

where $\delta_{j,j'}$ is the Kronecker delta. When $j = j'$, we have

$$\mathbf{E}(\overline{G_{h_{j'}}} G_{h_j}) - |F|^2 = \mathbf{E}(|G_h(x)|^2 | h \in H) - |\mathbf{E}(G_h(x)| h \in H)|^2$$

which is at most 1 since G_h is bounded (in fact one can sharpen this to $\frac{1}{4}$, but we will not need this). The claim follows. $\square$

Let $h_1, \ldots, h_{N_0^{100}}$ be as in the above Lemma. Then from (49) and (46) we see that

$$\|T^{\mu m} f_{UAP} - M\mathbf{E}(c_{\mu m, h_j} g_{h_j} | 1 \leq j \leq N_0^{100})\|_{L^2} \ll N_0^{-40} \text{ for all } 0 \leq m \leq (k-1)N_0. \tag{51}$$

We have now modeled a reasonably large number of shifts of our almost periodic function f_{UAP} in terms of a controlled number of functions $c_{\mu m, h_j}$. Next, we define a new factor $\mathcal{B}''$ finer than $\mathcal{B}'$ (and depending on μ, $h_1, \ldots, h_{N_0^{100}}$) by

$$\mathcal{B}'' := (\bigvee_{-(k-1)N_0 \leq m \leq (k-1)N_0} T^{\mu m}\mathcal{B}') \vee (\bigvee_{0\leq m\leq (k-1)N_0; 1\leq j\leq N_0^{100}} \mathcal{B}_{N_0^{-100}}(c_{\mu m, h_j}))$$

where $\mathcal{B}_\varepsilon(G)$ are the factors constructed by Proposition 6.2. Since the $c_{\mu m, h_j}$ are in UAP^{d-1} with norm at most 1, and $\mathcal{B}'$ was compact of order $d-1$ and complexity at most X', we see from (25) and Definition 6.4 that $\mathcal{B}''$ is also compact of order $d-1$ and complexity at most $O_{N_0,X'}(1)$. Also, from (20) we have

$$\|c_{\mu m,h_j} - \mathbf{E}(c_{\mu m,h_j}|\mathcal{B}'')\|_{L^\infty} \ll N_0^{-100} \text{ for all } 0 \leq m \leq (k-1)N_0$$

and hence (since the g_{h_j} are bounded)

$$\|M\mathbf{E}_{1\leq j\leq D}(c_{\mu m,h_j} g_{h_j}) - M\mathbf{E}(\mathbf{E}(c_{\mu m,h_j}|\mathcal{B}'')g_{h_j}|1 \leq j \leq D)\|_{L^2} \ll N_0^{-100}.$$

Combining this with (51) we see that

$$\|T^{\mu m} f_{UAP} - F_{\mu m}\|_{L^2} \ll_M N_0^{-40} \text{ for all } 0 \leq m \leq (k-1)N_0, \tag{52}$$

where F_n is defined for $n \in \mathbf{Z}_N$ by the formula

$$F_n := M\mathbf{E}_{1\leq j\leq D}\mathbf{E}(c_{n,h_j}|\mathcal{B}'')g_{h_j}. \tag{53}$$

We may then apply Proposition 9.1 (with $\mathcal{B}$ replaced by $\mathcal{B}''$) to estimate the quantity (48) as

$$(48) \gg \mathbf{E}_{1\leq\lambda\leq N_0/k_*}\mathbf{P}_{\mathbf{Z}_N} \bigcap_{m=1}^{k_*} E_{\mu\lambda} \tag{54}$$

where $k_* = O(1)$ and $E_{\mu\lambda m} = E_{\mu\lambda m}(k, \delta, k_*, \mathcal{B}'')$ was defined in that Proposition; recall that we are suppressing all dependence on the quantities d, k, δ, M.

Our attention thus turns to obtaining lower bounds for the size of $E_{\mu\lambda}$. We first use (52) to pass from $F_{\mu m}$ back to $T^{\mu m} f_{UAP}$ (modulo errors that can be made small by making D large). From (52) and Cauchy-Schwarz we have

$$\int_{\mathbf{Z}_N} \mathbf{E}(|T^{\mu m} f_{UAP} - F_{\mu m}||\mathcal{B}'') = \int_X |T^{\mu m} f_{UAP} - F_{\mu m}| \ll N_0^{-40} \text{ for all } 0 \leq m \leq (k-1)N_0,$$

so by Markov's inequality

$$\mathbf{P}_{\mathbf{Z}_N}(\mathbf{E}(|T^{\mu m} f_{UAP} - F_{\mu m}||\mathcal{B}'') \geq \frac{\delta}{16k}) \ll N_0^{-40} \text{ for all } 0 \leq m \leq (k-1)N_0.$$

In particular, from (38) and the triangle inequality we see (since $k_* = O(1)$) that

$$\mathbf{P}_{\mathbf{Z}_N} \bigcap_{m=1}^{k_*} E_{\mu\lambda m} \geq \mathbf{P}_{\mathbf{Z}_N} \bigcap_{m=1}^{k_*} E'_{\mu\lambda m} - O(N_0^{-30}), \tag{55}$$

where

$$\begin{aligned} E'_n := \{x \in \mathbf{Z}_N : \mathbf{E}(T^n f_{U^\perp}|\mathcal{B}'')(x) \geq \delta/2 \text{ and} \\ \mathbf{E}(|T^n f_{U^\perp} - T^n f_{UAP}||\mathcal{B}'')(x) \leq \frac{\delta}{16k}\}. \end{aligned}$$

The next step is to pull the shifts T^n out of the $\mathcal{B}''$ expectations. To do this we use the following observation.

Lemma 10.3 (Effective shift invariance of $\mathcal{B}''$). *Suppose that* $-(k-1)N_0 \leq m \leq (k-1)N_0$ *is such that*

$$\|\mathbf{E}(T^{\mu m} f_{U^\perp}|\mathcal{B}'') - T^{\mu m}\mathbf{E}(f_{U^\perp}|\mathcal{B}'')\|_{L^2} \geq N_0^{-100}$$

or

$$\|\mathbf{E}(T^{\mu m}|f_{U^\perp} - f_{UAP}|^2|\mathcal{B}'') - T^{\mu m}\mathbf{E}(|f_{U^\perp} - f_{UAP}|^2|\mathcal{B}'')\|_{L^2} \geq N_0^{-100}.$$

Then we are in the energy increment half of the dichotomy of Proposition 10.1.

Proof. We prove the first claim, as the second is analogous. Observe that $\mathbf{E}(T^{\mu m} f_{U^\perp}|\mathcal{B}'') = T^{\mu m}\mathbf{E}(f_{U^\perp}|T^{-\mu m}\mathcal{B}'')$, and so

$$\|\mathbf{E}(f_{U^\perp}|T^{-\mu m}\mathcal{B}'') - \mathbf{E}(f_{U^\perp}|\mathcal{B}'')\|_{L^2} \geq N_0^{-100}.$$

By the triangle inequality, we thus have either

$$\|\mathbf{E}(f_{U^\perp}|T^{-\mu m}\mathcal{B}'') - \mathbf{E}(f_{U^\perp}|\mathcal{B}')\|_{L^2} \geq \frac{1}{2}N_0^{-100}$$

or

$$\|\mathbf{E}(f_{U^\perp}|\mathcal{B}'') - \mathbf{E}(f_{U^\perp}|\mathcal{B}')\|_{L^2} \geq \frac{1}{2}N_0^{-100}.$$

But in either case we can use (31) (observing that $\mathcal{B}''$ and $T^{-\mu m}\mathcal{B}''$ are both finer than $\mathcal{B}'$, by construction of $\mathcal{B}''$) to obtain an energy increment (45), as desired. Similarly for the second claim (which uses the second component of $f = (f_{U^\perp}, |f_{U^\perp} - f_{UAP}|, |f_{U^\perp} - f_{UAP}|^2)$ rather than the first). $\square$

In light of this lemma, we may assume that

$$\|\mathbf{E}(T^n f_{U^\perp}|\mathcal{B}'') - T^n\mathbf{E}(f_{U^\perp}|\mathcal{B}'')\|_{L^2}, \quad \|\mathbf{E}(T^n|f_{U^\perp} - f_{UAP}|^2|\mathcal{B}'') - T^n\mathbf{E}(|f_{U^\perp} - f_{UAP}|^2|\mathcal{B}'')\|_{L^2} \leq N_0^{-100}$$

for all $n \in \mathbf{Z}_N$. In particular we have

$$\mathbf{P}_{\mathbf{Z}_N}(|\mathbf{E}(T^{\mu m} f_{U^\perp}|\mathcal{B}'') - T^{\mu m}\mathbf{E}(f_{U^\perp}|\mathcal{B}'')| \geq \delta/4) \ll N_0^{-100}$$

and

$$\mathbf{P}_{\mathbf{Z}_N}(|\mathbf{E}(|T^{\mu m} f_{U^\perp} - T^{\mu m} f_{UAP}||\mathcal{B}'') - T^{\mu m}\mathbf{E}(|f_{U^\perp} - f_{UAP}||\mathcal{B}'')| \geq \frac{\delta}{32k}) \ll N_0^{-50}$$

for all $0 \leq m \leq (k-1)N_0$. This allows us to estimate (since $k_* = O(1)$)

$$\mathbf{P}_{\mathbf{Z}_N}\bigcap_{m=1}^{k_*} E'_{\mu\lambda m} \geq \mathbf{P}_{\mathbf{Z}_N}\bigcap_{m=1}^{k_*} E''_{\mu\lambda m} - O(N_0^{-50}), \tag{56}$$

where

$$E''_n := \{x \in \mathbf{Z}_N : T^n\mathbf{E}(f_{U^\perp}|\mathcal{B}'')(x) \geq 3\delta/4 \text{ and } T^n\mathbf{E}(|f_{U^\perp} - f_{UAP}||\mathcal{B}'')(x) \leq \frac{\delta}{32k}\}.$$

Observe that $E''_n = T^n E''_0$. Combining this with (56), (55), (54) we obtain

$$(48) \gg \mathbf{E}_{1\leq\lambda\leq N_0/k_*} \int_X \prod_{m=1}^{k_*} T^{\mu\lambda m} \mathbf{1}_{E''_0} - O(N_0^{-30}). \tag{57}$$

The function $\mathbf{1}_{E''_0}$ is measurable in $\mathcal{B}''$, which is a compact factor of order $d-1$. At this point it it is tempting to apply the induction hypothesis (Theorem 3.3 for $d-1$) to $\mathbf{1}_{E''_0}$ (using Proposition 6.6) to obtain lower bounds for the right-hand side of (57). Unfortunately the problem is that the complexity of $\mathcal{B}''$ depends on X', whereas the range N_0/k_* of the variable λ is only allowed to depend on X, and so we cannot ensure that this expectation is even positive. To resolve this we must descend from the set $E''_0 \in \mathcal{B}''$ to the slightly modified set $E''_0 \cap \tilde{E}$, where

$$\tilde{E} := \{\mathbf{E}(f_{U^\perp}|\mathcal{B}) \geq \frac{7\delta}{8} \text{ and } \mathbf{E}(|f_{U^\perp} - f_{UAP}||\mathcal{B}) \leq \frac{\delta}{64k}\}.$$

Lemma 10.4. *Either we have*

$$\mathbf{P}_{\mathbf{Z}_N}(\tilde{E}\backslash E'') \ll \tau^2) \textit{ and } \mathbf{P}_{\mathbf{Z}_N}(E''_0 \cap \tilde{E}) \geq \delta/32$$

or we are in the energy increment half of the dichotomy.

Proof. We may assume without loss of generality that

$$\mathcal{E}_f(\mathcal{B}'') - \mathcal{E}_f(\mathcal{B}') \leq \tau^2$$

since otherwise we would be in the energy increment half of the dichotomy. From (32) we thus have

$$\mathcal{E}_f(\mathcal{B}'') - \mathcal{E}_f(\mathcal{B}) \leq 2\tau^2,$$

which implies from (31) and definition of f that

$$\int_{\mathbf{Z}_N} |\mathbf{E}(f_{U^\perp}|\mathcal{B}'') - \int_{\mathbf{Z}_N} f_{U^\perp}|\mathcal{B}')|^2 \leq 2\tau^2$$

and

$$\int_{\mathbf{Z}_N} |\mathbf{E}(|f_{U^\perp} - f_{UAP}||\mathcal{B}'') - \mathbf{E}(|f_{U^\perp} - f_{UAP}||\mathcal{B}')|^2 \leq 2\tau^2$$

In particular by Chebyshev's inequality we have

$$\mathbf{P}_{\mathbf{Z}_N}(|\mathbf{E}(f_{U^\perp}|\mathcal{B}'') - \mathbf{E}(f_{U^\perp}|\mathcal{B}')| \geq \frac{\delta}{8}) \ll \tau^2$$

and

$$\mathbf{P}_{\mathbf{Z}_N}(|\mathbf{E}(|f_{U^\perp} - f_{UAP}||\mathcal{B}'') - \mathbf{E}(|f_{U^\perp} - f_{UAP}||\mathcal{B}')| \geq \frac{\delta}{64k}) \ll \tau^2$$

and the first claim follows from the definitions of E''_0 and $\tilde{E}$.

Now we prove the second claim. From (5) we have

$$\int_{\mathbf{Z}_N} \mathbf{E}(f_{U^\perp}|\mathcal{B}) = \int_{\mathbf{Z}_N} f_{U^\perp} \geq \delta$$

and hence (by the boundedness of $\mathbf{E}(f_{U^\perp}|\mathcal{B}''))$

$$\mathbf{P}_{\mathbf{Z}_N}(\mathbf{E}(f_{U^\perp}|\mathcal{B}) \geq 7\delta/8 \geq \delta/8$$

while from (4) and Cauchy-Schwarz we have

$$\int_{\mathbf{Z}_N} \mathbf{E}(|f_{U^\perp} - f_{UAP}||\mathcal{B}'')| = \int_{\mathbf{Z}_N} |f_{U^\perp} - f_{UAP}| \leq \|f_{U^\perp} - f_{UAP}\|_{L^2} \leq \frac{\delta^2}{1024k}$$

and hence by Chebyshev's inequality

$$\mathbf{P}_{\mathbf{Z}_N}(\mathbf{E}(|f_{U^\perp} - f_{UAP}||\mathcal{B}'') > \frac{\delta}{64k}) \leq \delta/16.$$

By definition of E_0'', we thus have $\mathbf{P}_{\mathbf{Z}_N}(E_0'') \geq \delta/16$, and the second claim of the lemma thus follows from the first if τ is chosen sufficiently small. □

We may of course assume that we are not in the energy increment half of the dichotomy, in which case Lemma 10.4 implies that

$$\|\mathbf{1}_{E_0''\cap\tilde{E}} - \mathbf{1}_{\tilde{E}}\|_{L^2} \ll \tau.$$

Now observe that $\tilde{E} \in \mathcal{B}$ and $\mathcal{B}$ is compact of order $d-1$ and complexity at most X. Thus by Proposition 6.6 we can find a bounded nonnegative function $\tilde{f}_{UAP} \in UAP^{d-1}$ with

$$\|\mathbf{1}_{\tilde{E}} - \tilde{f}_{UAP}\|_{L^2} \ll \tau$$

and

$$\|\tilde{f}_{UAP}\|_{UAP^{d-1}} \ll_{X,\tau} 1.$$

In view of Lemma 10.4, we can thus apply the induction hypothesis of Theorem 10.4 with $f_{U^\perp}$ replaced by $\mathbf{1}_{E_0''\cap\tilde{E}}$, provided that τ is chosen sufficiently small. We conclude that

$$\mathbf{E}_{1\leq\lambda\leq N_0/k_*}\mathbf{E}_{\mathbf{Z}_N}\prod_{m=1}^{k_*} T^{\mu\lambda m}\mathbf{1}_{E_0} \gg_{\tau,X} 1$$

and so if we choose N_0 sufficiently large depending on τ, X, we see from (57) that

$$(48) \gg_{\tau,X} 1.$$

The claim (44) now follows from (47), and the proof of the Proposition is complete. □

The proof of Szemerédi's theorem is now complete.

References

[1] I. Assani, *Pointwise convergence of ergodic averages along cubes*, preprint.

[2] F. Beleznay, M. Foreman, *The complexity of the collection of measure-distal transformations.* Ergodic Theory Dynam. Systems **16** (1996), no. 5, 929–962.

[3] V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems,* J. Amer. Math. Soc. **9** (1996), 725–753.

[4] V. Bergelson, B. Host, R. McCutcheon, F. Parreau, *Aspects of uniformity in recurrence.* Colloq. Math. 84/85 (2000), part 2, 549–576.

[5] V. Bergelson and A. Leibman, *Set polynomials and a polynomial extension of the Hales-Jewett theorem*, Ann. of Math. (2) **150** (1999), no. 1, 33–75.

[6] J. Bourgain *A Szemerédi-type theorem for sets of positive density in* $\mathbf{R}^k$, Israel J. Math **54** (1986), no. 3, 307–316.

[7] J. Bourgain, *On triples in arithmetic progression*, GAFA **9** (1999), 968–984.

[8] P. Erdős, P. Turán, *On some sequences of integers*, J. London Math. Soc. **11** (1936), 261–264.

[9] P. Frankl, V. Rödl, *Extremal problems on set systems*, Random Structures and Algorithms **20** (2002), no. 2.

[10] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.

[11] H. Furstenberg, *Recurrence in Ergodic theory and Combinatorial Number Theory,* Princeton University Press, Princeton NJ 1981.

[12] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. d'Analyse Math. **34** (1978), 275–291.

[13] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for IP-systems and combinatorial theory*, J. d'Analyse Math. **45** (1985), 117–168.

[14] H. Furstenberg, Y. Katznelson, *A density version of the Hales-Jewett theorem*, J. d'Analyse Math. **57** (1991), 64–119.

[15] H. Furstenberg, Y. Katznelson, D. Ornstein, *The ergodic theoretical proof of Szemerédi's theorem*, Bull. Amer. Math. Soc. **7** (1982), 527–552.

[16] H. Furstenberg, B. Weiss, *A mean ergodic theorem for* $1/N\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$, Convergence in ergodic theory and probability (Columbus OH 1993), 193–227, Ohio State Univ. Math. Res. Inst. Publ., 5. de Gruyter, Berlin, 1996.

[17] J-Y. Girard, *Herbrand's theorem and proof theory*, Proceedings of the Herbrand symposium (Marseilles, 1981), 29–38, Stud. Logic Found. Math., 107, North-Holland, Amsterdam, 1982.

[18] T. Gowers, *Lower bounds of tower type for Szemerédi's uniformity lemma*, Geom. Func. Anal. **7** (1997), 322–337.

[19] T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, GAFA **8** (1998), 529–551.

[20] T. Gowers, *A new proof of Szemerédi's theorem*, GAFA **11** (2001), 465-588.

[21] T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem,* preprint.

[22] R. Graham, B. Rothschild, J.H. Spencer, Ramsey Theory, *John Wiley and Sons*, NY (1980).

[23] B.J. Green, T. Tao, *The primes contain arbitrarily long arithmetic progressions,* preprint.

[24] B.J. Green, T. Tao, *An inverse theorem for the Gowers $U^3(G)$ norm*, preprint.

[25] D.R. Heath-Brown, *Integer sets containing no arithmetic progressions*, J. London Math. Soc. **35** (1987), 385–394.

[26] B. Host, B. Kra, *Non-conventional ergodic averages and nilmanifolds,* to appear in Ann. Math.

[27] J. Kómlos, M. Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, Combinatorics, Paul Erdős is eighty, Vol 2. (Keszthely 1993), 295–352, Bolyai Soc. Math. Stud. **2**, János Bolyai Math. Soc. Budapest 1996.

[28] B. Nagle, V. Rödl, M. Schacht, *The counting lemma for regular k-uniform hypergraphs*, preprint.

[29] V. Rödl, J. Skokan, *Regularity lemma for k-uniform hypergraphs*, Random Structures and Algorithms, 25 (2004), no. 1, 1–42.

[30] V. Rödl, J. Skokan, *Applications of the regularity lemma for uniform hypergraphs*, Random Structures and Algorithms, 28 (2006), no. 2, 180–194.

[31] V. Rödl, M. Schacht, E. Tengan, N. Tokushige, *Density theorems and extremal hypergraph problems*, Israel J. Math., 152 (2006), 371–380.

[32] V. Rohlin, *Selected topics from the metric theory of dynamical systems*, Transl. Amer. Math. Soc. (2) **49** (1966), 171–209.

[33] K.F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 245-252.

[34] K.F. Roth, *Irregularities of sequences relative to arithmetic progressions*, IV. Period. Math. Hungar. **2** (1972), 301–326.

[35] I. Ruzsa, E. Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics (Proc. Fifth Hungarian Colloq. Keszthely, 1976), Vol. II., pp. 939–945., Colloq. Math. Soc. János Bolyai **18** (1978), North-Holland, Amsterdam-New York.

[36] S. Shelah, *Primitive recursive bounds for van der Waerden numbers*, J. Amer. Math. Soc. **1** (1988), 683–697.

[37] J. Solymosi, *Note on a question of Erdős and Graham*, Combinatorics Probability and Computing **13** (2004), 263–267.

[38] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104.

[39] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 299–345.

[40] E. Szemerédi, *Regular partitions of graphs,* in "Proc. Colloque Inter. CNRS" (J.-C. Bermond, J.-C. Fournier, M. Las Vergnas, D. Sotteau, eds.) (1978), 399–401.

[41] T. Tao, *Szemerédi's regularity lemma revisited*, preprint.

[42] T. Tao, *A quantitative ergodic theory proof of Szemerédi's theorem (abridged)*, unpublished.

[43] T. Tao, *The Gaussian primes contain arbitrarily shaped constellations*, preprint.

[44] B.L. Van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw. Arch. Wisk. **15** (1927), 212–216.

[45] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. **39** (1959), 358–360.

[46] M. Walters, *Combinatorial proofs of the polynomial Van der Waerden and the polynomial Hales-Jewett theorem*, J. London Math. Soc. **2** 61 (2000), 1–12.

[47] T. Ziegler, *Universal characteristic factors and Furstenberg averages*, J. Amer. Math. Soc. 20 (2007), 53-97.