Corso di Laurea in Matematica

Tesi di Laurea Magistrale

Non elementary methods in combinatorial number theory: Roth's and Sarkozy's theorems

Candidato
Francesco Di Baldassarre

Relatore
Prof. Mauro Di Nasso

Anno Accademico 2015/2016

Contents

Introduction

An area of research in combinatorial number theory deals with finding some kind of arithmetic structure in “large” sets. A set $A \subseteq \mathbb{N}$ is “large” if it has *positive (upper) density*, i.e.

$$\overline{d}(A) = \limsup_{N\to\infty} \frac{|A \cap [1, N]|}{N} > 0$$

A fundamental theorem in this field is

Szemeredi’s theorem (1975). *Let $A \subseteq \mathbb{N}$ such that $\overline{d}(A) > 0$. Then A contains arbitrarily long arithmetic progressions.*

In this work we focus on a weaker version of Szemeredi’s theorem:

Roth’s theorem (1953). *Let $A \subseteq \mathbb{N}$ such that $\overline{d}(A) > 0$. Then $x, x + r, x + 2r \in A$ for some $x, r \in \mathbb{N}$.*

Roth’s theorem can be formulated in a more quantitative manner. Let $r_3(N)$ be the cardinality of the largest subset of $[1, N]$ with no arithmetic progressions of length 3, then Roth’s theorem is equivalent to having

$$\lim_{N\to\infty} \frac{r_3(N)}{N} = 0$$

Some of the techniques used to prove Roth’s theorem focus on directly finding an estimate for $r_3(N)$ while others are focused on getting a decomposition of a set into a “structured” component and a “pseudo-random” component.

In this work we present two different arguments used to prove Roth’s theorem and we translate them to the framework of nonstandard analysis. We also adapt the first method to obtain a proof of

Sarkozy’s theorem (1978). *Let $A \subseteq \mathbb{N}$ such that $\overline{d}(A) > 0$. Then A contains two elements whose difference is a perfect square.*

The first approach we present, usually called *density increment*, takes a set with no arithmetic progressions of length 3 and aims to find an arithmetic progression on which the set has a higher density. If the set has positive density we can iterate this process enough times and obtain a contradiction. As a result, we obtain a certain estimate of $r_3(N)$.

This proof can be translated in the nonstandard setting where we obtain a slightly easier argument at the cost of losing the estimate. However we will be able to easily adapt this nonstandard proof to Sarkozy's theorem.

Another way to prove Roth's theorem relies on the Furstenberg correspondence principle that translates the problem into an ergodic framework. The *ergodic proof* consists in decomposing a set in a weak mixing component and a compact component and then restricting the problem to these components. The use of ergodic tools allows for generalizations but does not allow to obtain an estimate on $r_3(N)$.

Terence Tao was able to obtain an estimate using ideas from the ergodic approach with a method called *energy increment*. His proof relies on decomposing a given function in an almost period component plus a small remainder. The almost periodic component is found using the conditional expectation with respect to a particular σ-algebra.

The nonstandard version of this proof is interesting because one can directly use the properties of the integral over a measure space and, at the same time, rely on the hyperfinite nature of the objects under consideration. This way we can use both continuous and discrete techniques.

In the last chapter we use an estimate on Weyl sums and a property of quadratic recurrence to adapt the nonstandard density increment argument to obtain a proof of Sarkozy's theorem.

Landau asymptotic notation

We introduce here the Landau asymptotic notation which we will be using for the rest of the paper.

Let n be a positive variable and f, g real valued functions of n.

- $g(n) = O(f(n))$ means there exists $C > 0$ such that $|g(n)| \leq Cf(n)$ for every n
- $g(n) = \Omega(f(n))$ means there exists $c > 0$ such that $g(n) \geq cf(n)$ for all sufficiently large n
- $g(n) = \Theta(f(n))$ means both $g(n) = O(f(n))$ and $g(n) = \Omega(f(n))$
- $g(n) = o_{n\to\infty}(f(n))$ means $g(n) = O(a(n)f(n))$ for some sequence $a(n)$ such that $a(n) \to 0$ for $n \to \infty$
- $g(n) = \omega_{n\to\infty}(f(n))$ means $f(n) = o_{n\to\infty}(g(n))$

Parameter dependency is indicated by subscripts, e.g. $g(n) = O_k(f(n))$ means $|g(n)| \leq C_k f(n)$.

Basic definitions

In this chapter we recall the basic definitions and properties regarding Fourier transform, linear bias and Bohr sets. For details refer to [TV06].

Definition 0.1 (Bilinear form)**.** A *bilinear form* on an additive group Z is a map $(\xi, x) \mapsto \xi \cdot x$ from $Z \times Z \to \mathbb{R}/\mathbb{Z}$ which is an homomorphism in each of the variables separately.

The form is called *non-degenerate* if for every $\xi \neq 0$ the map $x \mapsto \xi \cdot x$ is not identically zero and is called *symmetric* if $\xi \cdot x = x \cdot \xi$.

Proposition 0.2. *Every finite additive group has at least one non-degenerate symmetric bilinear form.*

In this paper we will often consider the bilinear form $(\xi, x) \mapsto \frac{\xi x}{N}$ on $\mathbb{Z}_N$ which is symmetric and non-degenerate.

Let Z be a finite additive group with a non-degenerate bilinear form $\xi \cdot x$ and let $\mathbb{C}^Z$ be the space of all the complex valued functions $f \colon Z \to \mathbb{C}$.

Definition 0.3. Let $f \in \mathbb{C}^Z$. We define the *mean* of *expectation* of f to be the quantity

$$\mathbf{E}_Z(f) = \mathbf{E}_{x \in Z} f(x) = \frac{1}{|Z|} \sum_{x \in Z} f(x)$$

Similarly, if $A \subseteq Z$, we define the *density* or *probability* of A as

$$\mathbf{P}_Z(A) = \mathbf{P}_{x \in Z}(A) = \mathbf{E}_Z(\mathbb{1}_A) = \frac{|A|}{|Z|}$$

We will rely heavily on the exponential map $e \colon \mathbb{R}/\mathbb{Z} \to \mathbb{C}$, defined by

$$e(\theta) = e^{2\pi i \theta}$$

Definition 0.4 (Character)**.** Let $\xi \in Z$. We define the associated *character* $e_\xi \in \mathbb{C}^Z$ as

$$e_\xi(x) = e(\xi \cdot x)$$

Proposition 0.5 (Orthogonality properties)**.** *For any $\xi, \xi' \in Z$ we have*

$$\left\langle e_\xi, e_{\xi'} \right\rangle_{\mathbb{C}^Z} = \mathbf{E}_{x\in Z}\, e(\xi \cdot x)\overline{e(\xi' \cdot x)} = \mathbb{1}_{\xi=\xi'}$$

and, for any $x, x' \in Z$ we have

$$\sum_{\xi\in Z} e(\xi \cdot x)\overline{e(\xi \cdot x')} = |Z|\, \mathbb{1}_{x=x'}$$

We can now define the *Fourier transform*.

Definition 0.6 (Fourier transform)**.** Let $f \in \mathbb{C}^Z$. We define its *Fourier transform* $\hat{f} \in \mathbb{C}^Z$ as

$$\hat{f}(\xi) \overset{\text{def}}{=} \left\langle f, e_\xi \right\rangle_{\mathbb{C}^Z} = \mathbf{E}_{x\in Z}\, f(x)\overline{e(\xi \cdot x)}$$

We call $\hat{f}(\xi)$ the *Fourier coefficient* of f at frequency ξ.

Since the characters e_ξ form a complete orthonormal system we have

- the *Parseval identity*

$$\mathbf{E}_Z(|f|^2) = \sum_{\xi\in Z} \left|\hat{f}(\xi)\right|^2$$

- the *Plancherel theorem*

$$\langle f, g\rangle_{\mathbb{C}^Z} = \sum_{\xi\in Z} \hat{f}(\xi)\overline{\hat{g}(\xi)}$$

- the *Fourier inversion formula*

$$f = \sum_{\xi\in Z} \hat{f}(\xi) e_\xi$$

Definition 0.7 (Linear bias)**.** Let $A \subseteq Z$. We define the *linear bias* or *Fourier bias* of A the quantity

$$||A||_{u^2} = \sup_{\xi\in Z} \left|\hat{\mathbb{1}}_A(\xi)\right|$$

Sets with small linear bias are called *linearly uniform* or *pseudo-random*.

More generally we define the linear bias of a function $f \in \mathbb{C}^Z$ as

$$||f||_{u^2} = \sup_{\xi\in Z} \left|\hat{f}(\xi)\right|$$

Remark. The quantity $||A||_{u^2}$ is not monotone, i.e. $A \subseteq B$ does not imply $||A||_{u^2} \leq ||B||_{u^2}$.

We introduce now the norms used in this paper.
The first one is the norm on $L^2(Z)$ defined as

$$\|f\|_{L^2}^2 = \mathbf{E}_{z\in Z}\,|f(z)|^2$$

The second one is the norm on $\mathbb{R}/\mathbb{Z}$ defined as

$$\|\theta + \mathbb{Z}\|_{\mathbb{R}/\mathbb{Z}} = |\theta| \quad \text{if } -\frac{1}{2} \le \theta \le \frac{1}{2}$$

We have the well known bounds

$$4\,\|\theta\|_{\mathbb{R}/\mathbb{Z}} \le |e(\theta) - 1| \le 2\pi\,\|\theta\|_{\mathbb{R}/\mathbb{Z}} \tag{RZ bounds}$$

Theorem 0.8 (Kronecker approximation theorem)**.** *Let* $0 < \theta \le \frac{1}{2}$ *and* $\alpha \in \mathbb{R}$. *Then for any* $N > 0$ *we have*

$$\left|\{n \in (-N, N)\colon\ \|n\alpha\|_{\mathbb{R}/\mathbb{Z}} < \theta\}\right| \ge N\theta$$

In particular we have the following corollary

Corollary 0.9. *For any* $N > 0$ *and* $\alpha \in \mathbb{R}$ *if* $\theta \le \frac{1}{N}$ *then there exists an integer* $0 < h < N$ *such that* $\|h\alpha\|_{\mathbb{R}/\mathbb{Z}} \le \theta$*.*

We have a version of this property for quadratic recurrence too [Sch77] [CLR].

Corollary 0.10 (Quadratic recurrence)**.** *For all* $N \in \mathbb{Z}$ *sufficiently large and* $\alpha \in \mathbb{R}$ *there exists an integer* $1 \le h \le N$ *such that*

$$\left\|h^2\alpha\right\|_{\mathbb{R}/\mathbb{Z}} \le \frac{1}{N^{\frac{1}{10}}}$$

We can now give the definition of Bohr set.

Definition 0.11 (Bohr set)**.** Let $S \subseteq Z$ and $\rho > 0$. We define the *Bohr set* with set of frequencies S and radius ρ as

$$\mathrm{Bohr}(S, \rho) = \{x \in Z\colon \sup_{\xi\in S} \|\xi \cdot x\|_{\mathbb{R}/\mathbb{Z}} < \rho\}$$

The cardinality of S is called *rank* of the Bohr set.

Lemma 0.12 (Size bounds)**.** *If* $S \subseteq Z$ *and* $\rho > 0$ *then we have the lower bounds*

$$\mathbf{P}_Z(\mathrm{Bohr}(S, \rho)) \ge \rho^{|S|}$$

Nonstandard analysis

In this section we introduce the set of hypernaturals and give some basic properties. See [Gol98] for a complete introduction.

Let $\mathcal{U}$ be a non principal ultrafilter on $\mathbb{N}$. We define

$$^*\mathbb{N} = \mathbb{N}^{\mathbb{N}}/\mathcal{U} \quad \text{and} \quad ^*\mathbb{R} = \mathbb{R}^{\mathbb{N}}/\mathcal{U}$$

Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of natural (or real) numbers, we denote with $a = [a_n]$ the equivalence class of $(a_n)_{n\in\mathbb{N}}$.

Definition 0.13 (Internal set)**.** A set $A \subseteq {}^*\mathbb{N}$ is called *internal* if

$$[a_n] \in A \qquad \Longleftrightarrow \qquad \{n \colon a_n \in A_n\} \in \mathcal{U}$$

for some sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{P}(\mathbb{N})$. In that case we write $A = [A_n]$.

Definition 0.14 (Internal cardinality)**.** The *internal cardinality* of an internal subset $A = [A_n]$ is defined as

$$|A|_I = [|A_n|] \in {}^*\mathbb{N}$$

Definition 0.15 (Internal function)**.** A function $f \colon {}^*\mathbb{N} \to {}^*\mathbb{R}$ is called *internal* if

$$f([a_n]) = [f_n(a_n)]$$

for some sequence $(f_n)_{n\in\mathbb{N}}$, $f_n \colon \mathbb{N} \to \mathbb{R}$. In that case we write $f = [f_n]$.

Let $a \colon \mathbb{N} \to \mathbb{R}$ and ${}^*a = [a] \colon {}^*\mathbb{N} \to {}^*\mathbb{R}$. If $N = [N_n]$ we define

$$\sum_{n=1}^{N} {}^*a(n) = \left[\sum_{m=1}^{N_n} a(m)\right]$$

Definition 0.16 (Infinite and infinitesimal)**.** We say $a \in {}^*\mathbb{R}, a > 0$ is *infinitesimal* if $|a| \leq r$ for any $r \in \mathbb{R}^+$. Conversely a number $b \in {}^*\mathbb{R}$ is called *infinite* if $|b| \geq r$ for any $r \in \mathbb{R}$. A number is called *finite* if it is not infinite.

Two hyperreal numbers a, b are *infinitely close* if $a - b$ is infinitesimal. In that case we write $a \approx b$. Conversely we write $a \ll b$ if $a - b$ is infinite.

Theorem 0.17. *Every finite hyperreal a is infinitely close to exactly one real number called* standard part *of a and denoted with* $st\,(a)$.

Transfer principle

The transfer principle is a tool used to "transfer" properties from $\mathbb{R}$ to ${}^*\mathbb{R}$ and vice versa.

Let $\mathcal{R} = (\mathbb{R}, \mathcal{P}, \mathcal{F})$ be the full structure of $\mathbb{R}$ with $\mathcal{P}$ the set of all relations on $\mathbb{R}$ and $\mathcal{F}$ the set of all functions on $\mathbb{R}$ and let ${}^*\mathcal{R} = ({}^*\mathbb{R}, {}^*\mathcal{P}, {}^*\mathcal{F})$ with ${}^*\mathcal{P} = \{{}^*P \colon P \in \mathcal{P}\}$ and ${}^*\mathcal{F} = \{{}^*f \colon f \in \mathcal{F}\}$.

Let $\mathcal{L}_{\mathcal{R}}$ and $\mathcal{L}_{{}^*\mathcal{R}}$ be the languages associated with $\mathcal{R}$ and ${}^*\mathcal{R}$ respectively. We can transform any $\mathcal{L}_{\mathcal{R}}$-sentence ϕ into a $\mathcal{L}_{{}^*\mathcal{R}}$-sentence ${}^*\phi$ by replacing each relation symbol P with ${}^*P$ and each function symbol f with ${}^*f$.

Then we have

Transfer principle. *A $\mathcal{L}_{\mathcal{R}}$ sentence ϕ is true in R if and only if ${}^*\phi$ is true in ${}^*R$*

For instance we can apply this principle to the Kronecker approximation theorem and obtain the following result.

Corollary 0.18 (Kronecker approximation theorem)**.** *Let $\alpha \in {}^*\mathbb{R}$ and $\theta > 0$ a real number. For every $N \in {}^*\mathbb{N}$ such that $N\theta \leq 1$ there exists an hypernatural $0 < h < N$ such that $\|h\alpha\|_{{}^*\mathbb{R}/{}^*\mathbb{Z}} \leq \theta$.*

Proof. Apply the transfer principle to the formula

$$\forall \alpha \in \mathbb{R}\ \forall \theta \in \mathbb{R}_+\ \forall N \in \mathbb{N}\ \Big(N\theta \leq 1 \implies \exists h \in \mathbb{N}\ \ 0 < h < N\ \wedge\ \min_{n \in \mathbb{N}} |h\alpha - n| \leq \theta\Big)$$

which follows from the standard Kronecker approximation theorem. □

Similarly we can prove the nonstandard version of the simultaneous quadratic recurrence.

Corollary 0.19 (Quadratic recurrence)**.** *Let $N \in {}^*\mathbb{N}$ infinite and let $\alpha \in {}^*\mathbb{R}$. Then there exists an hypernatural $1 \leq h \leq N$ such that*

$$\left\|h^2\alpha\right\|_{{}^*\mathbb{R}/{}^*\mathbb{Z}} \leq \frac{1}{N^{\frac{1}{10}}}$$

Loeb measure

In this section we introduce the Loeb measure and recall some theorems regarding the integration with respect to this measure. For a complete exposition see [Cut01].

Let $N \in {}^*\mathbb{N}$ be infinite and let $\mathcal{P}_I([N])$ be the collection of internal subsets of $[1, N]$. We have that $\mathcal{P}_I([N])$ is closed under complement and finite unions and intersections, but it is not closed under countable operations.

Let $\mu\colon \mathcal{P}_I([N]) \to {}^*[0,1]$ be defined as

$$\mu(A) = \frac{|A|_I}{N}$$

then μ is an internal, finitely additive function on $\mathcal{P}_I([N])$. So we have $st(\mu)$ finitely additive on $\mathcal{P}_I([N])$.

Using Carathèodory's theorem we can extend $\mathcal{P}_I([N])$ and $st(\mu)$ to obtain a measure space.

Theorem 0.20. *There is a unique σ-additive extension of $st(\mu)$ to the σ-algebra $\sigma(\mathcal{P}_I([N]))$ generated by $\mathcal{P}_I([N])$. The completion of this measure is called* Loeb measure *and is denoted with μ_L. The completion of $\sigma(\mathcal{P}_I([N]))$ is called* Loeb σ-algebra *and is denoted with $\mathcal{L}$.*

A *Loeb-measurable* function $f\colon [1, N] \to \mathbb{R}$ is a function measurable with respect to $\mathcal{L}$. Loeb-measurable functions are not always internal, however the following theorem holds.

Theorem 0.21. *Let $f\colon [1, N] \to \mathbb{R}$, then the following are equivalent*

1. *f is Loeb-measurable*
2. *there is an internal function $F\colon [1, N] \to {}^*\mathbb{R}$ such that*

$$f(n) \approx F(n)$$

for almost all $n \in [1, N]$ (with respect to μ_L)

Definition 0.22 (Lifting)**.** Let $f\colon [1, N] \to \mathbb{R}$ be a Loeb-measurable function, then a function F as given by Theorem 0.21 is called *lifting* of f.

Remark. If $F\colon [1, N] \to {}^*\mathbb{R}$ is an internal function then F is a lifting of $st\,(F)$ and in particular $st\,(F)$ is Loeb measurable.

Given a Loeb space $([1, N], \mathcal{L}, \mu_L)$ and its originating space $([1, N], \mathcal{P}_I([N]), \mu)$ there are two possible integrals.

There are the internal integral

$$\int_{[1,N]} F(n) d\mu = \sum_{n=1}^{N} F(n)\mu(n) = \frac{1}{N} \sum_{n=1}^{N} F(n)$$

and the classical integral

$$\int_{[1,N]} f(n) d\mu_L$$

defined for any F internal and f Loeb-measurable.

As expected the integral of a Loeb-measurable function and the internal integral of a lifting are related.

Theorem 0.23. *Let F be an internal measurable function bounded by some finite real. Then*

$$st\left(\int_{[1,N]} F d\mu\right) = \int st(F) d\mu_L$$

Theorem 0.24. *If F is a lifting of a Loeb measurable function f and F is bounded by some finite real then*

$$\int_{[1,N]} f d\mu_L = st\left(\int_{[1,N]} F d\mu\right) = st\left(\frac{1}{N} \sum_{n=1}^{N} F(n)\right)$$

Chapter 1

Density Increment

In this chapter we will give a proof of Roth's theorem using a density increment argument [TV06]. We will prove that any sufficiently large set of natural numbers with no arithmetic progression of length 3 must have higher density on some subprogression and we will iterate this process multiple times to reach an absurdum. In the first section we deploy standard methods and obtain a certain estimate while in the second section we use a nonstandard approach which gives a slightly easier argument at the cost of losing the estimate.

1.1 Density Increment

In this section we will obtain the estimate

$$r_3([1,N]) = O\left(\frac{N}{\log\log N}\right)$$

where $r_3([1,N])$ is the cardinality of the largest subset of $[1,N]$ with no arithmetic progressions of length 3.

The steps we will follow are

1. We prove that any set with positive density and no arithmetic progression of length 3 has a strong correlation with some character e_ξ
2. We show that having large correlation with a character e_ξ implies a density increment on some arithmetic progression

If we take N large enough we can iterate the two points enough times and obtain an absurd since density cannot exceed 1.

Let Z be a finite group. It is convenient to define the trilinear form

$$\Lambda_3(f,g,h) = \mathbf{E}_{x,r\in Z}\, f(x)g(x+r)h(x+2r)$$

for any $f, g, h \colon Z \to \mathbb{C}$. In particular we have

$$\Lambda_3(\mathbb{1}_A, \mathbb{1}_A, \mathbb{1}_A) = \mathbf{P}_{x,r\in Z}(x, x+r, x+2r \in A)$$

so this form measures the number of arithmetic progressions of length 3 in A.

Lemma 1.1.1. *Let Z have odd order. For any function $f, g, h \colon Z \to \mathbb{C}$ we have*

$$\Lambda_3(f, g, h) = \sum_{\xi\in Z} \hat{f}(\xi)\hat{g}(-2\xi)\hat{h}(\xi)$$

We can then obtain the estimate

$$|\Lambda_3(f, g, h)| \le \|f\|_{u^2} \|g\|_{L^2}\|h\|_{L^2}$$

Proof. Using the Fourier inversion formula we have

$$f = \sum_{\xi_1\in Z} \hat{f}(\xi_1)e_{\xi_1} \qquad g = \sum_{\xi_2\in Z} \hat{g}(\xi_2)e_{\xi_2} \qquad h = \sum_{\xi_3\in Z} \hat{h}(\xi_3)e_{\xi_3}$$

then by linearity

$$\Lambda_3(f, g, h) = \sum_{\xi_1,\xi_2,\xi_3\in Z} \hat{f}(\xi_1)\hat{g}(\xi_2)\hat{h}(\xi_3)\Lambda_3(e_{\xi_1}, e_{\xi_2}, e_{\xi_3})$$

Using Proposition 0.5 we have

$$\begin{aligned}
\Lambda_3(e_{\xi_1}, e_{\xi_2}, e_{\xi_3}) &= \frac{1}{|Z|^2} \sum_{x,r\in Z} e_{\xi_1}(x)e_{\xi_2}(x+r)e_{\xi_3}(x+2r) = \\
&= \frac{1}{|Z|^2}\left(\sum_{x\in Z} e_{\xi_1}(x)e_{\xi_2}(x)e_{\xi_3}(x)\right)\left(\sum_{r\in Z} e_{\xi_2}(r)e_{\xi_3}(2r)\right) = \\
&= \frac{1}{|Z|^2}\left(\sum_{x\in Z} e((\xi_1+\xi_2)\cdot x)\overline{e(-\xi_3\cdot x)}\right)\left(\sum_{r\in Z} e(\xi_2\cdot r)\overline{e(-2\xi_3\cdot r)}\right) = \\
&= \mathbb{1}_{\xi_1+\xi_2=-\xi_3}\cdot \mathbb{1}_{\xi_2=-2\xi_3} = \mathbb{1}_{\xi_1=\xi_3,\xi_2=-2\xi_3}
\end{aligned}$$

which gives

$$\Lambda_3(f, g, h) = \sum_{\xi\in Z} \hat{f}(\xi)\hat{g}(-2\xi)\hat{h}(\xi)$$

Using Cauchy-Schwartz and the Parseval's identity we obtain the estimate

$$\begin{aligned}
|\Lambda_3(f, g, h)|^2 &\le \sup_{\xi\in Z}\left|\hat{f}(\xi)\right|^2\left(\sum_{\xi\in Z}\hat{g}(-2\xi)\hat{h}(\xi)\right)^2 \le \\
&\le \|f\|_{u^2}^2 \sum_{\xi\in Z}|\hat{g}(-2\xi)|^2\sum_{\xi\in Z}\left|\hat{h}(\xi)\right|^2 = \\
&= \|f\|_{u^2}^2 \|g\|_{L^2}^2\|h\|_{L^2}^2
\end{aligned}$$

□

1.1.1 Correlation to a linear phase

We prove that if a sufficiently large set A does not contain any arithmetic progression of length 3 then it has a strong correlation to a linear phase.

Proposition 1.1.2. *Let $P \subset \mathbb{N}$ be an arithmetic progression and let $A \subseteq P$ $|A| = \delta |P|$ for some $0 < \delta \leq 1$. Assume $|P| \geq \frac{100}{\delta^2}$ and that A does not contain any arithmetic progression of length 3. Then there exists $\xi \in \mathbb{R}/\mathbb{N}$ such that*

$$|\mathbf{E}_{n\in P}(\mathbb{1}_A(n) - \delta)e(n\xi)| = \Omega(\delta^2)$$

Proof. By rescaling we can take $P = [1, N]$, $N \geq \frac{100}{\delta^2}$. Using Bertrand's postulate we can pick a prime p between N and $2N$. We can then identify A with a subset of $\mathbb{Z}_p$ in the usual manner. Then, since in A we only have trivial arithmetic progressions $x, x + r, x + 2r$ with $r = 0$, we obtain

$$\Lambda_3(\mathbb{1}_A, \mathbb{1}_A, \mathbb{1}_A) \overset{\text{def}}{=} P_{x,r\in\mathbb{Z}_p}(x, x + r, x + 2r \in A) = \frac{|A|}{p^2} \leq \frac{\delta N}{N^2} = \frac{\delta}{N} \leq \frac{\delta^3}{100}$$

Let $\mathbb{1}_A = f_U + f_{U^\perp}$ with $f_{U^\perp} = \delta\mathbb{1}_{[1,N]}$ and $f_U = \mathbb{1}_A - f_{U^\perp}$. Then

$$\begin{aligned}
\Lambda_3(f_{U^\perp}, f_{U^\perp}, f_{U^\perp}) &= \mathbf{E}_{x\in\mathbb{Z}_p}\, \mathbf{E}_{r\in\mathbb{Z}_p}\, f_{U^\perp}(x)\, f_{U^\perp}(x + r)\, f_{U^\perp}(x + 2r) = \\
&= \delta^3\, \mathbf{E}_{x\in\mathbb{Z}_p}\, \mathbf{E}_{r\in\mathbb{Z}_p}\, \mathbb{1}_{[1,N]}(x)\, \mathbb{1}_{[1,N]}(x + r)\, \mathbb{1}_{[1,N]}(x + 2r) = \\
&= \frac{\delta^3}{p^2} \sum_{x\in\mathbb{Z}_p} \sum_{r\in\mathbb{Z}_p} \mathbb{1}_{[1,N]}(x)\, \mathbb{1}_{[1,N]}(x + r)\, \mathbb{1}_{[1,N]}(x + 2r) \geq \\
&\geq \frac{\delta^3}{p^2} \sum_{x=1}^{N/3} \sum_{r=1}^{N/3} \mathbb{1}_{[1,N]}(x)\, \mathbb{1}_{[1,N]}(x + r)\, \mathbb{1}_{[1,N]}(x + 2r) = \\
&= \frac{\delta^3}{p^2} \frac{N^2}{9} \geq \frac{\delta^3 N^2}{(2N)^2 9} = \frac{\delta^3}{36}
\end{aligned}$$

Putting the two inequalities together we obtain

$$\begin{aligned}
\frac{\delta^3}{25} &\leq |\Lambda_3(\mathbb{1}_A, \mathbb{1}_A, \mathbb{1}_A) - \Lambda_3(f_{U^\perp}, f_{U^\perp}, f_{U^\perp})| = \\
&= |\Lambda_3(f_U + f_{U^\perp}, f_U + f_{U^\perp}, f_U + f_{U^\perp}) - \Lambda_3(f_{U^\perp}, f_{U^\perp}, f_{U^\perp})| = \\
&= |\Lambda_3(f_U, f_U, f_U) + \Lambda_3(f_U, f_U, f_{U^\perp}) + \cdots + \Lambda_3(f_{U^\perp}, f_{U^\perp}, f_U)| \leq \\
&\leq |\Lambda_3(f_U, f_U, f_U)| + |\Lambda_3(f_U, f_U, f_{U^\perp})| + \cdots + |\Lambda_3(f_{U^\perp}, f_{U^\perp}, f_U)|
\end{aligned}$$

So at least one of the seven terms (all of which contain f_U) must be $\Omega(\delta^3)$. Assume, for instance

$$|\Lambda_3(f_U, f_U, f_{U^\perp})| = \Omega(\delta^3)$$

(the other cases are analogous). Using Lemma 1.1.1 we obtain

$$c\delta^3 \leq |\Lambda_3(f_U, f_U, f_U)| \leq \|f_U\|_u \, \|f_U\|_{L^2}\|f_{U^\perp}\|_{L^2}$$

for some constant c. We have that

$$\begin{aligned}
\|f\|_{L^2}^2 &= \frac{1}{p}\sum_{n=1}^{p}(\mathbb{1}_A(n) - \delta\mathbb{1}_{[1,N]}(n))^2 = \\
&= \frac{1}{p}\sum_{n=1}^{p}\mathbb{1}_A(n) + \frac{1}{p}\delta^2\sum_{n=1}^{p}\mathbb{1}_{[1,N]}(n) - 2\frac{\delta}{p}\sum_{n=1}^{p}\mathbb{1}_{[1,N]}(n) = \\
&= \frac{\delta N}{p} + \frac{\delta^2 N}{p} - \frac{2\delta^2 N}{p} = \frac{\delta N}{p} - \frac{\delta^2 N}{p} \leq \delta - \frac{\delta^2}{2} = O(\delta)
\end{aligned}$$

and $\|f_{U^\perp}\|_{L^2} = \|\delta\mathbb{1}_{[1,N]}\|_{L^2} = O(\delta)$. Thus we obtain, for some constant k,

$$c\delta^3 \leq \|f_U\|_u \, k\delta$$

hence $\|f_U\|_u = \Omega(\delta^2)$ and so there exists $\xi \in \mathbb{Z}_p$ such that

$$\left|\hat{f}_U(\xi)\right| = \left|\mathbf{E}_{n\in[1,N]}(\mathbb{1}_A(n) - \delta)e\left(\frac{n}{p}\xi\right)\right| = \Omega(\delta^2)$$

□

1.1.2 Fragmentation

We show that if a set A has a strong correlation to a linear phase then there exists an arithmetic progression on which A has an higher density. To do so, we need to fragment the interval $[1, N]$ in arithmetic progressions which are sufficiently large and on which the linear phase fluctuates only a little.

Lemma 1.1.3 (Linear fragmentation)**.** *Let $N \geq 1$, $0 < \epsilon \leq 1$, $\xi \in \mathbb{R}/\mathbb{Z}$ a linear character. Then there exists $N' = N'(N, \epsilon)$ such that $\lim_{N\to\infty} N' = \infty$ for any ϵ and a partition $[1, N] = \bigsqcup_{j=1}^{J} P_j \sqcup E$ such that*

1. *P_j arithmetic progression of length N'*
2. *$|E| = O(\epsilon N)$*
3. *$\left|e_\xi(x) - e_\xi(y)\right| = O(\epsilon)$ for any $x, y \in P_j$ and $j \in [1, J]$*

Proof. Let $N' = \sqrt{\epsilon N}$.We can use the Kronecker Approximation Theorem to find a phase $0 < h < N'$ such that $\|\xi \cdot h\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{\epsilon}{N'}$. We partition

$$[1, N] = \bigsqcup_{i=1}^{h} A_i$$

with A_i arithmetic progressions of spacing h, $A_i = \{i + hx\}_{x\in\{1,\cdots,s_i\}}$. We can then divide each A_i into subprogressions $P_{i,j}$ of spacing h and length N'

$$A_i = \left(\bigsqcup_{j=1}^{J_i} P_{i,j}\right) \sqcup E_i$$

Renaming the sets $\{P_{i,j}\}$ in $\{P_k\}$ and defining $E = \bigsqcup_{j=1}^{J} E_j$ we obtain

$$[1, N] = \left(\bigsqcup_{k=1}^{K} P_k\right) \sqcup E$$

By construction $|E_i| < N'$ so we have $|E| < N'h < (N')^2 = \epsilon N$.

Let $x, y \in P_j$, then $x = u + sh$, $y = u + th$. Using the RZ bounds we obtain

$$\begin{aligned}\left|e_\xi(x) - e_\xi(y)\right| &= \left|e_\xi(u + sh) - e_\xi(u + th)\right| = \left|e_\xi(u)(e_\xi(sh) - (e_\xi(th))\right| = \\ &= \left|e_\xi((s - t)h) - 1\right| \leq 2\pi \|\xi \cdot (s - t)h\|_{\mathbb{R}/\mathbb{Z}} \leq \\ &\leq 2\pi |s - t| \|\xi \cdot h\|_{\mathbb{R}/\mathbb{Z}} \leq 2\pi N' \frac{\epsilon}{N'} = O(\epsilon)\end{aligned}$$

□

We can now prove that if a function has a strong correlation with a linear character then it has an higher density in one of the pieces of the fragmentation.

Proposition 1.1.4. *Let P be an arithmetic progression and let $f : P \to \mathbb{R}$ be such that $|f(n)| \leq 1$ for all n and $\mathbf{E}_{n\in P} f(n) = 0$. If $\mathbf{E}_{n\in P} f(n)e_\xi(n) \geq \sigma$ for some $\xi \in \mathbb{R}/\mathbb{Z}$ and $\sigma > 0$, then there exist $P' \subseteq P$ arithmetic progression such that*

$$|P'| = \Omega(\sigma^2 |P|^{\frac{1}{2}}) \quad \text{and} \quad \mathbf{E}_{n\in P'} f(n) \geq \frac{\sigma}{4}$$

Proof. Assume $P = [1, N]$ by rescaling. Using the Fragmentation Lemma 1.1.3 we obtain a partition $[1, N] = \bigsqcup_{j=1}^{J} P_j \sqcup E$. Then we have

$$\sigma \leq \mathbf{E}_{n\in P} f(n)e_\xi(n) = \frac{1}{N}\left(\sum_{j=1}^{J} \sum_{x\in P_j} f(x)e_\xi(x) + \sum_{x\in E} f(x)e_\xi(x)\right)$$

Since $|E| = O(\epsilon N)$ and $|f(x)| \leq 1$ we have $\left|\sum_{x\in E} f(x)e_\xi(x)\right| = O(\epsilon N)$.

Let $h_j = \min P_j$, then

$$\begin{aligned}\left|\sum_{x\in P_j} f(x)e_\xi(x)\right| &= \left|\sum_{x\in P_j} f(x)(e_\xi(x) - e_\xi(h_j) + e_\xi(h_j))\right| \leq \\ &\leq \left|\sum_{x\in P_j} f(x)e_\xi(h_j)\right| + \left|\sum_{x\in P_j} f(x)(e_\xi(x) - e_\xi(h_j))\right| \leq \\ &\leq \left|\sum_{x\in P_j} f(x)\right| + O(\epsilon)\left|P_j\right|\end{aligned}$$

since $x, h_j \in P_j$ implies $\left|e_\xi(x) - e_\xi(h_j)\right| = O(\epsilon)$.

Putting the two estimates together we obtain, for some constant c,

$$\begin{aligned}\sigma \leq \left|\mathbf{E}_{n\in P} f(n)e_\xi(n)\right| &\leq \frac{1}{N}\sum_{j=1}^{J}\left|\sum_{x\in P_j} f(x)\right| + \frac{1}{N}O(\epsilon)\sum_{j=1}^{J}\left|P_j\right| + \frac{1}{N}O(\epsilon N) \leq \\ &\leq \frac{1}{N}\sum_{j=1}^{J}\left|\sum_{x\in P_j} f(x)\right| + c\epsilon\end{aligned}$$

Thus $\sum_{j=1}^{J}\left|\sum_{x\in P_j} f(x)\right| \geq N(\sigma - c\epsilon)$. Moreover, since

$$0 = \sum_{n=1}^{N} f(n) = \sum_{j=1}^{J}\sum_{x\in P_j} f(x) + \sum_{x\in E} f(x)$$

we obtain, for some constant $\overline{c}$,

$$\left|\sum_{j=1}^{J}\sum_{x\in P_j} f(x)\right| = \left|\sum_{x\in E} f(x)\right| = O(\epsilon N) \leq \overline{c}\epsilon N$$

Thus $\sum_{j=1}^{J}\sum_{x\in P_j} f(x) \geq -\overline{c}\epsilon N$.

Using the formula $|x| + x = 2\max(x, 0)$ we obtain

$$\begin{aligned}\sum_{j=1}^{J}\max\left(\sum_{x\in P_j} f(x), 0\right) &= \frac{1}{2}\left(\sum_{j=1}^{J}\left|\sum_{x\in P_j} f(x)\right| + \sum_{j=1}^{J}\sum_{x\in P_j} f(x)\right) \geq \\ &\geq \frac{1}{2}\left(N(\sigma - c\epsilon) - \overline{c}\epsilon N\right) = \frac{1}{2}N(\sigma - (c + \overline{c})\epsilon) \geq \\ &\geq \frac{\sigma}{4}N \geq \frac{\sigma}{4}\sum_{j=1}^{J}\left|P_j\right|\end{aligned}$$

when ϵ is sufficiently small.

By the pidgeonhole principle we can find an index j such that

$$\max\left(\sum_{x\in P_j} f(x), 0\right) \geq \frac{\sigma}{4}\left|P_j\right| > 0$$

Thus

$$\mathbf{E}_{x\in P_j} f(x) = \frac{1}{\left|P_j\right|}\sum_{x\in P_j} f(x) \geq \frac{\sigma}{4}$$

□

We have now the following useful corollary.

Corollary 1.1.5. *Let $A \subset P$, P arithmetic progression, $|A| = \delta|P|$ for some $0 < \delta \leq 1$. Assume $|P| \geq \frac{100}{\delta^2}$ and assume A does not contain an arithmetic progression of length 3. Then there exists a proper arithmetic progression $P' \subseteq P$ with $|P'| = \Omega(\delta^4 |P|^{\frac{1}{2}})$ such that we have the density increment*

$$\mathbf{P}_{P'}(A) \geq \mathbf{P}_P(A) + \Omega(\delta^2)$$

Proof. By Proposition 1.1.2 there exists a linear phase ξ such that

$$\mathbf{E}_{n\in P} f(n)e_{\xi}(n) \geq c\delta^2$$

with $f(n) = \mathbb{1}_A(n) - \mathbf{E}(\mathbb{1}_A)$. Then using Proposition 1.1.4 with $\sigma = c\delta^2$ we can find $P' \subseteq P$ such that

$$|P'| = \Omega(\delta^4 |P|^{\frac{1}{2}})$$

and

$$c\frac{\delta^2}{4} \leq \mathbf{E}_{P'} f(n) = \mathbf{P}_{P'}(A) - \mathbf{P}_P(A)$$

□

1.1.3 Roth's theorem

We can now prove Roth's theorem. More specifically we will prove the estimate

$$r_3([1,N]) = O\left(\frac{N}{\log\log N}\right)$$

Let $A \subseteq [1,N]$ be a non empty set with no arithmetic progression of length 3 and define

$$\delta(N) = \frac{|A \cap [1,N]|}{N}$$

We want to prove

$$\delta(N) = O\left(\frac{1}{\log\log N}\right)$$

Let $Z_0 = [1, N]$, $\delta_0(N) = \delta(N)$. Then applying the Corollary 1.1.5 we have the existence of $Z_1, Z_2, \ldots$ such that

$$\begin{aligned}
&\mathbf{P}_{Z_1}(A) \geq \mathbf{P}_{Z_0}(A) + \Omega(\delta_0^2) \geq \delta(N) + c\delta(N)^2 = \delta_1(N) \\
&\mathbf{P}_{Z_2}(A) \geq \mathbf{P}_{Z_1}(A) + \Omega(\delta_1^2) \geq \delta_1(N) + c\delta_1(N)^2 \geq \delta(N) + 2c\delta(N)^2 = \delta_2(N) \\
&\ldots \\
&\mathbf{P}_{Z_{n+1}}(A) \geq \mathbf{P}_{Z_n}(A) + \Omega(\delta_n^2) \geq \delta(N) + nc\delta(N)^2
\end{aligned}$$

So, after at most $\frac{1}{c\delta(N)}$ steps we have density at least $\delta(N) + \frac{1}{c\delta(N)}c\delta(N)^2 = 2\delta(N)$. Analogously we go from density $2\delta(N)$ to $4\delta(N)$ in $\frac{1}{2c\delta(N)}$ steps, hence, by induction, we reach density $2^l\delta(N)$ in at most

$$\frac{1}{c\delta(N)}\sum_{i=0}^{l}\frac{1}{2^i} \leq \frac{2}{c\delta(N)} = \frac{a}{\delta(N)} \quad \text{steps.}$$

In particular we reach a density greater than 1 in at most $\frac{a}{\delta(N)}$ steps.

From the corollary we have the following estimates on the size of the sub progressions

$$\begin{aligned}
&|Z_1| = \Omega(\delta_0^4\,|Z_0|^{\frac{1}{2}}) \geq k\delta(N)^4N^{\frac{1}{2}} = N_1 \\
&|Z_2| = \Omega(\delta_1^4\,|Z_1|^{\frac{1}{2}}) \geq k\delta(N)^4\delta(N)^2N^{\frac{1}{4}} \geq k\delta(N)^8N^{\frac{1}{4}} = N_2 \\
&|Z_3| = \Omega(\delta_2^4\,|Z_2|^{\frac{1}{2}}) \geq k\delta(N)^4\delta(N)^4N^{\frac{1}{8}} = k\delta(N)^8N^{\frac{1}{8}} = N_3 \\
&\ldots \\
&|Z_{n+1}| = \Omega(\delta_n^4\,|Z_n|^{\frac{1}{2}}) \geq k\delta(N)^8N^{\frac{1}{2^n}} = N_n
\end{aligned}$$

Now let n such that $\mathbf{P}_{Z_n}(A) = \delta_n(N) \geq 1$. Using the estimate above we can take $n \leq \frac{a}{\delta(N)}$. Since A does not contain an arithmetic progression of length 3 we have

$$|Z_{n+1}| < \frac{100}{\delta_n(N)^2}$$

otherwise we could apply the corollary and reach a contradiction. Using the estimates we obtain

$$k\delta(N)^8N^{\frac{1}{2^n}} \leq |Z_{n+1}| \leq 100$$

Hence

$$N^{\frac{1}{2^n}} \leq \frac{b}{\delta(N)^8}$$

We can then apply the logarithmic function to both terms and obtain

$$\frac{1}{2^n} \leq \frac{\log b - 8 \log \delta N}{\log N}$$

Using the inequality for n we have

$$2^{\frac{a}{\delta(N)}} \geq 2^n \geq \frac{\epsilon \log N}{-\alpha \log \delta(N)}$$

We can apply again the logarithmic function and obtain

$$\frac{a}{\delta(N)} \geq \overline{\epsilon} \log \log N - \log(-\alpha \log \delta(N))$$

Thus

$$\overline{\epsilon} \log \log N \leq \frac{a + \delta(N) \log(-\alpha \log \delta(N))}{\delta(N)} \leq \frac{k}{\delta(N)}$$

and so we can conclude

$$\delta(N) \leq \frac{\overline{k}}{\log \log N}$$

1.2 Nonstandard Density Increment

The density increment proof can be easily translated in the nonstandard setting: the strategy adopted is the same as in the standard case with the advantage that we can take an infinite subprogression on which we obtain an increased density. The statement we will prove is the following.

Theorem 1.2.1 (Roth's theorem)**.** *Let $N \in {}^*\mathbb{N}$ infinite and $A \subseteq [1, N]$ be an internal subset such that $\frac{|A|}{N} \not\approx 0$. Then A contains an arithmetic progression of length 3.*

Let us show that this statement implies the usual version of Roth's theorem. Let $A \subseteq \mathbb{N}$ be such that $\overline{d}(A) = \delta > 0$. Then there exists a progression $N_n \to \infty$ such that

$$\frac{|A \cap [1, N_n]|}{N_n} \geq \delta - \frac{1}{n}$$

Let $N = [N_n] \in {}^*\mathbb{N}$ and ${}^*A = [A]$. Then, by construction, N is infinite and $st\left(\frac{|A|}{N}\right) \geq \delta$. Hence $a, a + r, a + 2r \in {}^*A$ for some $a, r \in {}^*\mathbb{N}$ using nonstandard Roth. If we take $a = [a_n]$ and $r = [r_n]$ we obtain $a_n, a_n + r_n, a_n + 2r_n \in A$ $\mathcal{U}$-almost everywhere and in particular A contains an arithmetic progression of length 3.

We define Λ_3 in the nonstandard context the same way we have done in the standard case

$$\Lambda_3(f, g, h) = \mathbf{E}_{x,r\in[1,N]}\, f(x)g(x + r)h(x + 2r) = \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{r=0}^{N-1} f(x)g(x + r)h(x + 2r)$$

1.2.1 Correlation and fragmentation

We can use the transfer principle to prove the correlation in the same way we have done for the standard case.

Proposition 1.2.2. *Let $N \in {}^*\mathbb{N}$ be infinite and $A \subseteq [1, N]$ an internal set with $\frac{|A|}{N} = \delta > 0$ non infinitesimal. If A does not contain any arithmetic progression of length 3 then there exists $\alpha \in {}^*\mathbb{R}$ such that*

$$\frac{\mathbf{E}_{n\in[1,N]}\, |(\mathbb{1}_A(n) - \delta)e(\alpha n)|}{\delta^2} \not\approx 0$$

To fragment the interval $[1, N]$ we can proceed the same way we have done for the standard case with the advantage that we can take the subprogression of infinite length.

Lemma 1.2.3 (Linear fragmentation)**.** *Let N be an infinite hypernatural and $\xi \in {}^*\mathbb{R}/{}^*\mathbb{Z}$. Then there exists $\nu < N$ infinite and a partition*

$$[1, N] = \bigsqcup_{j=1}^{J} P_j \sqcup E$$

with

1. *P_j internal arithmetic progression of length at least ν*
2. *E internal and $\frac{|E|}{N}$ infinitesimal*
3. *$|e(\xi \cdot x) - e(\xi \cdot y)|$ is finite for any $x, y \in P_j$ and $j \in [1, J]$*

Proof. Let $\nu \ll \sqrt{N}$. Using the Kronecher's approximation in the nonstandard context (Theorem 0.18) we can find a phase $h < \nu$ such that $\|\xi \cdot h\|_{{}^*\mathbb{R}/{}^*\mathbb{N}} \leq \frac{1}{\nu}$. Similarly to the standard case we partition

$$[1, N] = \bigsqcup_{i=1}^{h} A_i$$

with N_i subprogressions of spacing h and then partition each A_i in subprogressions of length at least ν

$$A_i = \bigsqcup_{j=1}^{J_i} P_{i,j} \sqcup E_i \quad \text{and} \quad [1, N] = \bigsqcup_{j=1}^{h} P_j \sqcup E$$

By construction $|E| \leq h\nu \leq \nu^2 \ll N$. Using the same inequality as in the standard case we obtain

$$|e(\xi \cdot x) - e(\xi \cdot y)| \leq 2\pi$$

$\square$

We can now apply this lemma to obtain the density increment on an infinite subprogression.

Proposition 1.2.4. *Let $f\colon [1, N] \to {}^*\mathbb{R}$ such that*

$$st\left(\mathbf{E}_{n \in [1,N]} |f(n)e(\xi \cdot n)|\right) \geq \sigma > 0$$

Then there exists $P \subseteq [1, N]$ arithmetic progression of infinite length such that

$$st\left(\mathbf{E}_{n \in P} f(n)\right) = st\left(\frac{1}{|P|} \sum_{n \in P} f(n)\right) \geq \frac{\sigma}{4}$$

Again, the proof is the same as in the standard case.

1.2.2 Roth's theorem

We can now prove Roth's theorem.

Theorem 1.2.5 (Roth's theorem)**.** *Let $N \in {}^*\mathbb{N}$ infinite and $A \subseteq [1,N]$ internal subset with $\frac{|A|}{N} \not\approx 0$. Then A contains an arithmetic progression of length 3.*

Proof. Assume, ad absurdum, that A does not contain any arithmetic progression of length 3. Pick $\delta \in \mathbb{R}_+$ such that $\delta \geq st\left(\frac{|A|}{N}\right)$ and let $f(n) = \mathbb{1}_A(n) - \delta$. Then using Proposition 1.2.2 there exists ξ such that

$$\mathbf{E}_{n \in [1,N]} |f(n)e(\xi \cdot n)| \geq c\delta^2 > 0$$

Using Proposition 1.2.4 there exists $P \subseteq [1,N]$ arithmetic progression of length ν infinite such that

$$\frac{1}{|P|} \sum_{x \in P} (\mathbb{1}_A(x) - \delta \mathbb{1}_{[1,N]}(x)) \geq \frac{c\delta^2}{4}$$

hence

$$\frac{1}{|P|} \sum_{n \in P} \mathbb{1}_A(n) \geq \delta + \overline{c}\delta^2$$

Since ν is infinite and δ is not infinitesimal we can repeat the process enough times and obtain and absurdum since the density of a set cannot exceed one. $\square$

Chapter 2

Ergodic Proof

The aim of this chapter is to prove Roth's theorem using tools from the Ergodic theory [McC99]. We give now all the basic definitions and theorems used in the rest of the chapter. See [Pet89] for a more in depth exposition.

Definition 2.0.1 (Measure preserving system)**.** A *measure preserving system* (*mps* for short) is a tuple $(X, \mathcal{B}, \mu, T)$ where $(X, \mathcal{B}, \mu)$ is a complete probability space and $T : X \to X$ is bijective, T, T^{-1} are measurable and $\mu(A) = \mu(T(A))$ for every $A \in \mathcal{B}$.

We can consider T^n as an operator $L^2(X) \to L^2(X)$ defined as

$$T^n(f) = f \circ T^{-n}$$

This convention is chosen so that we have $T^n(\mathbb{1}_A) = \mathbb{1}_{T^n(A)}$.

Definition 2.0.2 (Ergodic system)**.** A measure preserving system $(X, \mathcal{B}, \mu, T)$ is *ergodic* if every T-invariant set has measure 0 or 1.

Example. Let $X = \mathbb{Z}_N$. Then the system $(X, 2^X, \mu, T)$ with μ uniform measure and $T(x) = x + 1$ is an ergodic measure preserving system.

Definition 2.0.3. Let $v_1, v_2, \ldots, v_n \in V$ normed vector space. We define

1. (convergence in norm) $\lim\limits_{n\to\infty} v_n = v$ if $\lim\limits_{n\to\infty} \|v_n - v\| = 0$
2. (convergence in density) $\mathcal{D}\text{-}\lim\limits_{n\to\infty} v_n = v$ if $\overline{d}\{n \in \mathbb{N} \colon \|v_n - v\| < \epsilon\} > 0$ for any $\epsilon > 0$
3. (Cesàro convergence) $\mathcal{C}\text{-}\lim\limits_{n\to\infty} v_n = v$ if $\lim\limits_{N\to\infty} \dfrac{1}{N} \sum\limits_{n=0}^{N-1} v_n = v$

Remark. Convergence in norm ⇒ convergence in density ⇒ Cesàro convergence

Proposition 2.0.4. *Let $v, v_0, v_1, \dots$ be a bounded sequence in a normed vector space V. Then the following are equivalent:*

1. $\underset{n\to\infty}{\mathcal{C}\text{-lim}}\ \|v_n - v\| = 0$
2. $\underset{n\to\infty}{\mathcal{C}\text{-lim}}\ \|v_n - v\|^2 = 0$
3. $\underset{n\to\infty}{\mathcal{D}\text{-lim}}\ \|v_n - v\| = 0$

Furthermore, any of these statements implies that $\mathcal{C}\text{-lim}_{\,n\to\infty}\, v_n = v$.

We introduce now an in important lemma regarding Cesàro limits.

Definition 2.0.5 (Cesàro supremum)**.** $\underset{n\to\infty}{\mathcal{C}\text{-sup}}\ v_n = \limsup_{N\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} v_n \right\|$

van der Corput lemma. *Let $v_0, v_1, \cdots$ be a bounded sequence of vectors in a Hilbert space. If*

$$\underset{h\to\infty}{\mathcal{C}\text{-lim}}\ \underset{n\to\infty}{\mathcal{C}\text{-sup}}\ \langle v_n, v_{n+h}\rangle = 0$$

then $\underset{n\to\infty}{\mathcal{C}\text{-lim}}\ v_n = 0$.

Another important convergence theorem is the Von Neumann mean ergodic theorem.

Definition 2.0.6 (Conditional expectation)**.** Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and let $\mathcal{B}' \subseteq \mathcal{B}$ be a subalgebra. The *conditional expectation* $\mathbf{E}(\,\cdot|\mathcal{B}')$ is the orthogonal projection map

$$\mathbf{E}(\,\cdot|\mathcal{B}')\colon L^2(X, \mathcal{B}, \mu) \to L^2(X, \mathcal{B}', \mu)$$

von Neumann mean ergodic theorem. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Let $\mathcal{B}^T = \{E \in \mathcal{B} \colon T(E) = E\}$. If $f \in L^2(X)$ then*

$$\mathbf{E}_{0\le n<N}(T^n f) = \frac{1}{N}\sum_{n=0}^{N-1} T^n f \longrightarrow \mathbf{E}(f|\mathcal{B}^T) \quad \textit{in } L^2(X)$$

In particular, if the system is ergodic, the limit is $\mathbf{E}(f)$.

In later sections we will often assume the systems to be ergodic. We remark that one can always decompose a generic system into ergodic components thanks to the following theorem.

Ergodic decomposition. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Let $\mathcal{E}(X)$ be the set of ergodic measures on X. Then there exists a map $\beta\colon X \to \mathcal{E}(X)$ such that the map $\Pi_A\colon x \mapsto \beta_x(A)$ is measurable for any $A \in \mathcal{B}$ and*

$$\mu(A) = \int_X \Pi_A(x) d\mu(x)$$

Furstenberg's correspondence principle

In order to translate combinatorial problems to the ergodic setting we will make use of the following principle [Fur77].

Furstenberg's correspondence principle. *Let $A \subseteq \mathbb{N}$, $\overline{d}(A) > 0$. Then there exist a measure preserving system $(X, \mathcal{B}, \mu, T)$ and an element $E \in \mathcal{B}$ such that $\mu(E) = \overline{d}(A)$ and for any $n_1, \ldots, n_k \in \mathbb{N}$*

$$\mu(E \cap T^{-n_1}(E) \cap \cdots \cap T^{-n_k}(E)) \leq \overline{d}(A \cap (A - n_1) \cap \cdots \cap (A - n_k))$$

2.1 General strategy

We have that Roth's theorem is a particular case of

Szemeredi's theorem. *Let $A \subseteq \mathbb{N}$ be a set with positive density. Then A contains arbitrarily long arithmetic progressions.*

Using Furstenberg's correspondence principle [Zha11] it is proved that Szemeredi's theorem is equivalent to

Furstenberg multiple recurrence theorem. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. For any integer k and any $E \in \mathcal{B}$ with $\mu(E) > 0$ there exists some $n > 0$ such that*

$$\mu(E \cap T^n E \cap T^{2n} E \cap \cdots \cap T^{(k-1)n} E) > 0$$

We consider now the (seemingly) stronger statement

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(E \cap T^n E \cap \cdots \cap T^{(k-1)n} E) > 0 \tag{2.1}$$

and consider the problem in terms of functions.

Definition 2.1.1 (SZ system)**.** We say that a measure preserving system is *SZ* (Szemeredi) of level k if

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int f \, T^n f \, \cdots \, T^{(k-1)n} f \, d\mu > 0 \tag{2.2}$$

for every $f \in L^\infty(X)$, f positive and $\mathbf{E}(f) > 0$.

We call the system SZ if it is SZ of every level.

Conditions (2.1) and (2.2) for all f are equivalent since we can take $f = \mathbb{1}_E$ in one direction and we can approximate f with simple functions in the other direction.

We can now state Roth's theorem in ergodic terms.

Roth ergodic. *Every measure preserving system is SZ of level 3.*

We show that this statement implies the usual version of Roth's theorem.

Let $A \subseteq \mathbb{N}$ such that $\overline{d}(A) = \delta > 0$. Using Furstenberg's correspondence principle there exists a measure preserving system $(X, \mathcal{B}, \mu, T)$ and a set $E \in \mathcal{B}$ such that

$$\mu(E \cap T^{-n_1}(E) \cap \cdots \cap T^{-n_k}(E)) \leq \overline{d}(A \cap (A - n_1) \cap \cdots \cap (A - n_k))$$

for any $n_1, \ldots, n_k \in \mathbb{N}$. Using the ergodic version of Roth's theorem we have that there exists an integer n such that

$$\mu(E \cap T^n E \cap T^{2n} E) > 0$$

hence

$$\overline{d}(A \cap (A - n) \cap (A - 2n)) > 0$$

and in particular there exists $x \in A \cap (A - n) \cap (A - 2n)$ and thus $x, x+n, x+2n \in A$.

To prove Roth's ergodic theorem we will go through three main steps:

1. We prove that a system that is either compact or weak mixing is SZ
2. We partition a generic system in a compact component and a weak mixing one
3. We show that one can always restrict the problem to only one of the two components

Essentially we will prove that every function can be decomposed in a "pseudo-random" component (weak mixing) and a "structured" one (almost periodic).

2.2 Weak mixing systems

The weak mixing component represents the pseudo-random part of our decomposition.

Definition 2.2.1 (Weak mixing system)**.** A measure preserving system $(X, \mathcal{B}, \mu, T)$ is *weak mixing* if

$$\mathcal{D}\text{-}\lim_{n\to\infty} \mu(T^n A \cap B) = \mu(A)\mu(B)$$

for every $A, B \in \mathcal{B}$. Or, equivalently, if

$$\mathcal{D}\text{-}\lim_{n\to\infty} \langle T^n f, g\rangle = \mathbf{E}(f)\,\mathbf{E}(g)$$

for any $f, g \in L^2(X)$.

Proposition 2.2.2. *Every weak mixing system is ergodic.*

Proof. Let A be such that $T(A) = A$, then $\mu(A) = \mu(T^n A \cap A) = \mu(A)^2$ by weak mixing. Hence $\mu(A)$ is either 0 or 1. □

We can give the notion of being weak mixing for a single function.

Definition 2.2.3 (Weak mixing function)**.** A function $f \in L^2(X)$ is *weak mixing* if $\mathcal{D}\text{-}\lim_{n\to\infty} \langle T^n f, f\rangle = 0$.

We have the following characterisation (see e.g. [Tao09]).

Theorem 2.2.4. *A system is weak mixing if and only if every function $f \in L^2(X)$ with mean zero is weak mixing.*

We introduce now a useful property of weak mixing functions.

Proposition 2.2.5. *Let $f \in L^2(X)$ be a weak mixing function. Then for any $g \in L^2(X)$ we have*

$$\mathcal{D}\text{-}\lim_{n\to\infty} \langle T^n f, g\rangle = 0 \quad \textit{and} \quad \mathcal{D}\text{-}\lim_{n\to\infty} \langle f, T^n g\rangle = 0$$

Proof. Since $\langle f, T^n g\rangle = \langle T^{-n} f, g\rangle$ by T-invariance it is sufficient to prove that $\mathcal{D}\text{-}\lim_{n\to\infty} \langle T^n f, g\rangle = 0$. We will prove that $C\text{-}\lim_{n\to\infty} |\langle T^n f, g\rangle|^2 = 0$ and the claim will follow thanks to Proposition 2.0.4.

We have

$$\frac{1}{N}\sum_{n=0}^{N-1} |\langle T^n f, g\rangle|^2 = \left\langle \frac{1}{N}\sum_{n=0}^{N-1} \langle T^n f, g\rangle\, T^n f, g \right\rangle$$

We prove that $C\text{-}\lim_{n\to\infty} \langle T^n f, g\rangle\, T^n f = 0$ (the claims then follow by Cauchy-Schwartz).

Using the van der Corput lemma on the sequence $v_n = \langle T^n f, g\rangle\, T^n f$ it suffices to show

$$\begin{aligned} C\text{-}\lim_{h\to\infty} C\text{-}\sup_{n\to\infty} \langle v_n, v_{n+h}\rangle &= C\text{-}\lim_{h\to\infty} C\text{-}\sup_{n\to\infty} \left\langle \langle T^n f, g\rangle\, T^n f, \left\langle T^{n+h} f, g\right\rangle T^{n+h} f\right\rangle \\ &= C\text{-}\lim_{h\to\infty} C\text{-}\sup_{n\to\infty} \langle T^n f, g\rangle \left\langle T^{n+h} f, g\right\rangle \left\langle T^n f, T^{n+h} f\right\rangle = 0 \end{aligned}$$

Now $\langle T^n f, g\rangle$ and $\left\langle T^{n+h} f, g\right\rangle$ are bounded (since $\|T^n f\|_{L^2} = \|T^{n+h} f\|_{L^2} = \|f\|_{L^2}$) and $\left\langle T^n f, T^{n+h} f\right\rangle = \left\langle f, T^h f\right\rangle$ and thus we obtain

$$\mathcal{C}\text{-}\lim_{h\to\infty} \mathcal{C}\text{-}\sup_{n\to\infty} \langle T^n f, g\rangle \left\langle T^{n+h} f, g\right\rangle \left\langle T^n f, T^{n+h} f\right\rangle \le K \cdot \mathcal{C}\text{-}\lim_{h\to\infty} \left|\left\langle f, T^h f\right\rangle\right| = 0$$

since f is weak mixing.

□

Proof of Theorem 2.2.4. If $(X, \mathcal{B}, \mu, T)$ is weak mixing and f has mean zero then $\mathcal{D}\text{-}\lim_{n\to\infty} \langle T^n f, f\rangle = \mathbf{E}(f)^2 = 0$ by definition. Conversely, if $f, g \in L^2(X)$, then from Proposition 2.2.5 we have

$$\mathcal{D}\text{-}\lim_{n\to\infty} \langle T^n(f - \mathbf{E}(f)), g\rangle = 0$$

since $f - \mathbf{E}(f)$ has mean zero and hence $\mathcal{D}\text{-}\lim_{n\to\infty} \langle T^n f, g\rangle = \mathbf{E}(f)\,\mathbf{E}(g)$. □

We can now prove that a weak mixing system is SZ.

Proposition 2.2.6. *Let $(X, \mathcal{B}, \mu, T)$ be a weak mixing system and k an integer, $k \ge 1$. Let $f_1, f_2, \cdots, f_k \in L^\infty(X)$ and let $a_1, a_2, \cdots, a_k$ be a sequence of distinct non-zero integers. Then*

$$\mathcal{C}\text{-}\lim_{n\to\infty} T^{a_1 n} f_1\, T^{a_2 n} f_2 \cdots T^{a_k n} f_k = \mathbf{E}(f_1)\,\mathbf{E}(f_2)\cdots\mathbf{E}(f_k)$$

in $L^2(X)$.

Proof. By induction on k.

Case $k = 1$. $(X, \mathcal{B}, \mu, T)$ weak mixing $\Rightarrow$ $(X, \mathcal{B}, \mu, T^{a_1})$ weak mixing $\Rightarrow$ ergodic. The thesis then follows from the von Neumann mean ergodic theorem.

Inductive step.

By replacing f_1 with $f_1 - \mathbf{E}(f_1)$ and by using the inductive hypothesis on $f_2, \cdots, f_k$ it suffices to prove $\mathcal{C}\text{-}\lim_{n\to\infty} T^{a_1 n} f_1\, T^{a_2 n} f_2 \cdots T^{a_k n} f_k = 0$.

We apply the van der Corput lemma on the sequence

$$v_r = T^{a_1 r} f_1\, T^{a_2 r} f_2 \cdots T^{a_k r} f_k$$

so it is sufficient to show $\mathcal{C}\text{-}\lim_{h\to\infty} \mathcal{C}\text{-}\sup_{n\to\infty} \langle v_n, v_{n+h}\rangle = 0$.

By definition we have

$$\begin{aligned}
\langle v_n, v_{n+h}\rangle &= \left\langle T^{a_1 n} f_1 \cdots T^{a_k n} f_k,\ T^{a_1(n+h)} f_1 \cdots T^{a_k(n+h)} f_k\right\rangle = \\
&= \left\langle T^{-a_k n}\left(T^{a_1 n} f_1 \cdots T^{a_k n} f_k\right),\ T^{-a_k n}\left(T^{a_1(n+h)} f_1 \cdots T^{a_k(n+h)} f_k\right)\right\rangle = \\
&= \left\langle T^{(a_1 - a_k)n} f_1 \cdots T^{(a_{k-1} - a_k)n} f_{k-1}\, f_k,\ T^{(a_1 - a_k)n} T^{a_1 h} f_1 \cdots T^{(a_{k-1} - a_k)n} T^{a_{k-1} h} f_{k-1}\, T^{a_k h} f_k\right\rangle \\
&= \int_X T^{(a_1 - a_k)n} f_{1,h} \cdots T^{(a_{k-1} - a_k)n} f_{k-1,h}\, f_{k,h}\, d\mu
\end{aligned}$$

where $f_{j,h} = f_j \, T^{a_j h} f_j$.

Using Cauchy-Schwartz and the fact that f bounded implies $f_{h,k}$ bounded it suffices to show

$$C\text{-}\lim_{h\to\infty} \; C\text{-}\sup_{n\to\infty} \; T^{(a_1-a_k)n} f_{1,h} \cdots T^{(a_{k-1}-a_k)n} f_{k-1,h} = 0$$

Applying the induction hypothesis we obtain

$$C\text{-}\lim_{n\to\infty} \; T^{(a_1-a_k)n} f_{1,h} \cdots T^{(a_{k-1}-a_k)n} f_{k-1,h} = \mathbf{E}(f_{1,h}) \cdots \mathbf{E}(f_{k-1,h})$$

Since $f_{j,h}$ are all bounded and f_1 is weak mixing we obtain

$$C\text{-}\lim_{h\to\infty} \; \mathbf{E}(f_{1,h}) = C\text{-}\sup_{h\to\infty} \; \mathbf{E}(f_1 T^{a_1 h} f_1) = C\text{-}\sup_{h\to\infty} \; \left\langle f_1, \; T^{a_1 h} f_1 \right\rangle = 0$$

Hence, putting the last two equations together, we conclude

$$\begin{aligned} C\text{-}\lim_{h\to\infty} \;\; C\text{-}\sup_{n\to\infty} \; T^{(a_1-a_k)n} f_{1,h} \cdots T^{(a_{k-1}-a_k)n} f_{k-1,h} &= \\ = C\text{-}\lim_{h\to\infty} \; \mathbf{E}(f_{1,h}) \cdots \mathbf{E}(f_{k-1,h}) &= 0 \end{aligned}$$

□

Corollary 2.2.7. *With the same assumptions of proposition 2.2.6 we have*

$$C\text{-}\lim_{n\to\infty} \int_X T^{a_1 n} f_1 \; T^{a_2 n} f_2 \cdots T^{a_k n} f_k \; d\mu = \mathbf{E}(f_1)\,\mathbf{E}(f_2) \cdots \mathbf{E}(f_k)$$

Theorem 2.2.8. *A weak mixing system is SZ.*

Proof. Let X be a weak mixing system and $f \in L^\infty(X)$, $f \geq 0$, $\mathbf{E}(f) > 0$.

Then, using the corollary above, we obtain

$$\liminf_{N\to\infty} \frac{1}{N} \int_X \sum_{n=0}^{N-1} f \; T^n f \cdots T^{kn} f \; d\mu = \mathbf{E}(f)\,\mathbf{E}(T^n f) \cdots \mathbf{E}(T^{kn} f) = \mathbf{E}(f)^{k+1} > 0$$

□

2.3 Compact systems

The compact component represents the structured part of our decomposition. A system is compact if its functions are almost periodic.

Definition 2.3.1 (Precompact)**.** Let $W \subseteq X$ with X metric space.

W is called *precompact* if $\overline{W}$ is compact.

W is called *totally bounded* if, for any $\epsilon > 0$, W can be covered with a finite collection of ϵ-balls.

Those two definitions are equivalent if X is complete.

Definition 2.3.2 (Almost periodic)**.** Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. We say $f \in L^2(X)$ is *almost periodic* if its orbit $\{T^n f \colon n \in \mathbb{Z}\}$ is precompact in $L^2(X)$ with the norm topology.

Equivalently, f is almost periodic if the set $\{n \in \mathbb{Z} \colon \|f - T^n f\| < \epsilon\}$ is syndetic for every $\epsilon > 0$.

Definition 2.3.3. A set $A \subseteq \mathbb{N}$ is called *syndetic* if there exists $N \in \mathbb{N}$ such that

$$\mathbb{N} = \bigcup_{n=1}^{N} (A - n)$$

or, equivalently, A has bounded gaps.

Definition 2.3.4 (Compact system)**.** The system $(X, \mathcal{B}, \mu, T)$ is *compact* if every function $f \in L^2(X)$ is almost periodic.

Theorem 2.3.5. *Every compact system is SZ.*

With the given definitions the above theorem is equivalent to the following proposition.

Proposition 2.3.6. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and k an integer, $k > 1$. Let $f \in L^2(X)$ be an almost periodic function such that $f \in L^\infty(X)$, $f > 0$ and $\mathbf{E}(f) > 0$. Then*

$$\liminf_{n\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f\, T^n f \cdots T^{(k-1)n} f \, d\mu > 0$$

Proof. We can assume, without loss of generality, $f < 1$.

Let $\epsilon > 0$. Since f is almost periodic, the set

$$N_\epsilon = \left\{ n \in \mathbb{N} \colon \|f - T^n f\| < \frac{\epsilon}{k2^k} \right\}$$

is syndetic. Let $n \in N_\epsilon$.

Since T is isometric we have

$$\|f - T^n f\|_{L^2} < \frac{\epsilon}{k2^k} \Rightarrow \|T^{in} f - T^{in} T^n f\|_{L^2} = \|T^{in} f - T^{(i+1)n} f\|_{L^2} < \frac{\epsilon}{k2^k}$$

Hence for any $j \in \{1, \ldots, k-1\}$ we have

$$\begin{aligned}\|f - T^{jn} f\|_{L^2} \leq \|f - T^n f\|_{L^2} + \|T^n f - T^{2n} f\|_{L^2} + \cdots + \|T^{(j-1)n} f - T^{jn} f\|_{L^2} < \\ < j \cdot \frac{\epsilon}{k2^k} \leq \frac{\epsilon}{2^k}\end{aligned}$$

So we can write $T^{jn} f = f + g_j$ with $\|g_j\|_{L^2} < \frac{\epsilon}{2^k}$.

Then

$$\begin{aligned}\int_X f\, T^n f \cdots T^{(k-1)n} f \, d\mu = \int_X f(f + g_1) \cdots (f + g_{k-1}) \, d\mu = \\ = \int_X f^k \, d\mu + \sum_i \int_X f^{s_i} \prod_{j \in T_i} g_j = \int_X f^k \, d\mu + \sum_i \langle F_i, \ g_{s_i} \rangle\end{aligned}$$

where F_i is a product of f and some g_j thus

$$\|F_i\|_{L^2} \leq \|f^r\|_{L^\infty} \prod \|g_{s_j}\|_{L^2} \leq 1$$

since $\|f\|_{L^\infty} \leq 1$ and $\|g_{s_j}\|_{L^2} \leq \frac{\epsilon}{2^k} \leq 1$.

So $|\langle F_i, \ g_{s_i} \rangle| \leq \|F_i\|_{L^2} \|g_{s_i}\|_{L^2} \leq \frac{\epsilon}{2^k}$.

Since we have at most 2^k such products we obtain

$$\int_X f\, T^n f \cdots T^{(k-1)n} f \, d\mu \geq \int_X f^k \, d\mu - 2^k \frac{\epsilon}{2^k} = \int_X f^k \, d\mu - \epsilon$$

Let d_ϵ be the maximum gap in N_ϵ (which is syndedic).

Since the above equation holds holds for $n \in N_\epsilon$ and $|[1, N] \cap N_\epsilon| \geq \frac{N}{d_\epsilon}$, we have

$$\begin{aligned}\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f\, T^n f \cdots T^{(k-1)n} f \, d\mu \geq \liminf_{N \to \infty} \frac{1}{N} \cdot \frac{N}{d_\epsilon} \left(\int_X f^k \, d\mu - \epsilon \right) = \\ = \frac{1}{d_\epsilon} \left(\int_X f^k \, d\mu - \epsilon \right) > 0\end{aligned}$$

for ϵ sufficiently small.

□

2.4 Compact operators

In this section we introduce a tool used to find almost periodic functions (see [Tao09]).

Definition 2.4.1 (Hilbert-Schmidt norm)**.** Let H, H' be separable Hilbert spaces with orthonormal basis $(e_a)_{a\in A}$ and $(f_b)_{b\in B}$ respectively. The *Hilbert-Schmidt norm* of a bounded linear operator $\Phi\colon H \to H'$ is the quantity $||\Phi||_{HS}$ defined as

$$\begin{aligned} ||\Phi||^2_{HS} &= \sum_{a\in A} ||\Phi(e_a)||^2_{H'} = \sum_{a\in A}\sum_{b\in B} |\langle \Phi(e_a),\ f_b\rangle|^2 = \\ &= \sum_{a\in A}\sum_{b\in B} |\langle e_a,\ \Phi^*(f_b)\rangle|^2 = \sum_{b\in B} ||\Phi(f_b)||^2_H \end{aligned}$$

where Φ^* is the adjoint of Φ.

Definition 2.4.2 (Hilbert-Schmidt operator)**.** Let H, H' be separable Hilbert spaces with orthonormal basis $(e_a)_{a\in A}$ and $(f_b)_{b\in B}$ respectively. A linear bounded operator $\Phi\colon H \to H'$ is a *Hilbert-Schmidt operator* (HS for short) if its Hilbert-Schmidt norm $||\Phi||_{HS}$ is finite.

Proposition 2.4.3 (Characterisation of HS operators)**.** *An operator*

$$\Phi\colon L^2(X) \to L^2(Y)$$

is a Hilbert-Schmidt operator if and only if there exists some kernel K in $L^2(X\times Y)$ such that

$$\Phi(f) = \Phi_k(f) = K * f \overset{\text{def}}{=} \int_X K(\cdot, y) f(y) dy$$

We have that Hilbert-Schmidt operators are compact.

Proposition 2.4.4. *A Hilbert-Schmidt operator between two separable Hilbert spaces is compact, i.e. the image of any bounded set is precompact.*

Proof. Let $\epsilon > 0$ and let $(e_n)_{n\in\mathbb{N}}$ orthogonal basis of H. Since $||\Phi||_{HS}$ is finite there exist $e_1, \dots, e_N$ such that

$$\sum_{n=1}^{N} ||\Phi(e_n)||^2_{H'} \geq ||\Phi||^2_{HS} - \epsilon^2$$

and in particular $\left\|\Phi(e_j)\right\|^2_{H'} < \epsilon^2$ for any e_j orthogonal to $\mathrm{Span}(e_1, \dots, e_N) = V$.

Since we can write any vector x in the unit ball as $x = v + v'$ with $v \in V$, $v \in V^\perp$ and $||v||_H \leq 1$, $||v'||_H \leq 1$ we have that

$$||\Phi(x)||^2_{H'} \leq ||\Phi(v)||^2_{H'} + ||\Phi(v')||^2_{H'} \leq ||\Phi(v)||^2_{H'} + \epsilon^2$$

Since V is finite dimensional we have that $\{\Phi(v) : \|v\| \leq 1\}_{v \in V}$ is precompact. Hence, from the inequality above, we can conclude that the image of the unit ball under Φ is precompact.

□

We can use compact operators to find almost periodic functions.

Proposition 2.4.5. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system.*
If $\Phi : L^2(X) \to L^2(X)$ is a compact operator which commutes with T then Φf is an almost periodic function for any $f \in L^2(X)$.

Proof. The orbit of f is bounded in $L^2(X)$ so the image set

$$\Phi(\{T^n f\}_{n\in\mathbb{Z}}) = \{\Phi(T^n f)\}_{n\in\mathbb{Z}} = \{T^n(\Phi f)\}_{n\in\mathbb{Z}}$$

is precompact which means Φf almost periodic. □

Since HS operators are compact we have the following corollary.

Corollary 2.4.6. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system.*
If $\Phi : L^2(X) \to L^2(X)$ is a HS operator which commutes with T then Φf is an almost periodic function for any $f \in L^2(X)$.

2.5 Decomposing $L^2(X)$

We now prove that we can decompose $L^2(X)$ in an almost periodic component plus a compact one. Let $AP(X)$ be the set of the almost periodic functions in $L^2(X)$ and $WM(X)$ the set of the weak mixing functions in $L^2(X)$.

Theorem 2.5.1. $L^2(X) = AP(X) \oplus WM(X)$

We divide the proof in three parts:

- (Lemma 2.5.2) $WM(X) \subseteq AP(X)^\perp$
- (Lemma 2.5.3) $AP(X)^\perp \subseteq WM(X)$
- (Lemma 2.5.4) $AP(X)$ is a closed subspace of $L^2(X)$

Lemma 2.5.2. *Let $f \in WM(X)$. Then $\langle f, g\rangle = 0$ for any $g \in AP(X)$.*

Proof. By T-invariance $\langle T^n f, T^n g\rangle = \langle f, g\rangle$ so it suffices to show

$$\mathcal{C}\text{-}\lim_{n\to\infty} |\langle T^n f, T^n g\rangle| = 0$$

Let $\epsilon > 0$. Since g is almost periodic there exist $g_1, \ldots, g_k \in L^2(X)$ such that for any n we have $\|T^n g_i - g\|_{L^2} < \epsilon$ for some i. Then we have

$$|\langle T^n f, T^n g\rangle| \leq |\langle T^n f, g_i\rangle| + \epsilon\|f\|_{L^2} \leq \sum_{i=1}^{k} |\langle T^n f, g_i\rangle| + \epsilon\|f\|_{L^2}$$

Since f is weak mixing we have $\mathcal{D}\text{-}\lim_{n\to\infty} |\langle T^n f, g\rangle| = 0$ for any $g \in L^2(X)$ (see Proposition 2.2.5), and so

$$\mathcal{C}\text{-}\lim_{n\to\infty} |\langle T^n f, T^n g\rangle| \leq \epsilon\|f\|_{L^2}$$

Since ϵ is arbitrary we obtain

$$|\langle f, g\rangle| = \mathcal{C}\text{-}\lim_{n\to\infty} |\langle T^n f, T^n g\rangle| = 0$$

□

Lemma 2.5.3. *Let $f \in AP(X)^\perp$. Then $f \in WM(X)$.*

Proof. We prove that, if f is not weak mixing, then there exists $g \in AP(X)$ such that $\langle f, g\rangle \neq 0$.

Let $\Phi_f \colon L^2(X) \to L^2(X)$, $\Phi_f(g) = \langle f, g\rangle f$. This is a HS operator since

$$\left\|\Phi_f\right\|_{HS}^2 = \sum_{n\in\mathbb{N}} \left\|\Phi_f(e_n)\right\|^2 = \sum_{n\in\mathbb{N}} |\langle f, e_n\rangle|^2 \|f\|_{L^2}^2 = \|f\|_{L^2}^4 \quad \text{is finite.}$$

Let U be the unitary operator on the space $HS(L^2(X) \to L^2(X))$ of the HS operators on $L^2(X)$ given by $U(S) = T \circ S \circ T^{-1}$. So we have $U(\Phi_f) = \Phi_{Tf}$.

By the von Neumann mean ergodic theorem we obtain that

$$\frac{1}{N}\sum_{n=0}^{N-1} U^n f \longrightarrow \Psi_f \quad \text{in norm}$$

with Ψ_f element of the U-invariant subspace of $HS(L^2(X) \to L^2(X))$.

Since Ψ_f is U-invariant we have that $\Psi_f = U\Psi_f = T \circ \Psi_f \circ T^{-1}$ so Ψ_f commutes with T.

Using the Corollary 2.4.6 we obtain that $\Psi_f(g)$ is almost periodic for any $g \in L^2(X)$ and in particular $\Psi_f(f)$ is almost periodic.

Now we have

$$\begin{aligned}\left\langle \Psi_f f, f\right\rangle &= \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \left\langle U^n \Phi_f f, f\right\rangle = \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \left\langle \Phi_{T^n f} f, f\right\rangle = \\ &= \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \langle \langle T^n f, f\rangle T^n f, f\rangle = \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} |\langle T^n f, f\rangle|^2 = \\ &= \mathcal{C}\text{-}\lim_{n\to\infty} |\langle T^n f, f\rangle|^2 \neq 0\end{aligned}$$

(f is not weak mixing $\Rightarrow \mathcal{D}$-lim $_{n\to\infty} \langle T^n f, f\rangle \neq 0$).

So we obtain $\left\langle \Psi_f f, f\right\rangle \neq 0$ and $\Psi_f f \in AP(X)$. □

Lemma 2.5.4. *$AP(X)$ is a closed T-invariant subspace of $L^2(X)$.*

Proof. f and Tf have the same orbit so $f \in AP(X)$ if and only if $Tf \in AP(X)$ hence $AP(X)$ is T-invariant.

We prove than $AP(X)$ is closed.

Let $(f_n) \subseteq AP(X)$, $f_n \longrightarrow f$ in $L^2(X)$. We prove that the orbit of f is totally bounded.

Let $\epsilon > 0$. Choose $m \in \mathbb{N}$ such that $\|f_m - f\| < \frac{\epsilon}{2}$.

Since $f_m \in AP(X)$ there exist $g_1, \dots, g_k \in L^2(X)$ such that for any $n \in \mathbb{N}$ we have $\|T^n f_m - g_i\| < \frac{\epsilon}{2}$ for some i.

So we obtain

$$\begin{aligned}\|T^n f - g_i\|_{L^2} &\leq \|T^n f - T^n f_m\|_{L^2} + \|T^n f_m - g_i\|_{L^2} = \\ &= \|f - f_m\|_{L^2} + \|T^n f_m - g_i\|_{L^2} \leq \frac{\epsilon}{2} + \frac{\epsilon}{2}\end{aligned}$$

Hence the orbit of f it totally bounded and so $f \in AP(X)$. □

We can now prove the main theorem.

Proof of Theorem 2.5.1. Since $AP(X)$ is a closed subspace of $L^2(X)$ and $AP^{\perp}(X) = WM(X)$ we have

$$L^2(X) = AP(X) \oplus AP^{\perp}(X) = AP(X) \oplus WM(X)$$

□

Using this decomposition we can prove the following characterisation of weak mixing systems.

Corollary 2.5.5. *A measure preserving system is weak mixing if and only if the almost periodic functions are constant almost everywhere.*

Proof. Let $f \in L^2(X)$ be an almost periodic function. Then $f - \mathbf{E}(f)$ is almost periodic and has mean zero. Since X is weak mixing we have that $f - \mathbf{E}(f)$ is also weak mixing. Hence, using the decomposition, we obtain $f - \mathbf{E}(f) = 0$ almost everywhere. □

We give now a property of closure of $AP(X)$ which will be used in the next section.

Proposition 2.5.6. *$AP(X)$ is closed under the following pointwise operations*

$$(f, g) \mapsto \max\{f, g\} \quad \textit{and} \quad (f, g) \mapsto \min\{f, g\}$$

Proof. Let $\phi : L^2(X) \times L^2(X) \to L^2(X)$ be a uniformly continuous function which commutes with T. We prove $AP(X)$ closed under ϕ. Let $f, g \in L^2(X)$. We prove that the orbit of $\phi(f, g)$ is totally bounded. Let $\epsilon > 0$. Since ϕ is uniformly continuous we can find $\delta > 0$ such that $\|\phi(f_1, g_1) - \phi(f_2, g_2)\|_{L^2} \leq \epsilon$ for any $\|f_1 - f_2\|_{L^2} < \delta$ and $\|g_1 - g_2\|_{L^2} < \delta$. Since f, g are almost periodic we can find $f_1, \ldots, f_k, g_1, \ldots, g_l \in L^2(X)$ such that for any n there exist indexes i, j such that $\|T^n f - f_i\|_{L^2} < \frac{\delta}{2}$ and $\|T^n g - g_j\|_{L^2} < \frac{\delta}{2}$ hence $\|\phi(T^n f, T^n g) - \phi(f_i, g_j)\|_{L^2} < \epsilon$. Thus the orbit can be covered with ϵ-balls centred at $\phi(f_i, g_j)$. The thesis follows since pointwise $\max\{f, g\}$ and $\min\{f, g\}$ are uniformly continuous functions which commute with T. □

2.6 Decomposing functions

We now give an explicit decomposition of a function in $L^2(X)$ using a Kronecker factor and the conditional expectation.

Definition 2.6.1 (Factor)**.** A *factor* in $(X, \mathcal{B}, \mu, T)$ is a T-invariant subalgebra $\mathcal{B}' \subseteq \mathcal{B}$.

A factor is called *trivial* if its elements have all measure 0 or 1.

A factor is called *compact* if $(X, \mathcal{B}', \mu, T)$ is a compact system.

Definition 2.6.2 (Kronecker factor)**.** Let $\mathcal{B}_{AP} = \{A \in \mathcal{B} \colon \mathbb{1}_A \in AP(X)\}$. Then $\mathcal{B}_{AP}$ is a factor (since $AP(X)$ is closed and T-invariant) and is called *Kronecker factor*.

We can now write a function in $L^2(X)$ as the sum of a structured part and a pseudo-random one.

Proposition 2.6.3. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and $f \in L^2(X)$. Then*

1. *$f \in AP(X)$ if and only if f is $\mathcal{B}_{AP}$-measurable*

2. *$f \in WM(X)$ if and only if $\mathbf{E}(f|\mathcal{B}_{AP}) = 0$*

3. *$f = f_{AP} + f_{WM}$ where $f_{AP} = \mathbf{E}(f|\mathcal{B}_{AP})$ and $f_{WM} = f - f_{AP}$*

Proof. (1) If f is $\mathcal{B}_{AP}$-measurable then f can be approximated by linear combinations of functions $\{\mathbb{1}_A\} \subseteq AP(X)$ and thus $f \in AP(X)$ (since $AP(X)$ is closed).

Conversely let $f \in AP(X)$. We prove $\{x\colon f(x) < a\} \in \mathcal{B}_{AP}$.

We have that

$$\lim_{n\to\infty} \min\{1, \max\{n(f-a), 0\}\} \longrightarrow \mathbb{1}_{\{f(x)<a\}} \quad \text{in } L^2(X)$$

Since $AP(X)$ is closed under pointwise min and max operations, we have $\mathbb{1}_{\{f(x)<a\}} \in AP(X)$ and so $\{f(x) < a\} \in \mathcal{B}_{AP}$.

(2) $f \in WM(X)$ if and only if $f \in AP^{\perp}(X)$ hence $\mathbf{E}(f|\mathcal{B}_{AP}) = 0$

(3) Follows from $L^2(X) = AP(X) \oplus WM(X)$

□

Corollary 2.6.4. *$\mathcal{B}_{AP}$ is the maximal compact factor of $(X, \mathcal{B}, \mu, T)$ and is not trivial if and only if there exists an almost periodic function which is not almost everywhere constant.*

This decomposition induces an interesting dichotomy:

A system is either pseudo-random or contains some structured piece

Proposition 2.6.5. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system.*
Then exactly one of the following holds:

1. *(Pseudo-random) X is weak mixing*

2. *(Structure) X has a non trivial factor*

Proof. If X is not weak mixing then, by Corollary 2.6.4, there exists $f \in AP(X)$ which is not constant a.e., hence $\mathcal{B}_{AP}$ is a non trivial factor. □

2.7 Roth's theorem

We can now prove Roth's theorem. The idea of the proof is that the system is either compact (hence SZ) or contains a non trivial Kronecker factor on which we can "project" the problem.

We begin with an estimate on the distance between two functions and their almost periodic part.

Proposition 2.7.1. *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system. Then for any $f_1, f_2 \in L^2(X)$ we have*

$$\underset{n\to\infty}{C\text{-lim}} \left(T^n f_1\, T^{2n} f_2 - T^n\, \mathbf{E}(f_1|\mathcal{B}_{AP})\, T^{2n}\, \mathbf{E}(f_2|\mathcal{B}_{AP})\right) = 0 \quad \text{in } L^2(X)$$

Proof. Since we can decompose f_1 and f_2 in their weak mixing and almost periodic components it is sufficient to prove

$$\underset{n\to\infty}{C\text{-lim}}\ T^n f_1\, T^{2n} f_2 = 0$$

if either f_1 or f_2 is weak mixing.

Let $v_n = T^n f_1\, T^{2n} f_2$.

Using the van der Corput lemma it suffices to show

$$\underset{h\to\infty}{C\text{-sup}}\ \underset{n\to\infty}{C\text{-sup}}\ \langle v_n, v_{n+h}\rangle = 0$$

Since $T\colon L^2(X) \to L^2(X)$ is linear and μ-invariant we have

$$\begin{aligned}\langle v_n, v_{n+h}\rangle &= \int_X T^n f_1\, T^{2n} f_2\, T^{n+h} f_1\, T^{2(n+h)} f_2\, d\mu = \\ &= \int_X f_1\, T^n f_2\, T^h f_1\, T^{n+2h} f_2\, d\mu = \int_X f_1\, T^h f_1\, T^n(f_2\, T^{2h} f_2)\, d\mu\end{aligned}$$

Since X is ergodic by the mean ergodic theorem we obtain

$$\underset{n\to\infty}{C\text{-lim}} \int_X f_1\, T^h f_1\, T^n(f_2\, T^{2h} f_2)\, d\mu = \int_X f_1\, T^h f_1 \left(\underset{n\to\infty}{C\text{-lim}}\ T^n(f_2\, T^{2h} f_2)\right) d\mu =$$
$$= \int_X f_1\, T^h f_1\ \mathbf{E}(f_2\, T^{2h} f_2)\, d\mu = \left\langle f_1, T^h f_1\right\rangle \left\langle f_2, T^{2h} f_2\right\rangle$$

Since either f_1 or f_2 is weak mixing at least one of the two inner products is zero, hence

$$\underset{h\to\infty}{C\text{-sup}}\ \underset{n\to\infty}{C\text{-sup}}\ \langle v_n, v_{n+h}\rangle = \underset{h\to\infty}{C\text{-sup}} \left\langle f_1, T^h f_1\right\rangle \left\langle f_2, T^{2h} f_2\right\rangle = 0$$

□

We can now prove Roth's theorem.

Theorem 2.7.2 (Roth's theorem)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Then for any $f \in L^\infty(X)$, $f \geq 0$, $\mathbf{E}(f) > 0$ we have*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f\, T^n f\, T^{2n} f\, d\mu > 0$$

Proof. Via ergodic decomposition we may assume that the system is ergodic.

Using the Proposition 2.7.1 we obtain

$$\begin{aligned}
\liminf_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1}\int_X f\, T^n f\, T^{2n} f\, d\mu &= \int_X f\, \underset{n\to\infty}{C\text{-lim}}\, T^n f\, T^{2n} f\, d\mu = \\
&= \int_X f\, \underset{n\to\infty}{C\text{-lim}}\, T^n\, \mathbf{E}(f|\mathcal{B}_{AP})\, T^{2n}\, \mathbf{E}(f|\mathcal{B}_{AP})\, d\mu = \\
&= \liminf_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1}\int_X f\, T^n\, \mathbf{E}(f|\mathcal{B}_{AP})\, T^{2n}\, \mathbf{E}(f|\mathcal{B}_{AP})\, d\mu = \\
&= \liminf_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1}\int_X \mathbf{E}\Big(f\, T^n\, \mathbf{E}(f|\mathcal{B}_{AP})\, T^{2n}\, \mathbf{E}(f|\mathcal{B}_{AP})\Big|\mathcal{B}_{AP}\Big)\, d\mu = \\
&= \liminf_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1}\int_X \mathbf{E}(f|\mathcal{B}_{AP})\, T^n\, \mathbf{E}(f|\mathcal{B}_{AP})\, T^{2n}\, \mathbf{E}(f|\mathcal{B}_{AP})\, d\mu > 0
\end{aligned}$$

since $\mathbf{E}(f|\mathcal{B}_{AP}) > 0$, is almost periodic and has positive mean.

□

Chapter 3

Energy Increment

In this chapter we present a proof of the finitary version of Roth's theorem using an energy increment argument [TV06]. Similarly to the ergodic proof we will decompose a given function in a periodic component and a pseudo-random remainder. To obtain the periodic component we will use the conditional expectation with respect to a particular algebra in such a way that the remainder will have small Fourier coefficients.

3.1 Energy Increment

The statement we will prove in this section is the following

Roth's theorem. *For all finite groups Z and for all $f\colon Z \to \mathbb{R}^+$ with $0 \le f(x) \le 1$ and $\mathbf{E}_Z(f) \ge \delta > 0$ we have*

$$\Lambda_3(f, f, f) = \Omega_\delta(1)$$

Let us show that this statement implies Roth's theorem in its usual form. Let $\overline{d}(A) = \delta > 0$, let $f = \mathbb{1}_A$ and let $N \in \mathbb{Z}$ large enough so that $\frac{|A \cap [1,N]|}{N} \ge \delta$.

Let $Z = \mathbb{Z}_{3N}$ and $\overline{A} = A \cap [1, N] \subset Z$. Then we have

$$\begin{aligned}
0 < c_\delta \le \Lambda_3(\mathbb{1}_{\overline{A}}, \mathbb{1}_{\overline{A}}, \mathbb{1}_{\overline{A}}) = \frac{1}{9N^2} \sum_{k=0}^{N-1} \sum_{n=0}^{N-1} \mathbb{1}_{\overline{A}}(n) \mathbb{1}_{\overline{A}}(n + r) \mathbb{1}_{\overline{A}}(n + 2r) \le \\
\le \frac{1}{9N^2} |\{(x, r)\colon\ x, x + r, x + 2r \in A \cap [1, N]\}| = \\
= \frac{1}{9N^2} (|A \cap [1, N]| + |\{(x, r), r \neq 0\colon\ x, x + r, x + 2r \in A \cap [1, N]\}|) = \\
= \frac{1}{9N^2} \Big(|A \cap [1, N]| + \big|\text{set of non trivial AP of length 3 in } A \cap [1, N]\big|\Big)
\end{aligned}$$

Therefore,

$$\left|\text{set of non trivial AP of length 3 in } A\right| \geq 9N^2 c_\delta - |A \cap [1, N]|$$
$$\geq N(9Nc_\delta - 1) > 1$$

if N is large enough.

In general terms the steps we will follow in the proof are:

1. We define quasi periodic and almost periodic functions
2. We show that functions which are measurable with respect to compact algebras are almost periodic
3. We define the energy of an algebra and prove that the lack of structure causes an increment in energy
4. We use the energy increment to prove we can decompose a function in an almost-periodic component and a "negligible" remainder

3.1.1 Almost periodicity

In this section we define the quasi periodic and almost periodic functions and show that almost periodic functions are recurrent.

Definition 3.1.1 (Quasi periodicity)**.** Let $k > 1$ be an integer. We say $f : Z \to \mathbb{C}$ is *k-quasiperiodic* if there exist frequencies $\xi_1, \ldots, \xi_k \in Z$ (possibly repeated) and $c_1, \ldots, c_k \in \mathbb{C}$ with $|c_1|, \ldots, |c_k| \leq 1$ such that

$$f(x) = \sum_{j=1}^{k} c_j \, e(x \cdot \xi_j)$$

Definition 3.1.2 (Almost periodicity)**.** Let $k \geq 1$ be an integer and $\sigma > 0$. A function $f : Z \to \mathbb{C}$ is (k, σ)-*almost periodic* if there exists a k-quasiperiodic function g such that

$$\|g - f\|_{L^2} \leq \sigma$$

We have a lower bound on $\Lambda_3(f, f, f)$ if f is (k, σ)-almost periodic with σ sufficiently small.

Theorem 3.1.3 (Almost periodicity implies recurrency)**.** *Let $f : Z \to \mathbb{R}^+$, $0 \leq f \leq 1$ and $\mathbf{E}(f) \geq \delta$. If f is (k, σ)-almost periodic for some $k \geq 1$ and $0 < \sigma < \frac{\delta^3}{8}$ then*

$$\Lambda_3(f, f, f) = \Omega\left(\left(\frac{\delta}{k}\right)^k \delta^3\right)$$

Proof. We assume 3 claims which are proved afterwards.

Since f is (k,σ)-almost periodic we have

$$f(x) = \sum_{j=1}^{k} c_j e(x \cdot \xi_j) + g(x)$$

with $\|g\|_{L^2} \leq \sigma$. Let $S = \{\xi_1, \dots, \xi_k\}$ and $\rho > 0$.

If $h \in \mathrm{Bohr}(S,\rho)$, then $e(h \cdot \xi_j) = 1 + O(\rho)$ for every $\xi_j \in S$.

Let $(Tf)(n) = f(n+1)$ be the shift operator. Then we have, for $u = 1, 2$,

$$\|T^{uh} f - f\|_{L^2} \leq O(k\rho) + 2\sigma \tag{Claim 1}$$

Since f is bounded we have $\|T^{uh} f\|_{L^\infty} \leq 1$ and then

$$\left\| f\, T^h f\, T^{2h} f - f^3 \right\|_{L^1} \leq O(k\rho) + 4\sigma \tag{Claim 2}$$

So we obtain

$$-O(k\rho) - 4\sigma \leq \mathbf{E}_Z(f\, T^h f\, T^{2h} f) - \mathbf{E}_Z(f^3) \leq O(k\rho) + 4\sigma$$

and hence

$$\mathbf{E}_Z(f^3) - O(k\rho) - 4\sigma \leq \mathbf{E}_Z(f\, T^h f\, T^{2h} f)$$

Using Hölder's inequality we have $\mathbf{E}_Z(f^3) \geq (\mathbf{E}_Z(f))^3 \geq \delta^3$ and so

$$\begin{aligned} \mathbf{E}_Z(f\, T^h f\, T^{2h} f) &\geq \delta^3 - O(k\rho) - 4\sigma \geq \\ &\geq \delta^3 - O(k\rho) - 4\frac{\delta^3}{8} = \frac{\delta^3}{2} - O(k\rho) \end{aligned}$$

Since this inequality holds for any $h \in \mathrm{Bohr}(S,\rho)$ we can prove

$$\Lambda_3(f,f,f) = \mathbf{E}_{x,h \in Z}(f\, T^h f\, T^{2h} f) \geq \rho^k \left(\frac{\delta^3}{2} - O(k\rho) \right) \tag{Claim 3}$$

We can now take $\rho = \frac{c\delta}{k}$ with $c > 0$ sufficiently small and obtain

$$\Lambda_3(f,f,f) \geq c^k \left(\frac{\delta}{k}\right)^k \left(\frac{\delta^3}{2} - O(c\delta) \right) \geq c^k \left(\frac{\delta}{k}\right)^k \frac{\delta^3}{4}$$

and hence

$$\Lambda_3(f,f,f) = \Omega\left(\left(\frac{\delta}{k}\right)^k \delta^3 \right)$$

□

Proof of claim 1.

$$
\begin{aligned}
T^{uh} f(x) &= \sum_{j=1}^{k} c_j e((x+uh)\cdot \xi_j) + g(x+uh) = \\
&= \sum_{j=1}^{k} c_j e(x\cdot \xi_j) e(uh \cdot \xi_j) + g(x+uh)
\end{aligned}
$$

If $u = 1$ we have $e(h \cdot \xi_j) = e(h \cdot \xi_j) = 1 + O(\rho)$
If $u = 2$ we have $e(2h \cdot \xi_j) = e(h \cdot \xi_j)e(h \cdot \xi_j) = (1 + O(\rho))(1 + O(\rho)) = 1 + O(\rho)$
Then we have

$$
\begin{aligned}
T^{uh} f(x) &= \sum_{j=1}^{k} c_j e(x\cdot \xi_j) e(jh \cdot \xi_j) + g(x+uh) = \\
&= \sum_{j=1}^{k} c_j e(x\cdot \xi_j)(1 + O(\rho)) + T^{uh} g(x) = \\
&= \sum_{j=1}^{k} c_j e(x\cdot \xi_j) + \sum_{j=1}^{k} c_j O(\rho) + T^{uh} g(x) = \\
&= f(x) - g(x) + \sum_{j=1}^{k} c_j O(\rho) + T^{uh} g(x)
\end{aligned}
$$

Hence

$$
\begin{aligned}
\|T^{uh} f - f\|_{L^2} &= \|O(\rho) \sum_{j=1}^{k} c_j + T^{uh} g(x) - g(x)\|_{L^2} \leq \\
&\leq O(\rho k) + \|T^{uh} g\|_{L^2} + \|g\|_{L^2} \leq O(\rho k) + 2\sigma
\end{aligned}
$$

□

Proof of claim 2.

$$
\begin{aligned}
\|f\, T^h f\, T^{2h} f - f^3\|_{L^1} &= \|f\, T^h f\, T^{2h} f - f\, f\, T^{2h} f + f\, f\, T^{2h} f - f^3\|_{L^1} \leq \\
&\leq \|f\, T^{2h} f\|_{L^\infty} \cdot \|T^h f - f\|_{L^1} + \|f^2\|_{L^\infty} \cdot \|T^{2h} f - f\|_{L^1} \leq \\
&\leq \|T^h f - f\|_{L^1} + \|T^{2h} f - f\|_{L^1} \leq \|T^h f - f\|_{L^2} + \|T^{2h} f - f\|_{L^2} \leq \\
&\leq O(k\rho) + 4\sigma
\end{aligned}
$$

□

Proof of claim 3. Using the positivity of f and Lemma 0.12 we obtain

$$\Lambda_3(f,f,f) = \mathbf{E}_{h\in Z}\,\mathbf{E}_{x\in Z}(f\,T^h f\,T^{2h} f) \geq \frac{1}{|Z|}\sum_{h\in \mathrm{Bohr}(S,\rho)} \mathbf{E}_{x\in Z}(f\,T^h f\,T^{2h} f) \geq$$
$$\geq \frac{1}{|Z|}\sum_{h\in \mathrm{Bohr}(S,\rho)} \left(\frac{\delta^3}{2} - O(k\rho)\right) = \mathbf{P}(\mathrm{Bohr}(S,\rho))\cdot\left(\frac{\delta^3}{2} - O(k\rho)\right) \geq$$
$$\geq \rho^{|S|}\left(\frac{\delta^3}{2} - O(k\rho)\right) = \rho^k\left(\frac{\delta^3}{2} - O(k\rho)\right)$$

The claim follows using the positivity of f. □

3.1.2 Compact algebras

In this section we prove that the characters e_ξ generate compact algebras $\mathcal{B}_{\sigma,\epsilon}$ and that, similarly to the ergodic case, functions which are $\mathcal{B}_{\sigma,\epsilon}$-measurable are almost periodic. To identify the almost periodic components we use the conditional expectation.

Definition 3.1.4 (Conditional expectation)**.** Let Z be a finite group, let $\mathcal{B}$ be an algebra of Z and let $f\colon Z\to\mathbb{C}$. We define the *conditional expectation* $\mathbf{E}(f|\mathcal{B})\colon Z\to\mathbb{C}$ to be the function

$$\mathbf{E}(f|\mathcal{B})(x) = \mathbf{E}_{B(x)}\,f = \frac{1}{|B(x)|}\sum_{y\in B(x)} f(y)$$

where $B(x)$ is the unique atom of $\mathcal{B}$ which contains x.

Remark. The algebras on Z can be identified with the possible partitions of Z.

There exist certain algebras $\mathcal{B}$ which are "compact" in the sense that the projections $\mathbf{E}(f|\mathcal{B})$ are almost periodic.

Theorem 3.1.5 (Characters generate compact algebras)**.** *Let $\xi\in Z$ and $0<\epsilon<1$. Then there exists an algebra $\mathcal{B}_{\epsilon,\xi}$ such that:*

1. *$\mathcal{B}_{\epsilon,\xi}$ contains $O_\epsilon(1)$ atoms*
2. *$\mathcal{B}_{\epsilon,\xi}$ contains "approximatively" e_ξ, i.e. $\|e_\xi - \mathbf{E}(e_\xi|\mathcal{B}_{\epsilon,\xi})\|_{L^\infty} \leq \epsilon\sqrt{2}$*
3. *every $\mathcal{B}_{\epsilon,\xi}$-measurable function f with $\|f\|_{L^\infty}\leq 1$ is $(O_{\epsilon,\sigma}(1), O(\sigma))$-almost periodic for every $\sigma>0$*

Proof. Pick α in the unit square (see the remark at the end of the proof)

$$Q = \{z \in \mathbb{C} \colon 0 \le \mathrm{Re}(z), \mathrm{Im}(z) < 1\}$$

We define, for $a, b \in \mathbb{Z}$,

$$\epsilon(Q + a + ib + \alpha) = \left\{z \in \mathbb{C} \colon \frac{z}{\epsilon} \in Q + a + ib + \alpha\right\}$$

and

$$A_{a,b,\epsilon,\alpha} = \left\{x \in Z \colon e(x \cdot \xi) = e_\xi(x) \in \epsilon(Q + a + ib + \alpha)\right\}$$

The sets $\epsilon(Q + a + ib + \alpha)$ form a partition of $\mathbb{C}$ in squares of side ϵ and thus $\{A_{a,b,\epsilon,\alpha} \colon a, b \in \mathbb{Z}\}$ is a partition of Z. Let $\mathcal{B}_{\epsilon,\xi}$ be the algebra generated by $\{A_{a,b,\epsilon,\alpha}\}_{a,b\in\mathbb{Z}}$.

(1) We prove $\mathcal{B}_{\epsilon,\xi}$ contains at most $O\left(\frac{1}{\epsilon}\right)$ atoms.

Let $A_{a,b,\epsilon,\alpha} \neq \emptyset$ and $x \in A_{a,b,\epsilon,\alpha}$. Then we have

$$\frac{e_\xi(x)}{\epsilon} \in Q + a + ib + \alpha$$

Since $e_\xi(x) \in \partial B(0, 1)$ we obtain

$$\partial B\left(0, \frac{1}{\epsilon}\right) \cap Q + a + ib + \alpha \neq \emptyset$$

By construction $\partial B\left(0, \frac{1}{\epsilon}\right)$ is covered with at most, say $\frac{10}{\epsilon} = O\left(\frac{1}{\epsilon}\right)$ squares $Q + a + ib + \alpha$ so we have at most $O\left(\frac{1}{\epsilon}\right)$ non empty atoms $A_{a,b,\epsilon,\alpha} \in \mathcal{B}_{\epsilon,\xi}$.

(2) We prove $\|e_\xi - \mathbf{E}(e_\xi|\mathcal{B}_{\epsilon,\xi})\|_{L^\infty} = O(\epsilon)$

$$\left|e_\xi(x) - \frac{1}{|B(x)|} \sum_{y\in B(x)} e_\xi(y)\right| \le \frac{1}{|B(x)|} \sum_{y\in B(x)} \left|e_\xi(x) - e_\xi(y)\right|$$

If $x, y \in B(x) = A_{a,b,\epsilon,\alpha}$ then $\frac{e_\xi(x)}{\epsilon}, \frac{e_\xi(y)}{\epsilon} \in Q + a + ib + \alpha$ hence

$$\left|e_\xi(x) - e_\xi(y)\right| = \epsilon \left|\frac{e_\xi(x)}{\epsilon} - \frac{e_\xi(y)}{\epsilon}\right| \le \epsilon \sqrt{2}$$

Then

$$\frac{1}{|B(x)|} \sum_{y\in B(x)} \left|e_\xi(x) - e_\xi(y)\right| \le \frac{1}{|B(x)|} |B(x)| \, \epsilon \sqrt{2} \le \epsilon \sqrt{2}$$

and so

$$\|e_\xi - \mathbf{E}(e_\xi|\mathcal{B}_{\epsilon,\xi})\|_{L^\infty} = O(\epsilon)$$

(3) We prove the last property

Let f be $\mathcal{B}_{\epsilon,\xi}$-measurable, then $f = \sum_{a,b} c_{a,b} \mathbb{1}_{A_{a,b,\epsilon,\alpha}}$. Since the number of atoms in $\mathcal{B}_{\epsilon,\xi}$ is $O\left(\frac{1}{\epsilon}\right)$ we have that f is the linear combination of at most $O\left(\frac{1}{\epsilon}\right)$ non trivial characteristic functions so it suffices to prove the claim for the functions $\mathbb{1}_{A_{a,b,\epsilon,\alpha}}$.

If $\sigma \geq 1$ the claim is immediate since we have $\|f\|_{L^2} \leq 1$ so let us suppose $0 < \sigma < 1$. By approximating it suffices to show the claim when $\sigma = \frac{1}{2^n}$ for some n. Assume we have

$$P(A_n) = P\left(\mathbb{1}_{A_{a,b,\epsilon,\alpha}} \text{ is } (O_{\epsilon,n}(1), O(2^{-n}))\text{-almost periodic}\right) = 1 - O(\epsilon 2^{-n}) \tag{3.1}$$

Then we obtain $P(A_n^c) = O(\epsilon 2^{-n})$ and $\sum_{n\in\mathbb{N}} P(A_n^c) < +\infty$. Hence, using the Borel-Cantelli lemma, we obtain

$$P\left(\limsup_{n\to\infty} A_n^c\right) = 0$$

where $\limsup_{n\to\infty} A_n^c = \cap_{n\in\mathbb{N}} \cup_{k\geq n} A_k^c$ and the claim follows.

Let us prove equation (3.1).

By definition of $A_{a,b,\epsilon,\alpha}$ we have

$$\mathbb{1}_{A_{a,b,\epsilon,\alpha}}(x) = \mathbb{1}_Q\left(\frac{e(x\cdot\xi)}{\epsilon} - a - ib - \alpha\right)$$

Let B_n be the $\epsilon 2^{-3n}$ neighbourhood of ∂Q

$$B_n = \left\{z \in \mathbb{C} \colon \exists q \in \partial Q \ |z - q| \leq \frac{\epsilon}{2^{3n}}\right\}$$

Using the Weierstrass approximation theorem we obtain

$$\mathbb{1}_Q(z) - P_{n,\epsilon}(z) = O(\mathbb{1}_{B_n}(z)) + O\left(\frac{1}{2^n}\right)$$

for a suitable polynomial $P_{n,\epsilon}$ and so we have

$$\begin{aligned}\mathbb{1}_{A_{a,b,\epsilon,\alpha}}(x) = P_{n,\epsilon}&\left(\frac{e(x\cdot\xi)}{\epsilon} - a - ib - \alpha\right) + \\ &+ O\left(I\left(\frac{e(x\cdot\xi)}{\epsilon} - a - ib - \alpha \in B_n\right)\right) + \\ &+ O\left(\frac{1}{2^n}\right)\end{aligned}$$

where $I\left(\frac{e(x\cdot\xi)}{\epsilon} - a - ib - \alpha \in B_n\right)$ is the characteristic function of the set

$$\left\{x \in Z \colon \frac{e(x\cdot\xi)}{\epsilon} - a - ib - \alpha \in B_n\right\}.$$

The first term is $O_{n,\epsilon}(1)$-quasi periodic since it is a polynomial of degree depending on n, ϵ only.

Let us consider the second term. Let $I = I\left(\frac{e(x\cdot\xi)}{\epsilon} - a - ib - \alpha \in B_n\right)$, then

$$\mathbf{E}(I) = \sum_{x\in Z} \mathbf{P}\left(\frac{e(x\cdot\xi)}{\epsilon} \in a + ib + \alpha + B_n\right) = O\left(\frac{\epsilon}{2^{3n}}\right)$$

Hence

$$\mathbf{E}(\|I\|_{L^2}^2) = O\left(\frac{\epsilon}{2^{3n}}\right)$$

By Markov's inequality (Proposition 3.1.6) we obtain

$$\mathbf{P}\left(\|I\|_{L^2}^2 \geq \frac{2^n}{\epsilon}\mathbf{E}(\|I\|_{L^2}^2)\right) \leq \frac{\epsilon}{2^n}$$

Since we know that, with probability $1 - O\left(\frac{\epsilon}{2^n}\right)$,

$$\|I\|_{L^2}^2 \leq \frac{2^n}{\epsilon}\mathbf{E}(\|I\|_{L^2}^2) = \frac{2^n}{\epsilon}\cdot O\left(\frac{\epsilon}{2^{3n}}\right) = O\left(\frac{1}{2^{2n}}\right)$$

we obtain

$$\|I\|_{L^2} = O\left(\frac{1}{2^n}\right).$$

Hence

$$\left\|\mathbb{1}_{A_{a,b,\epsilon,\alpha}} - P_{n,\epsilon}\left(\frac{e(x\cdot\xi)}{\epsilon} - a - ib - \alpha\right)\right\|_{L^2} = O\left(\frac{1}{2^n}\right)$$

which implies

$$\mathbb{1}_{A_{a,b,\epsilon,\alpha}} \text{ is } (O_{\epsilon,n}(1), O(2^{-n}))\text{-almost periodic}$$

with probability $1 - O\left(\frac{\epsilon}{2^n}\right)$. This proves (3.1) and the thesis is reached. □

Remark. Given ϵ and ξ we can construct several (possibly different) algebras $\mathcal{B}_{\epsilon,\xi}$ by taking every time a different α in the unit square.

We recall Markov's inequality.

Proposition 3.1.6 (Markov's inequality)**.** *Let X be a non negative random variable. Then for any positive real $\lambda > 0$*

$$\mathbf{P}(X \geq \lambda) \leq \frac{\mathbf{E}(X)}{\lambda}$$

One can extend the result to an algebra generated by multiple characters.

Corollary 3.1.7. *Let $\xi_1, \dots, \xi_n \in Z$ and $\epsilon_1, \dots, \epsilon_n > 0$. Let*

$$\mathcal{B} = \mathcal{B}_{\epsilon_1,\xi_1} \vee \cdots \vee \mathcal{B}_{\epsilon_n,\xi_n}$$

with $\mathcal{B}_{\epsilon_i,\xi_i}$ defined in the previous theorem. Then every $\mathcal{B}$-measurable function f with $\|f\|_{L^\infty} \leq 1$ is $(O_{\epsilon_1,\dots,\epsilon_n,n,\sigma}(1), O_n(\sigma))$-almost periodic for every $\sigma > 0$.

Proof. As for the main theorem, since $\mathcal{B}$ has at most $O_{\epsilon_1,\dots,\epsilon_n,n,\sigma}(1)$ atoms, it suffices to show the claim for $f = \mathbb{1}_A$ with A atom of $\mathcal{B}$. Since $\mathbb{1}_A$ is the product of $\mathbb{1}_{A_j}$ with A_j atom of $\mathcal{B}_{\epsilon_i,\xi_i}$ the claim follows from the previous theorem and the fact that product of bounded almost periodic functions is bounded and almost periodic. □

3.1.3 Energy increment

In this section we define the energy in respect to a function and show that lack of uniformity implies an energy increment.

Definition 3.1.8 (Energy)**.** We define the *energy* of $\mathcal{B}$ with respect to f to be

$$\mathcal{E}_f(\mathcal{B}) = \|\,\mathbf{E}(f|\mathcal{B})\|_{L^2}^2 = \mathbf{E}_{x\in Z}\,|\mathbf{E}(f|\mathcal{B})(x)|^2$$

We show that if $f - \mathbf{E}(f|\mathcal{B})$ has one large Fourier coefficient then we can find a new algebra with more energy with respect to f. More precisely

Theorem 3.1.9 (Lack of uniformity implies energy increment)**.** *Let $\epsilon, \mu > 0$, $\epsilon \leq \frac{\mu}{4}$ and $f\colon Z \to \mathbb{R}$, $0 \leq f \leq 1$. Let $\mathcal{B}$ be such that $\|f - \mathbf{E}(f|\mathcal{B})\|_{u^2} \geq \mu$. Then there exists a frequency $\xi \in Z$ such that*

$$\mathcal{E}_f(\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi}) \geq \mathcal{E}_f(\mathcal{B}) + \frac{\mu^2}{4}$$

Proof. By definition

$$\sup_{\xi\in Z}\left|\left\langle f - \mathbf{E}(f|\mathcal{B}), e_\xi\right\rangle_{L^2}\right| = \|f - \mathbf{E}(f|\mathcal{B})\|_{u^2} \geq \mu$$

Using the second point of Theorem 3.1.5 we obtain

$$\|e_\xi - \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi})\|_{L^\infty} \leq \|e_\xi - \mathbf{E}(e_\xi|\mathcal{B}_{\epsilon,\xi})\|_{L^\infty} \leq 2\epsilon$$

Since $\|f - \mathbf{E}(f|\mathcal{B})\|_{L^\infty} \leq 1$ we have

$$\begin{aligned}
\frac{\mu}{2} \geq 2\epsilon &\geq \left|\left\langle f - \mathbf{E}(f|\mathcal{B}), e_\xi - \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi})\right\rangle\right| = \\
&= \left|\left\langle f - \mathbf{E}(f|\mathcal{B}), e_\xi\right\rangle - \left\langle f - \mathbf{E}(f|\mathcal{B}), \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi})\right\rangle\right| \geq \\
&\geq \left|\left\langle f - \mathbf{E}(f|\mathcal{B}), e_\xi\right\rangle\right| - \left|\left\langle f - \mathbf{E}(f|\mathcal{B}), \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi})\right\rangle\right| \geq \\
&\geq \mu - \left|\left\langle f - \mathbf{E}(f|\mathcal{B}), \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi})\right\rangle\right|
\end{aligned}$$

Hence

$$\left|\left\langle f - \mathbf{E}(f|\mathcal{B}), \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi})\right\rangle\right| \geq \frac{\mu}{2}$$

Since $\mathbf{E}(f|\mathcal{B})$ is the projection on the subspace of $\mathcal{B}$-measurable functions we have

$$\langle f - \mathbf{E}(f|\mathcal{B}), g\rangle = 0 \quad \text{for every } g\ \mathcal{B}\text{-measurable}$$

In particular, if we consider the algebra $\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi}$,

$$\left\langle f - \mathbf{E}(f|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi}), \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi})\right\rangle = 0$$

and so

$$\begin{aligned}
\frac{\mu}{2} &\leq \left|\left\langle f - \mathbf{E}(f|\mathcal{B}), \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi})\right\rangle\right| = \\
&= | \left\langle f - \mathbf{E}(f|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi}), \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi})\right\rangle + \\
&+ \left\langle \mathbf{E}(f|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi}) - \mathbf{E}(f|\mathcal{B}), \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi})\right\rangle | = \\
&= \left|\left\langle \mathbf{E}(f|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi}) - \mathbf{E}(f|\mathcal{B}), \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi})\right\rangle\right| \leq \| \mathbf{E}(f|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi}) - \mathbf{E}(f|\mathcal{B})\|_{L^2}
\end{aligned}$$

Hence, using Pythagoras' theorem, we obtain

$$\begin{aligned}
\mathcal{E}_f(\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi}) &= \| \mathbf{E}(f|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi})\|_{L^2}^2 = \\
&= \| \mathbf{E}(f|\mathcal{B} \vee \mathcal{B}_{\epsilon,\xi}) - \mathbf{E}(f|\mathcal{B})\|_{L^2}^2 + \| \mathbf{E}(f|\mathcal{B})\|_{L^2}^2 \geq \\
&\geq \frac{\mu^2}{4} + \mathcal{E}_f(\mathcal{B})
\end{aligned}$$

□

3.1.4 Koopman-von Neumann decomposition

In this section we use the energy increment argument developed in the previous section to prove the Koopman-von Neumann decomposition.

Theorem 3.1.10 (Koopman-von Neumann decomposition)**.** *Let $f\colon Z \to \mathbb{R}$ such that $0 \leq f \leq 1$ and let $\sigma > 0$. Let $F\colon \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ be an arbitrary function. Then there exists $k = O_{\sigma,F}(1)$ and a decomposition $f = f_{U^\perp} + f_U$ such that*

1. $0 \leq f_{U^\perp} \leq 1$, $\mathbf{E}_Z(f_{U^\perp}) = \mathbf{E}_Z(f)$ *and $f_{U^\perp}$ is (k, σ)-almost periodic*
2. $\|f_U\|_{u^2} \leq \dfrac{1}{F(\sigma, k)}$

Proof. We construct $\mathcal{B}'$, $\mathcal{B}$ and $k \geq 1$ from the following algorithm.

Let us start with $\mathcal{B} = \{\emptyset, Z\}$.

Step 1 Let k be the smallest integer such that $E(f|\mathcal{B})$ is $(k, \frac{\sigma}{2})$-almost periodic (such k exists always thanks to the Fourier inversion formula). Set $\mathcal{B}' = \mathcal{B}$. Then we have (trivially)

$$\mathcal{E}_f(\mathcal{B}') \leq \mathcal{E}_f(\mathcal{B}) + \frac{\sigma^2}{4}$$

Step 2 If $\|f - \mathbf{E}(f|\mathcal{B}')\|_{u^2} \leq \frac{1}{F(\sigma,k)}$ then terminate.

Else define $\mu = \frac{1}{F(\sigma,k)}$ and $\epsilon = \frac{\mu}{4}$. Then, using the Theorem 3.1.9, we can find a frequency ξ and an algebra $\mathcal{B}_{\epsilon,\xi}$ such that

$$\mathcal{E}_f(\mathcal{B}' \vee \mathcal{B}_{\epsilon,\xi}) \geq \mathcal{E}_f(\mathcal{B}') + \frac{1}{4F(\sigma, k)^2}$$

We define $\mathcal{B}'' = \mathcal{B} \vee \mathcal{B}_{\epsilon,\xi}$ and go to step 3.

Step 3 If $\mathcal{E}_f(\mathcal{B}'') \leq \mathcal{E}_f(\mathcal{B}) + \frac{\sigma^2}{4}$ then define $\mathcal{B}' = \mathcal{B}''$ and go to step 2.

Else define $\mathcal{B} = \mathcal{B}''$ and go to step 1.

We prove the algorithm terminates.

We can go from step 3 to step 2 at most $4F(\sigma, k)^2$ many times since at every step the energy $\mathcal{E}_f(\mathcal{B}')$ increases by at least $\frac{1}{4F(\sigma,k)^2}$ and $\mathcal{E}_f(\mathcal{B}') \leq 1$ (since $\|f\|_{L^\infty} \leq 1$).

We can go from step 3 to step 1 at most $\frac{4}{\sigma^2}$ many times since at every step the energy $\mathcal{E}_f(\mathcal{B})$ increases of at least $\frac{\sigma^2}{4}$. Hence the algorithm terminates.

We now define $f_{U^\perp} = \mathbf{E}(f|\mathcal{B}')$ and $f_U = f - f_{U^\perp}$. By construction we have

1. $\|f_U\|_{u^2} = \|f - f_{U^\perp}\|_{u^2} = \|f - \mathbf{E}(f|\mathcal{B}')\|_{u^2} \leq \frac{1}{F(\sigma,k)}$
2. $0 \leq f_{U^\perp} \leq 1$
3. $\mathbf{E}_Z(f_{U^\perp}) = \mathbf{E}_Z(f)$

Since we have $\mathcal{E}_f(\mathcal{B}') \leq \mathcal{E}_f(\mathcal{B}) + \frac{\sigma^2}{4}$ we obtain, by Pythagoras' theorem,

$$\begin{aligned}\|f_{U^\perp} - \mathbf{E}(f|\mathcal{B})\|_{L^2}^2 &= \|\mathbf{E}(f|\mathcal{B}')\|_{L^2}^2 - \|\mathbf{E}(f|\mathcal{B})\|_{L^2}^2 = \\ &= \mathcal{E}_f(\mathcal{B}') - \mathcal{E}_f(\mathcal{B}) \leq \frac{\sigma^2}{4}\end{aligned}$$

hence $\|f_{U^\perp} - \mathbf{E}(f|\mathcal{B})\|_{L^2} \leq \frac{\sigma}{2}$.

Since by construction $\mathbf{E}(f|\mathcal{B})$ is $(k, \frac{\sigma}{2})$-almost periodic there exists a k-quasiperiodic function g such that $\|\mathbf{E}(f|\mathcal{B}) - g\|_{L^2} \leq \frac{\sigma}{2}$, so

$$\|f_{U^\perp} - g\|_{L^2} \leq \frac{\sigma}{2} + \frac{\sigma}{2} = \sigma$$

which means $f_{U^\perp}$ is (k,σ)-almost periodic.

We are left to verify $k = O_{\sigma,F}(1)$.

Let k_n be the value of k at the n-th step of the algorithm, i.e. the n-th time we are in step 1. We prove $k_n = O_{\sigma,F,n}(1)$ by induction.

In the first step we have $k_1 = 1$ trivially (since $\mathbf{E}(f|\mathcal{B})$ is constant) so let us consider the inductive step $n+1$. Let $\mathcal{B}_n$ be the algebra $\mathcal{B}$ obtained at then n-th step. Then

$$\mathcal{B}_{n+1} = \mathcal{B}_n \vee \mathcal{B}_{\epsilon_1,\xi_1} \vee \cdots \vee \mathcal{B}_{\epsilon_r,\xi_r} = \mathcal{B}_{\epsilon_1,\xi_1} \vee \cdots \vee \mathcal{B}_{\epsilon_m,\xi_m}$$

where r is the number of times we went from step 3 to step 2 plus one.

Then $r \le 4F(\sigma,k_n)^2 = O_{\sigma,F,n}(1)$ since $k_n = O_{\sigma,F,n}(1)$ by inductive hypothesis.

By inductive hypothesis we also have $m = O_{\sigma,F,n}(1)$.

By Corollary 3.1.7 we can find two constants $C_{\epsilon_1,\dots,\epsilon_m,m,\sigma'}$ and D_m such that $\mathbf{E}(f|\mathcal{B}_{n+1})$ is $(C_{\epsilon_1,\dots,\epsilon_m,m,\sigma'}, D_m\sigma')$-almost periodic for any σ'. If we take σ' such that $D_m\sigma' \le \frac{\sigma}{2}$ we obtain $k_n \le C_{\epsilon_1,\dots,\epsilon_m,m,\sigma'}$.

Since $m = O_{\sigma,F,n}(1)$ we have $\sigma' = O_{\sigma,F,n}(1)$ and since $\epsilon_1,\dots,\epsilon_r = \frac{1}{4F(\sigma,k_n)} = \Omega_{\sigma,F,n}(1)$ we obtain $k_n = O_{F,\sigma,n}(1)$.

We can now conclude that, since the algorithm can go from step 3 to step 1 at most $\frac{4}{\delta^2}$ many times, we have $k \le \max_{n\le\frac{4}{\delta^2}} k_n$, hence $k = O_{\sigma,F}(1)$. □

3.1.5 Roth's theorem

We can now prove the main theorem.

Theorem 3.1.11 (Roth's theorem)**.** *For all finite groups Z and for all $f\colon Z\to\mathbb{R}$ with $0\le f(x)\le 1$ and $\mathbf{E}_Z(f)\ge\delta>0$ we have*

$$\Lambda_3(f,f,f) = \Omega_\delta(1)$$

Proof. Define $\sigma = \frac{\delta^3}{8}$ and decompose $f = f_U + f_{U^\perp}$ (we fix F later).

Using the Theorem 3.1.3 we obtain

$$\Lambda_3(f_{U^\perp}, f_{U^\perp}, f_{U^\perp}) = \Omega\left(\left(\frac{\delta}{k}\right)^k \delta^3\right)$$

Using the linearity of Λ_3 we obtain

$$\begin{aligned}
&|\Lambda_3(f,f,f) - \Lambda_3(f_{U^\perp},f_{U^\perp},f_{U^\perp})| = \\
&= |\Lambda_3(f_U+f_{U^\perp}, f_U+f_{U^\perp}, f_U+f_{U^\perp}) - \Lambda_3(f_{U^\perp},f_{U^\perp},f_{U^\perp})| = \\
&= |\Lambda_3(f_U,f_U,f_U) + \Lambda_3(f_U,f_U,f_{U^\perp}) + \cdots + \Lambda_3(f_{U^\perp},f_{U^\perp},f_U)| \le \\
&\le |\Lambda_3(f_U,f_U,f_U)| + |\Lambda_3(f_U,f_U,f_{U^\perp})| + \cdots |\Lambda_3(f_{U^\perp},f_{U^\perp},f_U)|
\end{aligned}$$

Using Lemma 1.1.1 and the fact that $0 \leq f_U, f_{U^\perp} \leq 1$ we obtain

$$|\Lambda_3(f_{U^\perp}, f_U, f_U)| \leq \|f_U\|_{L^2} \|f_{U^\perp}\|_{L^2} \|f_U\|_{u^2} \leq \|f_U\|_{u^2}$$

and similarly for the other terms. Thus

$$|\Lambda_3(f, f, f) - \Lambda_3(f_{U^\perp}, f_{U^\perp}, f_{U^\perp})| \leq 7 \|f_U\|_{u^2} \leq \frac{7}{F(\sigma, k)}$$

Then, for F sufficiently quickly growing, we have

$$\Lambda_3(f, f, f) = \left(\Lambda_3(f, f, f) - \Lambda_3(f_{U^\perp}, f_{U^\perp}, f_{U^\perp})\right) + \Lambda_3(f_{U^\perp}, f_{U^\perp}, f_{U^\perp}) \geq$$
$$\geq -\frac{7}{F(\sigma, k)} + C\left(\frac{\delta}{k}\right)^k \delta^3 = \Omega\left(\left(\frac{\delta}{k}\right)^k \delta^3\right)$$

Since $k = O_{\sigma,F}(1) = O_\delta(1)$ (F is now fixed) the claim follows.

□

3.2 Nonstandard Energy Increment

In this section we translate Roth's theorem in the nonstandard setting and outline the main steps of the nonstandard proof.

We define a trilinear form $\Lambda_3 \colon L^\infty(\mu) \times L^\infty(\mu) \times L^\infty(\mu) \to \mathbb{C}$

$$\begin{aligned}\Lambda_3(f, g, h) &= \int_{[1,N]} \int_{[-N,N]} f(n)g(n + r)h(n + 2r)\, d\mu_L(r)d\mu_L(n) = \\ &= \mathbf{E}_{n\in[N]}\, \mathbf{E}_{r\in[-N,N]}\, f(n)g(n + r)h(n + 2r)\end{aligned}$$

with $f, g, h \colon [1, N] \to \mathbb{C}$ extended to 0 outside $[1, N]$. When dealing with internal functions we can write this linear form as a hyperfinite sum.

Proposition 3.2.1. *Let $F, G, H \colon [1, N] \to {}^*\mathbb{C}$ be internal functions bounded by a finite number, then*

$$\Lambda_3(st\,(F), st\,(G), st\,(H)) = st\left(\frac{1}{N(2N + 1)} \sum_{n=1}^{N} \sum_{r=-N}^{N} F(n)G(n + r)H(n + 2r)\right)$$

Proof. Let

$$A(n) = \frac{1}{2N + 1} \sum_{r=-N}^{N} F(n)H(n + r)H(n + 2r)$$

Since F, G, H are liftings of $st(F)$, $st(G)$, $st(H)$ respectively we have

$$a(n) = \int_{[-N,N]} F(n)G(n + r)H(n + 2r)\, d\mu_L(r) = st\,(A(n))$$

Since $A(n)$ is a lifting of $a(n)$ we have

$$\begin{aligned}\Lambda_3(st\,(F), st\,(G), st\,(H)) &= \int_{[1,N]} a(n)\, d\mu_L = st\left(\frac{1}{N} \sum_{n=1}^{N} A(n)\right) = \\ &= st\left(\frac{1}{N(2N + 1)} \sum_{n=1}^{N} \sum_{r=-N}^{N} F(n)G(n + r)H(n + 2r)\right)\end{aligned}$$

□

The statement we will prove is the following

Roth's theorem. *Let $f \in L^\infty(\mu)$, $f \colon [1, N] \to \mathbb{R}$ bounded, non negative, with $\mathbf{E}_{[N]}(f) > 0$. Then $\Lambda_3(f, f, f) > 0$.*

One can prove that this statement implies the usual formulation in the same way we have done for the density increment case.

The main steps of the proof of the nonstandard version of Roth's theorem are

1. We show that $\Lambda_3(f, f, f) > 0$ for any f almost-periodic
2. We show that there exists a σ-algebra $\mathcal{Z}^1$ such that $\mathcal{Z}^1$-measurable functions are almost periodic
3. We use the conditional expectation to decompose a generic function in a $\mathcal{Z}^1$-measurable component plus an orthogonal component $f^\perp$
4. We show that $f^\perp$ does not influence the value of Λ_3

3.2.1 Almost periodicity

A *character* $e_\xi \colon [1, N] \to \mathbb{C}$ for $\xi \in {}^*\mathbb{R}$ is defined as $e_\xi(n) = st\left(e^{2\pi i \frac{\xi n}{N}}\right)$.

Definition 3.2.2 (Quasiperiodic function)**.** Let k be a positive integer. A function $f \colon [1, N] \to \mathbb{C}$ is *k-quasiperiodic* if there exist $c_1, \ldots, c_k \in \mathbb{C}$ and $\xi_1, \ldots, \xi_k \in {}^*\mathbb{R}$ such that

$$f(n) = \sum_{i=1}^{k} c_i e_{\xi_i}(n)$$

Definition 3.2.3 (Almost periodic function)**.** Let k be a positive integer and let σ be a positive real. A function $g \colon [1, N] \to \mathbb{C}$ is (k, σ)*-almost periodic* if there exists a k-quasiperiodic function f such that

$$\|f - g\|_{L^1(\mu)} \leq \sigma$$

In analogy with the proof via energy increment we have that almost periodic functions are recurrent.

Theorem 3.2.4. *Let $f \in L^\infty(\mu)$, $0 \leq f \leq 1$, $\mathbf{E}_{[1,N]}(f) = \delta > 0$. If f is (k, σ)-almost periodic with $\sigma \leq \frac{\delta^3}{8}$ then*

$$\Lambda_3(f, f, f) > 0$$

Proof. We can write $f(n) = \sum_{i=1}^{k} c_i e_{\xi_i}(n) + h(n)$ with $\|h\|_{L^1(\mu)} \leq \sigma$.

Let $T(n) = n + 1$ be the shift function. It is easily proved that μ is T-invariant.

Let $\epsilon > 0$. We prove there exists an internal set $R_\epsilon \subseteq [-\epsilon N, \epsilon N]$ with "large" internal cardinality such that

$$\|f - T^r f\|_{L^1(\mu)} \leq \epsilon + 2\sigma \quad \text{for any } r \in R_\epsilon$$

$$\begin{aligned}\|f - T^r f\|_{L^1(\mu)} &= \left\| \sum_{i=1}^{k} c_i e_{\xi_i}(n) + h(n) - \sum_{i=1}^{k} c_i e_{\xi_i}(n+r) - h(n+r) \right\|_{L^1(\mu)} \leq \\ &\leq \sum_{i=1}^{k} |c_i| \left\| e_{\xi_i}(n)(1 - e_{\xi_i}(r)) \right\|_{L^1(\mu)} + 2\,\|h\|_{L^1(\mu)} \leq \\ &\leq \sum_{i=1}^{k} |c_i| \left| 1 - e_{\xi_i}(r) \right| + 2\sigma \leq kC \max_{i=1,\dots,k} \left| 1 - e_{\xi_i}(r) \right| + 2\sigma\end{aligned}$$

where $C = \max\{|c_i|\}_{i \in \{1,\dots,k\}}$.

Using the properties of the Bohr sets and the transfer principle we obtain that there exists an internal set R^ϵ such that

$$|R^\epsilon|_I \geq \omega N \quad \text{and} \quad R^\epsilon \subseteq [-\epsilon N, \epsilon N]$$

for some ω positive real number.

Now let $r \in R^\epsilon$. Since μ is T-invariant and $0 \leq f \leq 1$ we can use the Claim 2 in Proposition 3.1.3 to obtain

$$\left\| f\, T^r f\, T^{2r} f - f^3 \right\|_{L^1(\mu)} \leq 2Ck\epsilon + 4\sigma$$

Hence

$$\begin{aligned}\mathbf{E}_{[1,N]}(f\, T^r f\, T^{2r} f) &\geq -2Ak\epsilon - 4\sigma + \mathbf{E}_{[1,N]}(f^3) \geq \\ &\geq -2Ak\epsilon - 4\sigma + \delta^3 \geq -2Ak\epsilon + \frac{\delta^3}{2} \geq \frac{\delta^3}{4}\end{aligned}$$

for ϵ sufficiently small and for any $r \in R^\epsilon$.

Since f is positive we obtain

$$\begin{aligned}\Lambda_3(f,f,f) &= \mathbf{E}_{n \in [1,N]}\, \mathbf{E}_{r \in [-N,N]} \left(f(n)\, T^r f(n)\, T^{2r} f(n) \right) \geq \frac{\delta^3}{4} \mu(R^\epsilon) = \\ &= \frac{\delta^3}{4} st \left(\frac{|R^\epsilon|_I}{N} \right) \geq \frac{\delta^3}{4} st \left(\frac{\omega N}{N} \right) = \frac{\delta^3}{4} \omega > 0\end{aligned}$$

□

3.2.2 Compact factor

Similarly to the energy increment case we define $\mathcal{Z}_N^1$ to be the σ-algebra generated by the characters $\{e_\xi \colon \xi \in [1,N]\}$. We have that $\mathcal{Z}_N^1$-measurable functions are almost periodic. More precisely the following theorem holds.

Theorem 3.2.5. *Let $f\colon [1,N] \to \mathbb{C}$ be $\mathcal{Z}^1_N$-measurable and $f \in L^\infty(\mu)$. Then for any $\sigma > 0$ there exists k such that f is (k,σ)-almost periodic*

Proof. Case A. Let $f = \mathbb{1}_A$ with $A \in \mathcal{Z}^1_N$.

Case A.1. Let $A = \{n \in [1,N]\colon e_\xi(n) \in B(x_0,R)\}$ for some $\xi \in [1,N]$ and $R \in \mathbb{R}$.

We have $f(n) = \mathbb{1}_A(n) = \mathbb{1}_{B(x_0,R)}(e_\xi(n))$. Analogously to the energy increment case we can use the Urysohn's Lemma and the Stone-Weierstrass approximation theorem to find a polynomial P which is bounded (for instance $P \le 2$) and such that

$$\|f(n) - P_\sigma(e_\xi(n))\|_{L^1} \le \sigma$$

hence f is (k,σ)-almost periodic with k degree of P_σ.

Case A.2 Let $A = A_1 \cap A_2$.

Let $P_{A_i,\frac{\sigma}{4}}$ such that $\|P_{A_i,\frac{\sigma}{4}} - \mathbb{1}_{A_i}\|_{L^1} \le \frac{\sigma}{4}$. Then

$$\begin{aligned}\|\mathbb{1}_A - P_{A_1,\frac{\sigma}{4}}P_{A_2,\frac{\sigma}{4}}\|_{L^1} &= \|\mathbb{1}_{A_1}\mathbb{1}_{A_2} - \mathbb{1}_{A_1}P_{A_2,\frac{\sigma}{4}} + \mathbb{1}_{A_1}P_{A_2,\frac{\sigma}{4}} - P_{A_1,\frac{\sigma}{4}}P_{A_2,\frac{\sigma}{4}}\|_{L^1} \le \\ &\le \|P_{A_1,\frac{\sigma}{4}} - \mathbb{1}_{A_1}\|_{L^1} + 2\|P_{A_2,\frac{\sigma}{4}} - \mathbb{1}_{A_2}\|_{L^1} \le \frac{\sigma}{4} + \frac{\sigma}{2} \le \sigma\end{aligned}$$

The thesis follows since product of trigonometric polynomials is trigonometric.

Case A.3. Let $A = B^C$.

We have

$$\|\mathbb{1}_A - (1 - P_{B,\sigma})\|_{L^1} = \|\mathbb{1}_B - P_{B,\sigma}\|_{L^1} \le \sigma$$

Case A.4. Let $A = \cap_{n\in\mathbb{N}} A_n$ where we assume $A_n \downarrow$.

Let m be such that $\|\mathbb{1}_A - \mathbb{1}_{A_m}\|_{L^1} \le \frac{\sigma}{2}$. Then

$$\|\mathbb{1}_A - P_{A_m,\frac{\sigma}{2}}\|_{L^1} \le \|\mathbb{1}_A - \mathbb{1}_{A_m}\|_{L^1} + \|\mathbb{1}_{A_m} - P_{A_m,\frac{\sigma}{2}}\|_{L^1} \le \frac{\sigma}{2} + \frac{\sigma}{2} = \sigma$$

Thus we have the thesis for any $f = \mathbb{1}_A$ with $A \in \mathcal{Z}^1_N$.

Case B. Let f be simple, $f = \sum_{i=1}^k a_i \mathbb{1}_{A_i}$.

Pick M such that $|a_i| \le M$. Then we have

$$\begin{aligned}\|f - \sum_{i=1}^k a_i P_{A_i,\frac{\sigma}{kM}}\|_{L^1} &\le \|\sum_{i=1}^k a_i(\mathbb{1}_{A_i} - P_{A_i,\frac{\sigma}{kM}})\|_{L^1} \le \\ &\le \sum_{i=1}^k |a_i|\,\|\mathbb{1}_{A_i} - P_{A_i,\frac{\sigma}{kM}}\|_{L^1} \le Mk\frac{\sigma}{Mk} = \sigma\end{aligned}$$

Case C. Let f be Loeb measurable.

We can find $f_n \to f$ with f_n simple functions. Since $f_n, f \in L^\infty(\mu)$ we have $f_n \to f$ in $L^1(\mu)$ using Lebesgue dominated convergence.

Let N be such that $\|f_N - f\|_{L^1} \leq \frac{\sigma}{2}$ and P_N be such that $\|f_N - P_N\|_{L^1} \leq \frac{\sigma}{2}$. Then

$$\|f - P_N\|_{L^1} \leq \|f - f_N\|_{L^1} + \|f_N - P_N\|_{L^1} \leq \frac{\sigma}{2} + \frac{\sigma}{2} = \sigma$$

□

3.2.3 Roth's theorem

Using the σ-algebra $\mathcal{Z}_N^1$ defined in the previous section we can decompose a Loeb measurable function f as $f = f_U + f_{U^\perp}$ with $f_{U^\perp} = \mathbf{E}(f|\mathcal{Z}_N^1)$ and $f_U = f - f_{U^\perp}$. Before proving Roth's theorem we need to show that f_U does not impact the value of Λ_3.

Lemma 3.2.6. *Let $f, g, h\colon [1, N] \to \mathbb{C}$, $f, g, h \in L^\infty(\mu)$. If $\mathbf{E}(f|\mathcal{Z}_N^1) = 0$ then $\Lambda_3(f, g, h) = 0$.*

Proof. We can take $f, g, h\colon Z \to \mathbb{C}$ with

$$Z = {}^*\mathbb{Z}/\overline{N}{}^*\mathbb{Z}$$

$\overline{N} = 3N$ and f, g, h extended to 0 outside $[1, N]$.

Without loss of generality we can assume that f, g, h are bounded by 1.

Let $F, G, H\colon Z \to {}^*\mathbb{C}$ be liftings of f, g, h respectively.

Then, similarly to the energy increment case, we obtain

$$\begin{aligned}
\Lambda_3(f, g, h) &= \mathbf{E}_{n\in Z}\,\mathbf{E}_{r\in Z}\, f(n)g(n+r)h(n+2r) = \\
&= st\left(\frac{1}{\overline{N}^2}\sum_{n=1}^{\overline{N}}\sum_{r=1}^{\overline{N}} F(n)G(n+r)H(n+2r)\right) = \\
&= st\left(\frac{1}{\overline{N}^2}\sum_{n=1}^{\overline{N}}\sum_{r=1}^{\overline{N}}\sum_{\xi\in Z}\hat{F}(\xi)e^{\xi n}\sum_{\xi\in Z}\hat{G}(\xi)e^{\xi(n+r)}\sum_{\xi\in Z}\hat{H}(\xi)e^{\xi(n+2r)}\right) = \\
&= st\left(\sum_{\xi_1,\xi_2,\xi_3\in Z}\hat{F}(\xi_1)\hat{G}(\xi_2)\hat{H}(\xi_3)\,\mathbf{E}_{n\in Z}\,\mathbf{E}_{r\in Z}\left(e^{\xi_1 n+\xi_2(n+r)+\xi_3(n+2r)}\right)\right) = \\
&= st\left(\sum_{\xi\in Z}\hat{F}(\xi)\hat{G}(-2\xi)\hat{H}(\xi)\right) \leq st\left(\|G\|_{L^2}\|H\|_{L^2}\,\|F\|_{u^2}\right) \leq st\left(\max_{\xi\in[1,N]}\left|\hat{F}(\xi)\right|\right)
\end{aligned}$$

Since $\mathbf{E}(f|\mathcal{Z}_N^1) = 0$ and F lifting of f we obtain that for every $\xi \in [1, N]$

$$0 = \mathbf{E}_{n\in Z}\left(f(n)e_\xi(n)\right) = st\left(\frac{1}{|Z|}\sum_{n\in Z}F(n)e(\xi\cdot n)\right) = st\left(\hat{F}(\xi)\right)$$

Hence $\hat{F}(\xi)$ infinitesimal and

$$|\Lambda_3(f,g,h)| \leq st\left(\max_{\xi\in[1,N]}\left|\hat{F}(\xi)\right|\right) = 0$$

□

Remark. This lemma holds if at least one among f, g, h has conditional expectation zero with respect to $\mathcal{Z}_N^1$.

We can now prove Roth's theorem.

Theorem 3.2.7 (Roth's theorem)**.** *Let $f \in L^\infty(\mu)$, $f\colon [1,N] \to \mathbb{R}$ bounded, non negative, with $\mathbf{E}_{[1,N]}(f) > 0$. Then $\Lambda_3(f,f,f) > 0$.*

Proof. Let $f = f_U + f_{U^\perp}$ with $f_{U^\perp} = \mathbf{E}(f|\mathcal{Z}_N^1)$ and $f_U = f - f_{U^\perp}$. Then

$$\begin{aligned}\Lambda_3(f,f,f) =& \Lambda_3(f_U,f_U,f_U) + \Lambda_3(f_U,f_U,f_{U^\perp}) + \\ & \cdots + \Lambda_3(f_{U^\perp},f_{U^\perp},f_U) + \Lambda_3(f_{U^\perp},f_{U^\perp},f_{U^\perp}) = \\ & = \Lambda_3(f_{U^\perp},f_{U^\perp},f_{U^\perp}) > 0\end{aligned}$$

using Lemma 3.2.6 and Theorem 3.2.4.

□

Chapter 4

Sarkozy's Theorem

In this chapter we prove Sarkozy's theorem using a density increment argument [Lya] [Gre].

Sarkozy's theorem. *Let $A \subseteq \mathbb{N}$, $\overline{d}(A) > 0$. Then A contains two elements whose difference is a perfect square.*

We will prove the equivalent nonstandard version.

Sarkozy's theorem. *Let $N \in {}^*\mathbb{N}$ be infinite and let $A \subseteq [1, N]$ be an internal subset such that its internal cardinality $|A|_I \geq \delta N$ with $\delta > 0$. Then A contains two elements whose difference is a perfect square.*

4.1 General strategy

Similarly to what we have done with Roth's theorem we will prove that if A does not contain two element whose difference is a perfect square then A has a strong correlation to a linear phase. We will then use this correlation to prove that A has higher density on a subprogression P whose spacing is a perfect square, i.e. $P = \{a + nh^2\}_{n \in {}^*\mathbb{N}}$. To do so we introduce the bilinear form

$$\Lambda_2(f, g) = \frac{1}{N^2} \sum_{n=1}^{N} \sum_{m=1}^{N} f(n) g(m) \mathbb{1}_S(m - n)$$

where $S = \{d^2 \colon 1 \leq d \leq \sqrt{N}\}$.

4.2 Weyl sums

In this section we give some estimates regarding Weyl sums [Lya].

Definition 4.2.1 (Weyl sum)**.** Let $M \in \mathbb{N}$, $\alpha \in Z_N$. We define *Weyl sum* the quantity

$$S_M(\alpha) = \sum_{m=1}^{M} e^{2\pi i\alpha\cdot m^2} = \sum_{m=1}^{M} e(\alpha\cdot m^2)$$

For k, M natural numbers we define

$$R_{2k}(M) = \left|\{(m_1, \ldots, m_k, n_1, \ldots, n_k) \in [1, M]^{2k} \colon m_1^2 + \cdots + m_k^2 = n_1^2 + \cdots + n_k^2\}\right|$$

We have the following estimate on the size of $R_{2k}(M)$.

Theorem 4.2.2. *If $k \geq 3$ then there exists a constant $c_0 > 0$ such that*

$$R_{2k}(M) \leq c_0 M^{2k-2}$$

For the proof of this theorem see [Lya].
Using the facts that

$$|S_M(n)|^2 = \sum_{1\leq m,r\leq M} e(n\cdot(m^2 - r^2))$$

and

$$\mathbf{E}_{n\in\mathbb{Z}_N}\, e(n\cdot m) = \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{otherwise} \end{cases}$$

we obtain

$$\begin{aligned} R_{2k}(M) &= \sum_{1\leq m_j,r_j\leq M} \mathbf{E}_{n\in\mathbb{Z}_N}\, e(n\cdot(m_1^2 + \cdots + m_k^2 - r_1^2 - \cdots - r_k^2)) = \\ &= \mathbf{E}_{n\in\mathbb{Z}_N} \sum_{1\leq m_j,r_j\leq M} e(n\cdot(m_1^2 - r_1^2))\cdots e(n\cdot(m_k^2 - r_k^2)) = \\ &= \mathbf{E}_{n\in\mathbb{Z}_N} \left(\sum_{1\leq m,r\leq M} e(n\cdot(m^2 - r^2))\right)^k = \mathbf{E}_{n\in\mathbb{Z}_N} |S_M(n)|^{2k} \end{aligned}$$

Let $S = \{d^2 \colon 1 \leq d \leq \sqrt{N}\}$. Since

$$\left\|\hat{\mathbb{1}}_S\right\|_{L^k}^k = \frac{1}{N} \sum_{r\in\mathbb{Z}_N} \left|\hat{\mathbb{1}}_S(r)\right|^k$$

and

$$\hat{\mathbb{1}}_S(r) = \frac{1}{N}\sum_{n\in\mathbb{Z}_N} \mathbb{1}_S(n)\overline{e(r\cdot n)} = \frac{1}{N}\sum_{1\leq 1\leq\sqrt{N}} e(-r\cdot n^2) = \frac{S_{\sqrt{N}}(-r)}{N}$$

we obtain

$$\left\|\hat{\mathbb{1}}_S\right\|_{L^{2k}}^{2k} = \frac{1}{N}\sum_{r\in\mathbb{Z}_N} \frac{S_{\sqrt{N}}(-r)^{2k}}{N^{2k}} = \frac{1}{N^{2k}}\mathbf{E}_{n\in\mathbb{Z}_N}\left|S_{\sqrt{N}}(r)\right|^{2k} = \frac{R_{2k}(\sqrt{N})}{N^{2k}}$$

Thus we have the following corollary.

Corollary 4.2.3. *If $k \geq 3$ then there exists a constant $c_0 > 0$ such that*

$$\left\|\hat{\mathbb{1}}_S\right\|_{L^{2k}} \leq \frac{c_0}{N^{\frac{k+1}{2k}}}$$

4.3 Correlation to a linear phase

We can prove the correlation to a linear phase in the same way as we have done in the proof of Roth's theorem.

Proposition 4.3.1. *Let P be an arithmetic progression with quadratic spacing, $A \subseteq P$ with $|A| \geq \delta|P|$, $\delta > 0$. If A does not contain two elements whose difference is a perfect square then there exists $\xi \in {}^*\mathbb{R}$ such that*

$$|\mathbf{E}_{n\in P}(\mathbb{1}_A(n) - \delta)e(\xi n)| = \Omega(\delta^2)$$

Proof. By rescaling we can assume $P = [1, N]$. Assume $A \subseteq \mathbb{Z}_p$ for a prime p with $N \leq p \leq 2N$ and let $S = \{d^2 \colon 1 \leq d \leq \sqrt{N}\}$. Then, by hypothesis, we have

$$\Lambda_2(\mathbb{1}_A, \mathbb{1}_A) = 0$$

We can decompose $\mathbb{1}_A = f_U + f_{U^\perp}$ with $f_{U^\perp} = \delta\mathbb{1}_{[1,N]}$, $f_U = \mathbb{1}_A - f_{U^\perp}$. Then

$$\begin{aligned}\Lambda_2(f_{U^\perp}, f_{U^\perp}) &= \frac{\delta^2}{p^2}\sum_{m,r\in\mathbb{Z}_p}\mathbb{1}_{[1,N]}(m)\mathbb{1}_{[1,N]}(r)\mathbb{1}_S(m-r) \geq \\ &\geq \frac{\delta^2}{4N}\sqrt{N} = \frac{\delta^2}{4\sqrt{N}}\end{aligned}$$

Thus, since

$$\Lambda_2(\mathbb{1}_A, \mathbb{1}_A) = \Lambda_2(f_U, f_U) + \Lambda_2(f_U, f_{U^\perp}) + \Lambda_2(f_{U^\perp}, f_U) + \Lambda_2(f_{U^\perp}, f_{U^\perp})$$

we have

$$\frac{\delta^2}{4\sqrt{N}} \le |\Lambda_2(\mathbb{1}_A, \mathbb{1}_A) - \Lambda_2(f_{U^\perp}, f_{U^\perp})| \le |\Lambda_2(f_U, f_U) + \Lambda_2(f_U, f_{U^\perp}) + \Lambda_2(f_{U^\perp}, f_U)|$$

Since

$$\begin{aligned}\Lambda_2(f, g) &= \frac{1}{p^2}\sum_{n=1}^{p}\sum_{m=1}^{p} f(n)g(m)\mathbb{1}_S(m-n) = \sum_{\xi\in\mathbb{Z}_p} \hat{f}(-\xi)\hat{g}(\xi)\hat{\mathbb{1}_S}(\xi) = \\ &= p\ \mathbf{E}_{\xi\in\mathbb{Z}_p} \hat{f}(-\xi)\hat{g}(\xi)\hat{\mathbb{1}}_S(\xi)\end{aligned}$$

we obtain, using Hölder's inequality, the Parseval identity and the estimate from Corollary 4.2.3, the inequality

$$\begin{aligned}|\Lambda_2(f_U, f_{U^\perp})| &\le 2N\, \|f_U\|_{u^2}\, \|\hat{f}_{U^\perp}\|_{L^2} \|\hat{\mathbb{1}}_S\|_{L^2} \le \\ &\le 2N\, \|f_U\|_{u^2}\, \frac{\delta^{\frac{1}{2}}}{N^{\frac{1}{2}}}\frac{c_0}{N} \le \frac{2c_0\,\|f_U\|_{u^2}}{\sqrt{N}}\end{aligned}$$

Since we can obtain the same estimate for the other two terms we obtain

$$\frac{c\delta^2}{\sqrt{N}} \le \frac{2c_0\,\|f_U\|_{u^2}}{\sqrt{N}}$$

for some $c > 0$. Hence

$$\frac{c\delta^2}{2c_0} \le \|f_U\|_{u^2} = \left|\mathbf{E}_{n\in\mathbb{Z}_p}(\mathbb{1}_A(n) - \delta)e(\xi\cdot n)\right| \le \left|\mathbf{E}_{n\in[1,N]}(\mathbb{1}_A(n) - \delta)e\left(\frac{\xi n}{p}\right)\right|$$

for some $\xi \in {}^*\mathbb{N}$.

□

4.4 Quadratic fragmentation

To obtain a partition of $[1, N]$ in subprogressions with square spacing on which e_ξ has little fluctuation we need to use the nonstandard version of the quadratic recurrence theorem.

Lemma 4.4.1 (Quadratic fragmentation)**.** *Let $N \in {}^*\mathbb{N}$ infinite, $\xi \in {}^*\mathbb{R}/{}^*\mathbb{Z}$. Then there exists $\nu \le N$ infinite and a partition*

$$[1, N] = \bigsqcup_{i=1}^{J} P_i \sqcup E$$

such that

1. *$P_j = \{a + nh^2\}_{n\in A_j}$ arithmetic progression of length at least ν*
2. *P_j, E internal and $\frac{|E|}{N}$ infinitesimal*
3. *$\left|e_\xi(x) - e_\xi(y)\right| = O(1)$ for any $x, y \in P_j$ and $j \in [1, J]$*

Proof. Let $\mu = \sqrt[6]{N}$. By quadratic recurrence (theorem 0.19) we can find $1 \le h \le \mu$ such that

$$\left\|\xi \cdot h^2\right\|_{\mathbb{R}/\mathbb{Z}} \le \frac{k}{\sqrt[10]{\mu}}$$

Let $\nu = \sqrt[10]{\mu}$. We have the partition

$$[1, N] = \bigsqcup_{i=1}^{h^2} N_i$$

with N_i subprogressions of spacing h^2. We can then partition each N_i in subprogressions of length at most ν

$$N_i = \bigsqcup_{j=1}^{J_i} P_{i,j} \sqcup E_i \quad \text{and} \quad [1, N] = \bigsqcup_{j=1}^{h} P_j \sqcup E$$

By construction $|E| \le h^2\nu \le \mu^2\nu \le \mu^3 \le \sqrt{N} = o(N)$.

We prove $\left|e_\xi(x) - e_\xi(y)\right| = O(1)$.

Let $x, y \in P_j$. Then $x = i + sh^2$, $y = i + th^2$ and

$$\begin{aligned}\left|e_\xi(x) - e_\xi(y)\right| &= \left|e_\xi(i + sh^2) - e_\xi(i + th^2)\right| = \left|e_\xi(i)(e_\xi(sh^2) - (e_\xi(th^2))\right| = \\ &= \left|e_\xi((s - t)h^2) - 1\right| \le 2\pi \left\|\xi \cdot (s - t)h^2\right\|_{\mathbb{R}/\mathbb{Z}} \le \\ &\le 2\pi\, |s - t| \left\|\xi \cdot h^2\right\|_{\mathbb{R}/\mathbb{Z}} \le 2\pi\nu\frac{k}{\nu} = O(1)\end{aligned}$$

□

We can apply this lemma to obtain the density increment on an infinite subprogression.

Proposition 4.4.2. *Let $f\colon [1, N] \to {}^*\mathbb{R}$ such that*

$$st\left(\mathbf{E}_{n\in N} \left|f(n)e(\xi \cdot n)\right|\right) \ge \sigma > 0$$

Then there exists $P \subseteq [1, N]$ arithmetic progression of infinite length and square spacing such that

$$st\left(\mathbf{E}_{n\in P}\, f(n)\right) = st\left(\frac{1}{|P|}\sum_{n\in P} f(n)\right) \ge \frac{\sigma}{4}$$

The proof is the same of Proposition 1.2.4 except that this one uses the Quadratic fragmentation Lemma 4.4.1 instead of the Linear fragmentation Lemma to obtain the subprogression.

4.5 Sarkozy's theorem

We can now prove Sarkozy's theorem.

Sarkozy's theorem. *Let $N \in {}^*\mathbb{N}$ infinite and let $A \subseteq [1, N]$ internal subset such that $|A|_I = \delta N$ with $\delta > 0$. Then A contains two elements whose difference is a perfect square.*

Proof. Assume ad absurdum that A does not contain two elements whose difference is a perfect square. By using Proposition 4.3.1 we can find a phase ξ such that

$$\left|\mathbf{E}_{n\in[1,N]}(\mathbb{1}_A(n) - \delta)e(\xi n)\right| \geq c\delta^2$$

Using Proposition 4.4.2 we can find a subprogression with spacing h^2, $h \geq 1$ and infinite length ν such that

$$\frac{1}{|P|}\sum_{n\in P}(\mathbb{1}_A(n) - \delta\mathbb{1}_{[1,N]}(n)) \geq \frac{c\delta^2}{4}$$

hence

$$\frac{1}{|P|}\sum_{n\in P}\mathbb{1}_A(n) \geq \delta + \frac{c\delta^2}{4}$$

Let $P = \{a + nh^2\}_{n\in[1,\nu]}$, then we define

$$B = \frac{(A \cap P) - a}{h^2}$$

We prove B does not contain two elements whose difference is a perfect square. Let $x, y \in B$ such that $x - y = r^2$, then $\overline{x} = h^2x + a$ and $\overline{y} = h^2y + a$ are in A and $\overline{x} - \overline{y} = h^2r^2 = (rh)^2 = 0$ since A does not contain two elements whose difference is a perfect square. Since $h \leq 1$ we have $r = 0$ and thus $x = y$.

We can then repeat this process enough times until we obtain an absurdum since the density of B increases by at least $k\delta^2$ each step and the density cannot exceed one. $\square$

Conclusions

We now give some final remarks and indicate some possible developments.

The proof of Roth's theorem in the nonstandard setting using the *density increment* approach is easy to obtain and reduces the length of computations but does not give any estimate on $r_3(N)$.

The proof of Roth's theorem in the nonstandard setting using the *energy increment* approach is more interesting since it provides an easy way to obtain the decomposition. On the other hand, the introduction of the Loeb measure prevents one from using nonstandard tools in several steps of the proof. One could obtain a "pure" nonstandard proof by providing a subspace of $\mathbb{C}^N$ (whose dimension is hyperfinite) which contains all the almost periodic functions. For instance, one could consider the subspace generated by the eigenvectors in $\mathbb{C}^N$ with respect to the shift function and adapt the process used in the standard case by replacing the characters e_ξ with said eigenvectors.

To prove Sarkozy's theorem we have used the nonstandard version of the density increment argument, however it is possible to prove it with the standard density increment in a similar way. The standard proof requires slightly longer computations, but at least it gives an estimate (albeit poor).

Using an approximation of polynomial Weyl sums and a theorem of polynomial recurrence, one should be able to extend both proofs to patterns of the form $x, x + P(n)$ with $P(n)$ polynomial such that $P(0) = 0$.

Bibliography

[CLR] Ernie Croot, Neil Lyall, and Alex Rice. A purely combinatorial approach to simultaneous polynomial recurrence modulo 1. *Proc. Amer. Math. Soc. 143 (2015), no. 8, 3231–3238.*

[Cut01] Nigel J. Cutland. *Loeb Measures in Practice: Recent Advances*. Springer, 2001.

[Fur77] Harry Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *Journal d'Analyse Mathématique*, 1977.

[Gol98] Robert Goldblatt. *Lectures on the Hyperreals*. Springer, 1998.

[Gre] Ben Green. On arithmetic structures in dense sets of integers. *Duke Math. Jour. 114 (2002), no. 2, 215-238.*

[Lya] Neil Lyall. The Weil inequality and Sarkozy's theorem. Lecture notes.

[McC99] Randall McCutcheon. *Elemental Methods in Ergodic Ramsey Theory*. Springer, 1999.

[Pet89] Karl Petersen. *Ergodic Theory*. Cambridge University Press, 1989.

[Sch77] Wolfgang M. Schmidt. *Small Fractional Parts of Polynomials*. American Mathematical Society, 1977.

[Tao09] Terence Tao. *Poincaré's Legacies, Part I: pages from year two of a mathematical blog*. American Mathematical Society, 2009.

[TV06] Terence Tao and Van H. Vu. *Additive Combinatorics*. Cambridge University Press, 2006.

[Zha11] Yufei Zhao. Szemerédi's theorem via ergodic theory, 2011. Unpublished notes.