Two proofs of Roth's theorem

Seungki Kim

Under the direction of Drs. Matthew Emerton and Bryna Kra

Northwestern University

April 25, 2010

Acknowledgements

I would like to express my highest gratitude to Professor Matthew Emerton, who has always been my guide in mathematics and enriched my mathematical experience greatly since my very first day at Northwestern. In fact, it is he who first recommended me to explore the subject matter of this paper. And I cannot thank enough Professor Bryna Kra, one of the world's best experts in this area of mathematics, who willingly agreed to have weekly discussion sessions and share her insights despite her crowded schedule, and encouraged me when I was about to give up writing this paper shortly before the deadline for submission. I also thank very much the professors who supported me during my educational crisis in summer 2009: (in the alphabetical order, titles omitted) Frank Calegari, John Coates, Kevin Costello, Matthew Emerton (I thank him again for his consistent concern and help in this matter), John Franks, Stefan Kaufmann, Min-Hyung Kim, Eugene Kushnirsky, Chiu-Chu Melissa Liu, Janet Pierrehumbert, Andrew Rivers, Mike Stein, Boris Tsygan, and Sara Vaux.

Contents

1 Introduction **4**

1.1 A note on my writing style 4

1.2 The topic 4

2 Gowers's proof **5**

2.1 Pseudorandom sets and the Gowers U^2-norm 6

2.2 Roth's theorem for pseudorandom functions 9

2.3 Roth's theorem for non-pseudorandom functions 10

3 Furstenberg's proof **14**

3.1 Some background knowledge 14

3.2 Multiple recurrence and correspondence principle 15

3.3 Outline of the proof 16

3.4 The "structured" case 18

3.5 The "random" case 19

4 Synthesis of the two arguments **22**

1 Introduction

1.1 A note on my writing style

This paper is my undergraduate thesis in mathematics. I think this paper should be demonstrative, rather than expository, of what I have learned and how I think of mathematics. Besides, there are tons of expository papers out there written by first-rate mathematicians, anyway. For this reason, I did not write anything that I did not really understand or was not convinced by; for instance, when I failed to understand a certain lemma, I proved something else that achieves the same effect. But more importantly, I did write what I thought upon seeing an interesting mathematical object or phenomenon.

I have observed repeatedly that mathematicians, despite their richness of thoughts and imaginations, are somehow very terse in writing. This is no less true of expository papers or textbooks, which I think are supposed to *expose* the ideas that are behind the definitions they propose and the theorems they prove. I know there are good reasons to be objective — to be careful so as not to mislead the readers, for example — but anyway, in my own paper, I decided to be explicit about it. So the readers will face two full pages of my attempt at motivating the definition of the Gowers U^2 norm (which is done in a single phrase in [17]), reports of my unsuccessful ideas, and declarations of lessons learned. These are really what I wanted to write about; a mere reproduction of well-known proofs is not only boring but useless as well. This paper might be applicable to both adjectives, too (I seriously hope not!), but that is precisely why I attempt to find its meaning in being demonstrative.

1.2 The topic

Our ultimate interest lies in the following surprising fact, once a conjecture of Erdös and Turán [2] in 1936, first proved by Szemerédi [16] in 1975:

Theorem 1 (Szemerédi [16])**.** *Let $A \subset \mathbb{Z}$ be a set of positive upper density.*[1] *Then A contains an arithmetic progression of any finite length.*

Szemerédi's theorem has inspired many brilliant works in mathematics, including Szemerédi's own proof, several different proofs ([3], [5]) all with immense contribution to mathematics, and the famous Green-Tao theorem ([9]), which asserts that the the set of the prime numbers contains an arithmetic progression of any finite length. But in this paper, we focus on something less ambitious (nevertheless not easy at all):

[1] The positive upper density of a set is defined as $d^*(A) = \limsup_{N\to\infty} |A \cap [-N, N]|/|[-N, N]$.

Theorem 2 (Roth [15])**.** *A subset of $\mathbb{Z}$ with positive upper density contains an arithmetic progression of length three.*

This special case of the Szemerédi's theorem is named after Klaus Roth, who proved it in 1953. The general case is substantially more difficult and longer (all the known proofs consume fifty pages or more), and I will not explain it here. However, understanding Roth's theorem is a good first step to understanding Szemerédi's theorem. The outlines of their proofs are basically identical; the only problem is that a tool used in a certain part of the proof of Roth's theorem needs to be sharpened in order to be useful in the general case. The natural questions at this point are:

1. What is exactly that "tool," and what is exactly that "certain part?"
2. How does the "sharpening" go about?

If one knows the answers to these two questions, then one understands Szemerédi's theorem. In the present exposition we are mainly concerned with the first question. For this end, we examine two different proofs of Roth's theorem: Gowers's proof [5] that utilizes the *Gowers U^2-norm*, and Furstenberg's work [3], [4] using ergodic theory. They give slightly different answers to the above questions, but there are also many parallels and similarities in important respects. I will start by explaining each approach, and give a third proof that combine their ideas and highlight the interesting points of comparison and contrast.[2]

2 Gowers's proof

Our goal is to show

Proposition 1. *For any positive integer N and a nonempty subset $A \subset \mathbb{Z}/N\mathbb{Z}$, let $\delta := |A|/N$. Then*

$$\mathbb{E}_{x,d\in\mathbb{Z}/N\mathbb{Z}} 1_A(x)1_A(x+d)1_A(x+2d) \geq c(\delta).$$

for some $c(\delta) > 0$ depending only on δ.

[2] I worked out the third proof myself, but I am pretty sure that I am not the only person who had the same idea. And it is not really my original idea, although I have never seen it done in exactly the same way. The combinations of Gowers's and Furstenberg's insights are seen all over in this part of mathematical literature, such as [7], [9], [11], [17].

The summation on the left is intended to count the number of 3-term arithmetic progressions in A. But it also counts the trivial progressions of the form (x, x, x), and the pathological ones like $(x, x+d, x)$ with $2d = N$; fortunately, their contribution to the sum is no greater than $2\delta N^{-1}$, which vanishes as $N \to \infty$. A more serious problem is that, in this cyclic group environment, a 3-progression can "wrap around"; i.e. things like $(N-1, N, 1)$ and $(N-7, N-2, 3)$ count as 3-progressions in $\mathbb{Z}/N\mathbb{Z}$.

But it is not so difficult to lift this result to $\mathbb{Z}_{>0}$. If $A \subset \mathbb{Z}_{>0}$ with positive upper density δ, fix $0 < \varepsilon << \delta$, and pick a sufficiently large N such that $|A \cap [1, N]|/N \geq \delta - \varepsilon$. Then at least for one $i \in 0, 1, 2$, $A \cap [iN/3 + 1, (i+1)N/3]$ has density $\delta' \geq (\delta - \varepsilon)/3$ in $[1, N]$. Applying the above proposition with $A \cap [iN/3 + 1, (i+1)N/3] \subset \mathbb{Z}/N\mathbb{Z}$ and δ', one only gets to count the progressions that do not wrap around. Therefore it really suffices to prove this proposition, which we now do.

2.1 Pseudorandom sets and the Gowers U^2-norm

In his proof of Szemerédi's theorem, Gowers [5] starts by remarking that "random sets" have many arithmetic progressions. Although he does not mention it explicitly, a random set here means a set A of density $\delta > 0$ with the following property: for any positive integer d, if $x \in A$, then $x + d \in A$ with probability δ. For example, if $A \subset \mathbb{Z}/N\mathbb{Z}$ and $\delta = |A|/N$, then the number of 3-term arithmetic progressions in A is approximately $\delta N \cdot \delta N \cdot \delta = \delta^3 N^2$.

At this point, one may take a flight of imagination, and consider randomness as a property that we can investigate and manipulate. Then it is natural to come to the following thoughts:

1. If a set is *sufficiently random*, then it must have around $\delta^3 N^2$ 3-term progressions.
2. If a set is not sufficiently random, it must be *structured*, whatever that means (it will be discussed later in the paper).

This is the core idea behind this proof of Szemerédi's theorem. In fact, this way of thinking in terms of random sets and non-random (or *structured*) sets is present in any proof of Szemerédi's theorem, and also in the Green-Tao theorem. Terence Tao said during his lecture in ICM 2006 [18]:

> Firstly, for a given class of objects, one quantifies what it means for an object to be "(pseudo-)random" and an object to be "structured". Then, one establishes a *dichotomy between randomness and structure*, which typically looks something like this:

If an object is not (pseudo-)random, then it (or some non-trivial component of it) correlates with a structured object.

(Also, see [6])

Tao, who has an extraordinary naming sense,[3] calls this "the dichotomy of randomness and structure."

In this section, we make rigorous the first thought, and the second thought will be explained in the next part. Instead of the term "sufficiently random," the word *pseudorandom* is often used in the literature of Szemerédi's theorem and related works. Now our task is to devise an appropriate mathematical expression of pseudorandomness, or, in other words, a way to measure the degree of randomness of a set.

Recall what is meant by "random" here: for a set A of density $\delta > 0$, we say A is random if for any nonzero integer d, if $x \in A$, then $x + d \in A$ with probability δ. It is not so clear if any subset of $\mathbb{Z}_{>0}$ or $\mathbb{Z}/N\mathbb{Z}$ has this property, but that is not what is claimed here anyway. This notion of randomness is there just as some kind of yardstick.

For a set $A \subset \mathbb{Z}/N\mathbb{Z}$ and an integer $0 \leq d \leq N$, the probability that $x \in A \Rightarrow x + d \in A$ is $\mathbb{E}_{x \in A} 1_A(x) 1_A(x + d)$, but we can also measure it (although on a different scale) by the more convenient $\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} 1_A(x) 1_A(x + d)$. But then summing this again over all $d \in \mathbb{Z}/N\mathbb{Z}$, we only get the constant $\delta N \cdot \delta N \cdot N^{-2}$. In order to salvage the lost information, we sum up the square of each individual term:[4]

$$\mathbb{E}_{d \in \mathbb{Z}/N\mathbb{Z}} |\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} 1_A(x) 1_A(x + d)|^2.$$

This is the fourth power of the *Gowers U^2-norm* of 1_A, written $\|1_A\|_{U^2}$. It is indeed a norm over the space of functions on $\mathbb{Z}/N\mathbb{Z}$.

Note that when A is random, the term inside $\mathbb{E}_x 1_A(x) 1_A(x + d)$ equals δ^2 (except when $d = 0$). Ignoring the pesky case $d = 0$ (in which case the sum vanishes as $N \to \infty$ anyway), we may heuristically call A *pseudorandom* if

[3] We will soon see another instance of his naming sense.

[4] This is not an unnatural thing to do. In statistics, when one tries to measure the average distance between each variable x and the mean M by $\mathbb{E}_x(x - M)$, one obtains nothing more than zero. Hence we instead measure $\mathbb{E}_x|x - M|^2$, i.e. the variance.

$$\mathbb{E}_{d\in\mathbb{Z}/N\mathbb{Z}}|\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}1_A(x)1_A(x+d)-\delta^2|^2$$

is close to zero. Indeed, fixing the density δ, a set A of density δ is random if and only if the above value is the smallest possible.

We may further refine this expression into

$$\mathbb{E}_{d\in\mathbb{Z}/N\mathbb{Z}}|\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}f(x)f(x+d)|^2$$

where $f(x)=1_A(x)-\delta$ is the *balanced function* of A. Evidently, this is the fourth power of the Gowers norm of f, or, we could also say, the Gowers norm of the set A.

A little discussion about the motivation of the Gowers norm: there are many ways to understand how it comes about — see [5], [7], and [17]. The one above is what I personally am most comfortable with. Gowers himself introduced it [5] in [5] while trying to use Fourier analysis to prove Roth's theorem. In the language of Fourier analysis, where $\omega=e^{2\pi i/N}$,

$$\begin{aligned}\mathbb{E}_{x,d\in\mathbb{Z}/N\mathbb{Z}}1_A(x)1_A(x+d)1_A(x+2d) &= \mathbb{E}_{a,b,c\in A}\mathbb{E}_{r\in\mathbb{Z}/N\mathbb{Z}}\omega^{r(a-2b+c)}\\ &= \sum_{r\in\mathbb{Z}/N\mathbb{Z}}\hat{1}_A(r)\hat{1}_A(-2r)\hat{1}_A(r)\end{aligned}$$

and

$$\mathbb{E}_{d\in\mathbb{Z}/N\mathbb{Z}}|\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}f(x)\overline{f(x+d)}|^2=\mathbb{E}_{r\in\mathbb{Z}/N\mathbb{Z}}|\hat{f}(r)|^4,$$

(note that f here is the balanced function of the set A) which, when sufficiently small, somehow[6] gives a nonzero upper bound to the earlier sum that counts the 3-progressions, thereby giving sense to our motto pseudorandom sets contain 3-progressions. Also note the complex conjugation sign on the left-hand side; Gowers apparently cared about how his norm would apply to complex-valued functions, which I didn't, as it was not immediately necessary, and will not be until the end of this proof. But for the sake of delivering the correct information, I will declare the definition as he did anyway, in the complex version:

[5] In fact, Wiener [22] is the first person to invent it under the name of the *autocorrelation function*, although his purpose then had nothing to do with Szemerédi's theorem, then a conjecture of Erdös and Turán [2].

[6] The details are very simple, but kind of messy — see Gowers's original proof [5].

Definition (Gowers U^2-norm)**.** *The Gowers U^2-norm of a function $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ is*

$$\|f\|_{U^2} = [\mathbb{E}_{d\in\mathbb{Z}/N\mathbb{Z}}|\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}f(x)\overline{f(x+d)}|^2]^{1/4} = [\mathbb{E}_{x,n,m\in\mathbb{Z}/N\mathbb{Z}}f(x)\overline{f(x+n)f(x+m)}f(x+m+n)]^{1/4}. \tag{1}$$

Indeed,

Proposition 2. *The Gowers U^2-norm is a norm on the space of functions $\mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$.*

Checking this amounts to a few straightforward computations, and so is omitted (for a proof, see [5] or [9]).

2.2 Roth's theorem for pseudorandom functions

Much of the power of the Gowers norm comes from the following statement.

Lemma 1 (Generalized von Neumann Theorem[7] [8])**.** *If $f_1, f_2, f_3 : \mathbb{Z}/N\mathbb{Z} \longrightarrow D^2 \in \mathbb{C}$, we have*

$$\mathbb{E}_{x,d}f_1(x)f_2(x+d)f_3(x+2d) \leq \|f_i\|_{U^2}$$

for any $i \in \{1, 2, 3\}$.

Proof. Here we provide the proof for the case $i = 1$; the other cases are very similar. Rewrite

$$\mathbb{E}_{x,d}f_1(x)f_2(x+d)f_3(x+2d) = \mathbb{E}_{s,t}f_1(2s-t)f_2(s)f_3(t).$$

By the Cauchy-Schwarz inequality and the fact that $|f_2(x)| \leq 1$,

$$|\mathbb{E}_{x,d}f_1(x)f_2(x+d)f_3(x+2d)|^2 \leq \mathbb{E}_s\mathbb{E}_{t,t'}f_1(2s-t)\overline{f_1(2s-t')}f_3(t)\overline{f_3(t')}.$$

Again by the Cauchy-Schwarz inequality and the fact that $f_3(t)\overline{f_3(t')} \leq 1$,

[7] This is how Tao named it. The apparent lack of similarity between this lemma and anything that is called von Neumann(-Koopman) theorem in ergodic theory nonplussed everyone that works on Szemerédi-related problems.

$$|\mathbb{E}_{x,d} f_1(x) f_2(x+d) f_3(x+2d)|^4 \leq \mathbb{E}_{s,s'} \mathbb{E}_{t,t'} f_1(2s-t) \overline{f_1(2s-t') f_1(2s'-t)} f_1(2s'-t').$$

But then, the right-hand side here coincides with $\|f_1\|_{U^2}^4$, as desired. □

Corollary. *If $f : \mathbb{Z}/N\mathbb{Z} \longrightarrow D^2$, then $\mathbb{E}_{x,d \in \mathbb{Z}/N\mathbb{Z}} f(x) f(x+d) f(x+2d) \leq \|f\|_{U^2}$.*

Below is the main result of this tiny section. It is basically Proposition 1 with the restriction that A is pseudorandom.

Proposition 3 (Sets with sufficiently small Gowers norms have 3-progressions)**.** *Let $A \subset \mathbb{Z}/N\mathbb{Z}$ with $|A| = \delta$, and f be the balanced function of A. If*

$$\|f\|_{U^2} \leq \delta^3/8,$$

then

$$\mathbb{E}_{x,d \in \mathbb{Z}/N\mathbb{Z}} 1_A(x) 1_A(x+d) 1_A(x+2d) \geq \delta^3/8.$$

Proof. Write $f_1(x) = \delta$ and $f_2(x) = f(x)$. Then

$$\begin{aligned}
&\mathbb{E}_{x,d} 1_A(x) 1_A(x+d) 1_A(x+2d) \\
&= \mathbb{E}_{x,d} (f_1(x) + f_2(x))(f_1(x+d) + f_2(x+d))(f_1(x+2d) + f_2(x+2d)) \\
&= \mathbb{E}_{x,d} f_1(x) f_1(x+d) f_1(x+2d) + \sum_{i,j,k \in 1,2, \text{not all } 1} \mathbb{E}_{x,d} f_i(x) f_j(x+d) f_k(x+2d) \\
&\geq \delta^3 - 7\|f_2\|_{U^2} \\
&\geq \delta^3/8,
\end{aligned}$$

by the Generalized von Neumann theorem. □

Remark. *In the above argument, the smaller we assume $\|f\|_{U^2}$ is, the higher the lower bound becomes, closer to δ^3, which is the expected value for random sets. In addition, the supremum of $\|f\|_{U^2}$ we may assume is $\delta^3/7$ not inclusive. So there is nothing special about the denominator 8 in our estimate above.*

2.3 Roth's theorem for non-pseudorandom functions

To complete the proof of Roth's theorem, we need to know what we can say about sets with Gowers norms $> \delta^3/8$. Indeed, we have the following result.

Lemma 2 (Gowers inverse theorem [7]). *Let $\alpha \in [0,1]$. Then $\|f\|_{U^2} \geq \alpha \Longrightarrow \|\hat{f}\|_\infty \geq \alpha^2$.*

Proof. Note that

$$\|f\|_{U^2}^4 = \mathbb{E}_r |\hat{f}(r)|^4,$$

which in fact gives a better estimate $\|\hat{f}\|_\infty \geq \alpha$.

□

To see why this information is helpful, consider the explicit expression of the Fourier transform $\hat{f}(r)$:

$$\hat{f}(r) = \mathbb{E}_x f(x)\omega^{-rx}$$

This is a weighted sum of certain unit vectors. Visualizing how the sum accumulates on the complex plane as x runs through $\mathbb{Z}/N\mathbb{Z}$ (to facilitate thinking here assume $r|N$), one can actually prove a statement of kind 'if $|\hat{1}_A(r)| > M$ for some $M = M(\delta)$, then A contains a 3-progression,' although such M is far much larger than $\delta^3/8$ and thus not so useful in our situation. On the other hand, if A does have a 3-progression, then for some r a partial sum of $\hat{f}(r)$ *must* accumulate large enough so as to imply the existence of a 3-progression, although the whole sum does not have to. We could detect this partial accumulation by, for example, convolving $\hat{f}(r)$ with something, and we could prove Roth's theorem this way. In fact, Bryna Kra [13] told me that B. Host and she took on this approach and proved a stronger claim than Szemerédi's theorem (namely, that A contains a positive density of k-progressions with the same common difference) for cases $k = 3$ and 4; however, the claim is false for $k \geq 5$, for the reason provided by I. Ruzsa: it somehow has to do with the terrible non-abelianness of the alternating group A_k for $k \geq 5$. It must be very interesting and deep, but since one of our goals for studying Roth's theorem is to eventually understand Szemerédi's theorem, in this paper I am not stepping onto this track.

I mentioned these abortive Fourier-analytic approaches to make a point that Fourier coefficients do contain some information about 3-progressions. Now we look at the approach that actually works:

Proposition 4 (Density increment [8]). *For $0 < \delta < 1$ and a sufficiently large N depending on δ, if $A \in \mathbb{Z}/N\mathbb{Z}$ with density δ, and if the Gowers norm of A is greater than $\delta^3/8$, then there exists an arithmetic progression $P \subset [1,N] = 1,\ldots,N$ whose length diverges to infinity as $N \to \infty$ and $\varepsilon = \varepsilon(\delta)$, such that $|A \cap P|/|P| \geq (\delta + \varepsilon)$.*

(Here we shift to the context of $[1, N]$ to avoid the wraparound issues.)

If in addition $\varepsilon(\delta)$ may be defined as a nondecreasing function of δ, then finitely many applications of this statement imply Szemerédi's theorem.

Let f be the balanced function of A. By the Gowers inverse theorem, we already know that there exists $r \in \mathbb{Z}/N\mathbb{Z}$ such that $|\hat{f}(r)| = \mathbb{E}_x f(x)\omega^{-rx} \geq \delta^6/64$. A couple of lemmas are in order.

Lemma 3 (Dirichlet's pigeonhole principle/Weyl equidistribution theorem [8])**.** *Take r as in the above discussion, and suppose $0 < c < 1$. Then there exists a positive integer $d \leq 1/c$ such that $\|dr\|_{\mathbb{R}/\mathbb{Z}} \leq c$.*

Proof. Consider $0, r, 2r, \ldots, mr$ where $m = \lfloor 1/c \rfloor$. By the pigeonhole principle, there exist j, j' such that $\|jr - j'r\|_{\mathbb{R}/\mathbb{Z}} \leq c$. Take $d = |j - j'|$. □

Lemma 4. *Take the earlier r, and let $0 < \eta < 1$. Suppose that $N > C\eta^{-6}$ for some appropriate constant C. Then there exists a partition of $[1, N]$ into arithmetic progressions P_i, $i = 1, \ldots, n$ each of length at least $N^{1/3}$, such that $\sup_{x,x' \in P_i} |\omega^{rx} - \omega^{rx'}| \leq \eta$ for each i.*

Proof. Applying the previous lemma with $c = \eta N^{-1/3}/4\pi$, we can find a $d \leq 4\pi N^{1/3}/\eta$ such that $\|dr\|_{\mathbb{R}/\mathbb{Z}} \leq \eta N^{-1/3}/4\pi$. If P is any progression with common difference d and length at most $2N^{1/3}$, then, by $2N^{1/3}$ applications of the triangle inequality

$$\sup_{x,x' \in P} |\omega^{rx} - \omega^{rx'}| \leq 2N^{1/3} |\omega^{dr} - 1|.$$

Since $|\omega^t - 1| = 2|\sin \pi t| \leq 2\pi \|t\|_{\mathbb{R}/\mathbb{Z}}$,

$$\sup_{x,x' \in P} |\omega^{rx} - \omega^{rx'}| \leq 4\pi N^{1/3} \|dr\|_{\mathbb{R}/\mathbb{Z}} \leq \eta.$$

If $N > C\eta^{-6}$, then d is at most $\sqrt{N}$ (note: C can be whatever that makes this estimate correct; a simple computation shows that we can fix $C = (4\pi)^6$), and now it is tedious but not so hard to see that $[1, N]$ may be partitioned into progressions P_i of difference d and length between $N^{1/3}$ and $N^{2/3}$. □

Proof of density increment. Let $\eta = \delta^6/128$, and make N sufficiently large to satisfy $N > C\eta^{-6}$. Then apply the above lemma to obtain the progressions P_i.

Recall we had $|\sum_x f(x)\omega^{-rx}| > \delta^6 N/64$. By the triangle inequality,

$$\sum_{i=1}^{n}\Big|\sum_{x\in P_i} f(x)\omega^{-rx}\Big| > \delta^6 N/64.$$

Therefore, fixing some $x_i \in P_i$ for each $i = 1, \ldots, k$, we obtain

$$\begin{aligned}
\sum_{i=1}^{n}\Big|\sum_{x\in P_i} f(x)\Big| &= \sum_{i=1}^{n}\Big|\sum_{x\in P_i} f(x)\omega^{-rx_i}\Big| \\
&\geq \sum_{i=1}^{n}\Big|\sum_{x\in P_i} f(x)\omega^{-rx}\Big| - \sum_{i=1}^{n}\Big|\sum_{x\in P_i} f(x)(\omega^{-rx_i} - \omega^{-rx})\Big| \\
&\geq \delta^6 N/64 - \sum_{i=1}^{n}\Big|\sum_{x\in P_i} f(x)\Big|\eta \\
&\geq \delta^6/128 \cdot \sum_{i=1}^{n}|P_i|.
\end{aligned}$$

Note that we used $N = \sum_{i=1}^{n}|P_i|$ above. To remove the modulus sign on the left side, we use the fact that $\sum_{i=1}^{n}\sum_{x\in P_i} f(x) = 0$, which implies

$$\sum_{i=1}^{n}(\Big|\sum_{x\in P_i} f(x)\Big| + \sum_{x\in P_i} f(x)) \geq \delta^6/128 \cdot \sum_{i=1}^{n}|P_i|.$$

By the pigeonhole principle, there exists i so that

$$\begin{aligned}
&\Big|\sum_{x\in P_i} f(x)\Big| + \sum_{x\in P_i} f(x) \geq \delta^6|P_i|/128 \\
&\Rightarrow \sum_{x\in P_i} f(x) \geq \delta^6|P_i|/256
\end{aligned}$$

$$\Rightarrow \sum_{x\in P_i}(1_A(x)-\delta) \geq \delta^6|P_i|/256$$
$$\Rightarrow |A\cap P_i| \geq (\delta+\delta^6/256)|P_i|,$$

as desired. This completes the proof of Density increment and of Roth's theorem. □

3 Furstenberg's proof

3.1 Some background knowledge

I will start by providing some basic facts in ergodic theory. For details, see [20].

Definition. *A* probability measure space *is a triple $(X,\mathcal{B},\mu)$ where X is a set, $\mathcal{B}$ is a σ-algebra over X, and μ is a probability measure on $(X,\mathcal{B})$, i.e. $\mu(X)=1$. A* probability measure preserving system *$(X,\mathcal{B},\mu,T)$ is a probability measure space equipped with a* measure preserving transformation *T; that is, $T:X\to X$ is measurable and $\mu(A)=\mu(T^{-1}A)$ for all $A\in\mathcal{B}$.*

*A set $A\in\mathcal{B}$ is called T-*invariant *if $T^{-1}A=A$ almost everywhere. T, or $(X,\mathcal{B},\mu,T)$ is called* ergodic *if the only T-invariant sets are $\emptyset$ and X up to null sets.*

The following result, the von Neumann ergodic theorem, suggests one among many cool properties of an ergodic transformation.

Theorem 3 (von Neumann [19])**.** *Let U be a unitary operator on the Hilbert space $\mathcal{H}$, $\mathcal{M}=\{x\mid Ux=x\}$, P the orthogonal projection onto $\mathcal{M}$. Let $S_N=\frac{1}{N}\sum_{i=0}^{N-1}U^i$. Then for every $x\in\mathcal{H}$, $S_Nx\to Px$ in the norm topology.*

One may check that if $(X,\mathcal{B},\mu,T)$ is a probability measure preserving space, then the operator $U_T:L^2(X,\mathcal{B},\mu)\to L^2(X,\mathcal{B},\mu)$ by $U_Tf=f\circ T$ is unitary.[8] So we can apply the above theorem with $\mathcal{H}=L^2$ and $U=U_T$. If T is ergodic, then $\mathcal{M}$ is the subspace spanned by constant functions, which implies

$$\frac{1}{N}\sum_{i=0}^{N-1}U_T^if\to\int f d\mu$$

[8]Often we abuse the notation and write T for U_T.

in L^2, which is very useful in many situations — we will see one soon.

On a side note, it is a theorem of Birkhoff [1] that if T is ergodic, then $\frac{1}{N}\sum_{i=0}^{N-1} f(T^i x)$ converges to $\int f d\mu$ for almost every x. In this paper, we only need the von Neumann ergodic theorem.

3.2 Multiple recurrence and correspondence principle

Our discussion of an ergodic proof of Roth's theorem starts with a basic result in ergodic theory, proved by Poincaré.

Theorem 4 (Poincaré recurrence [14])**.** *If $(X, \mathcal{B}, \mu, T)$ is a probability measure preserving system and $A \in \mathcal{B}$ with $\mu(A) > 0$, then there exist $n \in \mathbb{N}$ such that $\mu(A \cap T^{-n}A) > 0$.*

In other words, there are many elements of A that comes back to A after n iterations of T (hence the name "recurrence").

Example 1. *Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ with 0 and 1 identified be the unit circle, and let $\mathcal{L}$ and λ be the Lebesgue σ-algebra and the Lebesgue measure on $\mathbb{T}$, respectively. Define a transformation T on $\mathbb{T}$ by $Tx = x + \alpha$ (mod 1), where $\alpha \in [0, 1]$. We now have a system $(\mathbb{T}, \mathcal{L}, \lambda, T)$.*

Pick $x \in \mathbb{T}$ and $\varepsilon > 0$. Define $A = [x - \varepsilon, x + \varepsilon]$. By Poincaré recurrence, there exists n such that $\mu(A \cap T^{-n}A) > 0$, i.e. $T^n x$ comes back arbitrarily close to x. This is not unrelated to the Dirichlet lemma we used in the earlier proof of Roth's theorem: $\mathbb{T}$ can also be parametrized as the unit circle $\{e^{i\theta} | \theta \in [0, 2\pi]\}$ on $\mathbb{C}$, and this result is independent of parametrizations.

Example 2. *The above example can be generalized as follows. Take the system $(\mathbb{T} \times \ldots \times \mathbb{T}, \mathcal{L} \times \ldots \times \mathcal{L}, \lambda \times \ldots \times \lambda, T)$, where the direct sums are of $m < \infty$ terms and $T(x_1, \ldots, x_m) = T(x_1 + \alpha_1, \ldots, x_m + \alpha_m)$, $\alpha_i \in [0, 1]$.*

Fix $(x_1, \ldots, x_m) \in \mathbb{T} \times \ldots \times \mathbb{T}$ and $\varepsilon > 0$, and define $A = [x_1 - \varepsilon, x_1 + \varepsilon] \times \ldots \times [x_m - \varepsilon, x_m + \varepsilon]$. Again by Poincaré recurrence, there exists n such that $\mu(A \cap T^{-n}A) > 0$, i.e. $T^n(x_1, \ldots, x_m)$ comes back arbitrarily close to $(x_1, \ldots, x_m)$, in the sense that $\|(x_i + n\alpha_i) - x_i\|_{\mathbb{R}/\mathbb{Z}} < 2\varepsilon$ for every i.

A similar result can be proved using Fourier analysis; that is the Weyl equidistribution theorem *[21].*

Looking at Poincaré recurrence, one may wonder what is going to happen to $\mu(A \cap T^{-n}A \cap T^{-2n}A)$, $\mu(A \cap T^{-n}A \cap T^{-2n}A \cap T^{-3n})$, etc. In fact, the same thing is going to happen:

Theorem 5 (Furstenberg; multiple recurrence [3]). *If $(X, \mathcal{B}, \mu, T)$ is a probability measure preserving system and $A \in \mathcal{B}$ with $\mu(A) > 0$, then for any $k \geq 1$ there exists $n \in \mathbb{N}$ such that $\mu(A \cap T^{-n}A \cap \ldots \cap T^{-kn}A) > 0$.*

This statement looks very analogous to Szemerédi's theorem. In fact, it is implies to Szemerédi's theorem,[9] via the following theorem:

Theorem 6 (Furstenberg; correspondence principle [4]). *Suppose $E \subset \mathbb{Z}$ has positive upper density $d^*(E) > 0$. Then there exists an ergodic system[10] $(X, \mathcal{B}, \mu, T)$ and a set $A \in \mathcal{B}$ with $\mu(A) = d^*(E)$ such that*

$$\mu(T^{-m_1}A \cap \ldots \cap T^{-m_k}A) \leq d^*((E + m_1) \cap \ldots \cap (E + m_k))$$

for all $k \in \mathbb{N}$ and all $m_1, \ldots, m_k \in \mathbb{Z}$.

For a proof, see [12].

We only care about the case $k = 2$ in this paper. Also, since the correspondence principle gives an ergodic system (in fact it is a Bernoulli shift), to prove Szemerédi it suffices to worry about Theorem 5 for ergodic systems only. Therefore, for our goals here, it suffices to prove the following statement.

Theorem 7 (Furstenberg [4]). *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic system. If $f \in L^\infty(X, \mathcal{B}, \mu)$ is nonnegative and not a.e. 0, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int f(x) f(T^n x) f(T^{2n} x) d\mu(x) > 0.$$

If $A \subset X$ with $\mu(A) > 0$, letting $f(x) = 1_A(x)$ and applying the above theorem yields the desired result.

3.3 Outline of the proof

As in Gowers's proof of Roth's theorem, Furstenberg's proof is in two parts. However, the ideas are different. Gowers uses the fact that random sets have arithmetic progressions, and

[9] Actually a little work shows that it is equivalent to Szemerédi's theorem; see [12].

[10] In the original proof of Furstenberg [3], the system was not ergodic.

then shows that nonrandom sets have linear bias i.e. it is distributed more densely on a certain portion of $\mathbb{N}$ than on the rest. Furstenberg first proves that "structured" sets have 3-term arithmetic progressions, and in the case of general sets, he finds arithmetic progressions in their "structured component." Let me explain this more carefully. First, the notion of "structure" used here is as follows:

Definition (Kronecker factor [4], [12]). *Let $(X, \mathcal{B}, \mu, T)$ be a probability measure preserving system. The* Kronecker factor $\mathcal{E}(X, \mathcal{B}, \mu, T)$ *of $(X, \mathcal{B}, \mu, T)$ is the subspace of $L^2(X, \mathcal{B}, \mu)$ spanned by the eigenvectors of T.*

The Kronecker factor is also called the *rotation factor*. The name probably comes from the fact that T acts like a rotation on each eigenvector; $Tf = \lambda f \Rightarrow |\lambda| = 1$, so T works like rotating the image of f on the complex plane. In effect, T acts on $\mathcal{E}(X, \mathcal{B}, \mu, T)$ like a rotation on a direct sum of $\mathbb{T}$.

A rotation is a fairly well-understood object in ergodic theory; Example 2 already suggests that for any $f \in \mathcal{E}(X, \mathcal{B}, \mu, T)$, there exists some n such that $T^n f$ and $T^{2n} f$ comes arbitrarily close to f. In fact, the first part of Furstenberg's proof is a rigorous expression of this intuition.

Proposition 5 (Furstenberg [3]). *If $\psi \in L^\infty(X) \cap \mathcal{E}(X, \mathcal{B}, \mu, T)$, then for every $\varepsilon > 0$ there exists a syndetic set of n*[11] *such that*

$$\int \psi(x)\psi(T^n x)\psi(T^{2n}x)d\mu > \int \psi(x)^3 d\mu - \varepsilon.$$

For any $f \in L^2(X, \mathcal{B}, \mu)$, denote by $\tilde{f}$ the orthogonal projection of f onto $\mathcal{E}(X, \mathcal{B}, \mu, T)$. (This makes sense because $\mathcal{E}(X, \mathcal{B}, \mu, T)$ is a closed subspace of L^2.) $\tilde{f}$ may be interpreted as the "structured component" of f. The second part of Furstenberg's proof consists of the following deep theorem:

Proposition 6 (Furstenberg [3]). *If $f, g, h \in L^\infty(X, \mathcal{B}, \mu) \cap L^2(X, \mathcal{B}, \mu)$, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int f(x)g(T^n x)h(T^{2n}x)d\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \tilde{f}(x)\tilde{g}(T^n x)\tilde{h}(T^{2n}x)d\mu.$$

Proof of Theorem 7. By Proposition 6, it suffices to show that if $f \in L^\infty$ is nonnegative and not a. e. 0, then

[11] A subset of $\mathbb{Z}$ is syndetic if the gap between any of its two adjacent element is bounded by a constant.

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int \tilde{f}(x)\tilde{f}(T^n x)\tilde{f}(T^{2n}x)d\mu(x) > 0.$$

By Proposition 5, for syndetic n,

$$\int \tilde{f}(x)\tilde{f}(T^n x)\tilde{f}(T^{2n}x)d\mu > \int \tilde{f}(x)^3 d\mu - \varepsilon.$$

So all we need to show is $\int \tilde{f}(x)^3 d\mu > 0$.

Lemma 5. *If $f \geq 0$, then $\tilde{f} \geq 0$.*

Proof. The claim follows from these two points: (i) $\tilde{f} = \max(\tilde{f}, 0)$ because they both minimize the distance from $\mathcal{E}(X, \mathcal{B}, \mu, T)$ to f (ii) $\max(\tilde{f}, 0) \in \mathcal{E}(X, \mathcal{B}, \mu, T)$ as $\max(\tilde{f}, 0) = (f + |f|)/2$. □

Lemma 6. *$\tilde{f}$ is not a.e. 0.*

Proof. This is true because $\int \tilde{f} d\mu = \int f d\mu \neq 0$. □

Corollary. $\int \tilde{f}(x)^3 d\mu > 0$.

This completes the proof of Roth's theorem. □

3.4 The "structured" case

Lemma 7. *If $\psi \in \mathcal{E}(X, \mathcal{B}, \mu, T)$, then for any $\varepsilon > 0$ there exists a syndetic set of n such that $\|T^n\psi - \psi\|_2 < \varepsilon$.*

Proof. It suffices to prove this for a dense set of $\mathcal{E}(X, \mathcal{B}, \mu, T)$, in particular, the algebraic span (i.e. linear combination of finitely many terms) of the eigenvectors of T. Suppose $\psi = \sum_{j=1}^{m} c_i e_i$, $\|\psi\|_2 = 1$ without loss of generality, where $c_i \in \mathbb{C}$ and e_i are eigenvectors with eigenvalue λ_i.

By the Weyl equidistribution theorem (or see Example 2 above), there exists a syndetic set of n such that $|\lambda_i^n - 1| < \varepsilon$ for all $i = 1, \ldots, m$. This implies $\|T^n\psi - \psi\|_2 = \|\sum_{j=1}^m c_i(\lambda_i^n - 1)e_i\|_2 < \varepsilon$, as desired. □

Proof of Proposition 5.

$$\begin{aligned}
&\Big|\int \psi(x)\psi(T^n x)\psi(T^{2n}x)d\mu - \int \psi(x)^3 d\mu\Big| \\
&\leq \int |\psi(x)||\psi(T^n x)||\psi(T^{2n}x) - \psi(x)|d\mu + \int |\psi(x)|^2|\psi(T^n x) - \psi(x)|d\mu \\
&\leq \|\psi T^n\psi\|_2\|T^{2n}\psi - \psi\|_2 + \|\psi\|_2\|T^n\psi - \psi\|_2 \\
&\leq \|\psi\|_\infty^2(\|T^{2n}\psi - T^n\psi\|_2 + \|T^n\psi - \psi\|_2) + \|\psi\|_\infty\|T^n\psi - \psi\|_2 \\
&\leq 3\|\psi\|_\infty^2\|T^n\psi - \psi\|_2.
\end{aligned}$$

The proposition now follows by the previous lemma. □

3.5 The "random" case

We are now up to the second and last part of the proof. There are (at least) two ways to do this, one as in [4], and another as in [12]. They are basically the same proofs, but I will follow the latter one because it is shorter.

Lemma 8 (van der Corput; special case)**.** *Let $\{u_n\}$ be a seqence in a Hilbert space with $\|u_n\| \leq 1$ for all $n \in \mathbb{N}$. For $m \in \mathbb{N}$, set*

$$\gamma_m = \limsup_{N\to\infty}\Big|\frac{1}{N}\sum_{n=0}^{N-1}\langle u_{n+m}, u_n\rangle\Big|.$$

If $\lim_{M\to\infty}\frac{1}{M}\sum_{m=0}^{M-1}\gamma_m = 0$ for all m, then

$$\limsup_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=0}^{N-1}u_n\Big\|^2 = 0.$$

Proof. Given $\varepsilon > 0$ and a fixed $M \in \mathbb{N}$, for N sufficiently large we have that

$$\Big|\frac{1}{N}\sum_{n=0}^{N-1} u_n - \frac{1}{N}\frac{1}{M}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} u_{n+m}\Big| < \varepsilon.$$

By convexity,

$$\begin{aligned}
\Big\|\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{M}\sum_{m=0}^{M-1} u_{n+m}\Big\|^2 &\leq \frac{1}{N}\sum_{n=0}^{N-1}\Big\|\frac{1}{M}\sum_{m=0}^{M-1} u_{n+m}\Big\|^2 \\
&= \frac{1}{N}\frac{1}{M^2}\sum_{n=0}^{N-1}\sum_{m_1,m_2=0}^{M-1}\langle u_{n+m_1}, u_{n+m_2}\rangle \\
&\leq \frac{1}{M^2}\sum_{m_1,m_2=0}^{M-1}\frac{1}{N}\sum_{n=0}^{N-1}\langle u_{n+m_1}, u_{n+m_2}\rangle \\
&\leq \frac{1}{M^2}\sum_{m_1,m_2=0}^{M-1}\gamma_{|m_1-m_2|} \\
&\leq \frac{2}{M}\sum_{m=1}^{M}\frac{1}{M}\sum_{l=0}^{m-1}\gamma_l,
\end{aligned}$$

which approaches 0 as $M \to \infty$, by assumption. □

Proposition 7 (Furstenberg; double convergence [3])**.** *Suppose $(X, \mathcal{B}, \mu, T)$ is ergodic and $g, h \in L^\infty(X, \mathcal{B}, \mu) \cap L^2(X, \mathcal{B}, \mu)$. Then*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} T^n g \cdot T^{2n} h = \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} T^n \tilde{g} T^{2n}\tilde{h}$$

in L^2.

Proof. Without loss of generality, we may assume that $\tilde{g} = 0$.

Let $u_n = T^n g \cdot T^{2n} h$. By the van der Corput lemma and the above assumption, it suffices to show that $\lim_{M\to\infty}\frac{1}{M}\sum_m \gamma_m = 0$ for all m.

$$\langle u_n, u_{n+m}\rangle = \int T^n g \cdot T^{2n} h \cdot T^{n+m}\overline{g} \cdot T^{2n+2m}\overline{h} d\mu$$

$$= \int (g \cdot T^m \overline{g}) \cdot T^n (h \cdot T^{2m} \overline{h}) d\mu.$$

Therefore,

$$\frac{1}{N} \sum_{n=0}^{N-1} \langle u_n, u_{n+m} \rangle = \int (g \cdot T^m \overline{g}) \frac{1}{N} \sum_{n=0}^{N-1} T^n (h \cdot T^{2m} \overline{h}) d\mu.$$

By the von Neumann ergodic theorem applied to the second term, the limit

$$\gamma_m = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle u_n, u_{n+m} \rangle$$

exists, and is equal to

$$\int g \cdot T^m \overline{g} d\mu \cdot \int h \cdot T^{2m} \overline{h} d\mu$$

because T is ergodic. Now, since $\int h \cdot T^{2m} \overline{h} d\mu$ is bounded, say by B,

$$\begin{aligned} \Big| \frac{1}{M} \sum_{m=0}^{M-1} \gamma_m \Big| &\leq |B| \Big| \frac{1}{M} \sum_{m=0}^{M-1} \int g \cdot T^m \overline{g} d\mu \Big| \\ &= |B| \Big| \int g \cdot \frac{1}{M} \sum_{m=0}^{M-1} T^m \overline{g} d\mu \Big| \end{aligned}$$

which vanishes as $M \to \infty$, by the von Neumann ergodic theorem and the fact that $\tilde{g} = 0$. □

Corollary. *If $f, g, h \in L^\infty(X, \mathcal{B}, \mu) \cap L^2(X, \mathcal{B}, \mu)$, then*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int f(x) g(T^n x) h(T^{2n} x) d\mu = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \tilde{f}(x) \tilde{g}(T^n x) \tilde{h}(T^{2n} x) d\mu.$$

Proof. Take the weak limit of $\frac{1}{N} \sum_{n=0}^{N-1} T^n g \cdot T^{2n} h$. □

This is precisely Proposition 6.

4 Synthesis of the two arguments

It is very interesting to compare Gowers's and Furstenberg's proofs of Roth's theorem. Both of these proofs are split into two parts, the case for random (loosely speaking) sets, and the case for structured (again loosely speaking) sets. However, the difficulty of each proof lies in different parts. In Gowers's proof, it is relatively easy to prove the theorem for pseudorandom functions, and the main difficulty lies in proposing and proving the density increment argument. In contrast, the structured case in Furstenberg's proof is rather an easy consequence of Poincaré recurrence, whereas Proposition 6 requires lots of insight.

This observation tempts one to take the easy part of each proof and fuse them together, in the hope of creating the "easiest proof." Such a proof will be conceptually nice as well. Gowers's proof gives a good idea of pseudorandomness in terms of Gowers U^2-norm. But it expresses the idea of structure in a very indirect way, i.e. if the balanced function of a set $A \subset \mathbb{Z}/N\mathbb{Z}$ has a large Fourier coefficient, then A has a "linear bias." Furstenberg has a rigorous definition of what it means to be structured: a set $E \subset \mathbb{Z}$ is structured if its corresponding set A via the correspondence principle is in the rotation factor. However, his notion of randomness is defined covertly, as whatever that is not in the rotation factor. A combination of these two arguments will be the one that tells us explicitly about both randomness and structure, and how the two notions are related.

Let me state the goal here clear: we want a proof of Roth's theorem that utilizes both the Gowers U^2 norm and the Kronecker factor, in the way they were used in the above proofs.

One immediate difficulty is that Gowers's and Furstenberg's arguments are made in very different environments. Roth's theorem is a statement about a subset of $\mathbb{Z}$. Gowers reduces it to a statement about a subset of $\mathbb{Z}/N\mathbb{Z}$; Furstenberg uses the correspondence principle to turn it into a problem about some ergodic system $(X, \mathcal{B}, \mu, T)$. Although $\mathbb{Z}/N\mathbb{Z}$ can be made into an ergodic system equipped with the σ-algebra $2^{\mathbb{Z}/N\mathbb{Z}}$ (the power set of $\mathbb{Z}/N\mathbb{Z}$), the normalized counting measure, and the transformation $Sx = x + 1$, Furstenberg's argument is not very applicable to this system, since it does not satisfy the correspondence principle. It is futile to ignore this and apply Propositions 6 and 5 anyways. Proposition 6 do not give any new information, as the Kronecker factor of $\mathbb{Z}/N\mathbb{Z}$ is simply the entire L^2-space. Proposition 5 does not help either; the syndetic set of n could very well be $\{0, N, 2N, \ldots\}$, so all it tells us is that $\int 1_A(x)^3 d\mu > \int 1_A(x)^3 d\mu - \varepsilon$, which is useless.

So we can only hope that the Gowers U^2-norm is applicable to the space $L^\infty(X, \mathcal{B}, \mu)$ induced by the ergodic system $(X, \mathcal{B}, \mu, T)$. The bad news is that the Gowers norm as defined here is only applicable to functions supported on finite sets. The good news is that there exists a generalization of the Gowers norm that is defined on $L^\infty(X, \mathcal{B}, \mu)$. It is called the

Host-Kra norm,[12] named after its inventors [10], [11]. It is a very difficult object to define and understand, but the one that corresponds to the Gowers U^2-norm has a simple expression, which is a natural extension of the Gowers norm:

$$\|f\|_{HK^2}^4 = \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1}\Big(\int f\cdot T^n f d\mu\Big)^2.$$

Moreover, there is a Host-Kra version of the Generalized von Neumann theorem.

Lemma 9. *Suppose $(X,\mathcal{B},\mu,T)$ is ergodic. If $f_1, f_2 \in L^\infty(X,\mathcal{B},\mu)$ with $\|f_1\|_\infty, \|f_2\|_\infty \le 1$, then, for $i = 1, 2$,*

$$\limsup_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=0}^{N-1} T^n f_1 T^{2n} f_2\Big\|_{L^2(\mu)} \le i\|f_i\|_{HK^2}.$$

If in addition $\|f\|_\infty \le 1$, then

$$\limsup_{N\to\infty}\Big|\frac{1}{N}\sum_{n=0}^{N-1}\int f T^n f_1 T^{2n} f_2 d\mu\Big| \le i\|f_i\|_{HK^2}.$$

Proof. For the first inequality, see [12]. The second inequality follows from the first and Hölder's inequality. □

With this, we can do the same thing as in Proposition 3 to show that pseudorandom sets have 3-progressions.

For the general case, given a set A with $\mu(A) = \delta > 0$, let $f_1 = \tilde{1}_A$ be the orthogonal projection of 1_A into the Kronecker factor, and let $f_2 = 1_A - f_1$. Then

$$\begin{aligned}&\frac{1}{N}\sum_{n=0}^{N-1}\int 1_A T^n 1_A T^{2n} 1_A d\mu\\ &\quad= \frac{1}{N}\sum_{n=0}^{N-1}\int f_1 T^n f_1 T^{2n} f_1 d\mu + \sum_{i,j,k\in\{1,2\},\text{not all }1}\frac{1}{N}\sum_{n=0}^{N-1}\int f_i T^n f_j T^{2n} f_k d\mu.\end{aligned}$$

[12] It is in fact just a seminorm; below we will see an instance of a nonzero function with zero Host-Kra norm.

By Proposition 5, the first term on the right-hand side is arbitrarily close to $\int f_1^3 d\mu > 0$ (because $\int f_1 d\mu = \int 1_A d\mu = \delta$) for syndetic n; for other values of n, it is at least nonnegative. All that remains to show is that $\|f_2\|_{HK^2}$ is small. To do this, we could rely on the Host-Kra version of the Gowers inverse theorem. There does exist such a theorem, and indeed the Host-Kra U^2 norm is as deeply related to the l^4-norm of the Fourier transform as is the Gowers U^2-norm (see [10], [11]). But it will take a lot of space to introduce, so instead we do something much simpler: just note that, by the same argument at the end of Proposition 7, $\|f_2\|_{HK^2} = 0$. The proof is complete.

In fact, what we just proved is precisely Proposition 6, except that here we have $f = g = h$. Note the extreme similarity between the way we proved it then and the way we proved it here using the Host-Kra norm. In particular, the Host-Kra norm is an instance of γ_m in the statement of the van der Corput lemma (which is said to have inspired the Gowers norms, according to [17]), and the work we have done there is done here in the Host-Kra version of the Generalized von Neumann theorem.

Perhaps the greatest lesson from this proof is that the Host-Kra norm measures the amount of the rotation factor of a given function. If f has zero rotation factor, then f has zero Host-Kra norm. And if f has any rotation factor, then we have $\|f\|_{HK^2} = \|\tilde{f}\|_{HK^2}$ by the triangle inequality, and $\|\tilde{f}\|_{HK^2} > 0$ by Proposition 5 and the Generalized von Neumann theorem. Examining Proposition 5 more carefully, we see that the "larger" the rotation factor is, the greater $\|\tilde{f}\|_{HK^2}$ becomes. This unifies Gowers's notion of pseudorandomness and Furstenberg's idea of structure.

References

[1] G. D. Birkhoff, *Proof of the ergodic theorem*, Proc. Natl. Acad. Sci. U.S.A. **17** (1931), 656-660.

[2] P. Erdös and P. Turán, *On some sequences of integers*, J. London Math. Soc. **11** (1936), 261-264.

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

[4] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, M. B. Porter Lectures, Princeton Univ. Press, Princeton, NJ, 1981.

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

[6] B. Green, *Long arithmetic progressions of primes*, Clay Math. Proc. **7** (2007) Amer. Math. Soc., Providence, RI, 149-167.

[7] B. Green, *Montréal notes on quadratic Fourier analysis*, CRM Proc. Lecture Notes **43** (2007) Amer. Math. Soc., Providence, RI, 69-102.

[8] B. Green, *Roth's theorem on progressions of length 3*, Lecture notes for additive combinatorics course (2009), http://www.dpmms.cam.ac.uk/ bjg23/add-combinatorics.html.

[9] B. Green and T. C. Tao, *The primes contain arbitrarily long arithmetic progressions"*, Ann. of Math. **167** (2008), 481-547.

[10] B. Host and B. Kra. *Averaging along cubes*, Modern dynamical systems and applications, Cambridge Univ. Press, Cambridge, 2004, 123-144.

[11] B. Host and B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), No. 1, 397-488.

[12] B. Kra, *Ergodic Methods in Additive Combinatorics*, CRM Proc. Lecture Notes **43** (2007) Amer. Math. Soc., Providence, RI, 103-143.

[13] B. Kra, *personal communication.*

[14] H. Poincaré, *Les méthodes nouvelles de la mécanique céleste.* I. Gathiers-Villars, Paris, 1892; II, 1893; III, 1899.

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

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

[17] T. C. Tao, *A quantitative ergodic theory proof of Szemerédi's theorem*, Electron. J. Combin. **13** (2006). 1 No. 99, 1-49.

[18] T. C. Tao, *The dichotomy between structure and randomness, arithmetic progressions, and the primes*, Proc. of the ICM, Madrid 2006, Vol. 1, 581-608.

[19] J. von Neumann, *Proof of the quasi-ergodic hypothesis*, Proc. Nat. Acad. Sci. U.S.A. **18** (1932), 70-82.

[20] P. Walters, *An introduction to ergodic theory*, Graduate texts in mathematics, Springer-Verlag New York, Inc., New York, NY, 1982.

[21] H. Weyl, *Über die Gibbs'sche Erscheinung und verwandte Konvergenzphänomene*, Rendiconti del Circolo Matematico di Palermo **330** (1910), 377-407.

[22] N. Wiener, *Generalized harmonic analysis*, Acta Math. **55** (1930) No. 1, 117-258.