The Function Theory and Structural Information

Mingcheng Sheng

Advisor: Amir Mohammadi

UCSD

Honor Thesis

June 6, 2023

ACKNOWLEDGMENT

From explaining the complicated theory to editing words in this honor thesis, my advisor Amir Mohammadi has guided me meticulously throughout the whole project. He lead me into the world of Mathematics and I am gratitude to connect to him as my advisor.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Background

Let A be any set and define the sum set as $A+A := \{x+y : x,y \in A\}$, what structural information about A can we deduce from the behavior of the sum set? For example, if $A \subset \mathbb{Z}$ and $|A| = m$, then the Freiman's Theorem states, roughly, that when $|A+A|$ is not much larger than A, A must effectively contained in a generalised arithmetic progression. The exact statement is as follows.

1.1.1 Theorem. *Let C be a constant. There exist constants d_0 and K depending only on C such that whenever A is a subset of $\mathbb{Z}$ with $|A| = m$ and $|A+A| \leq Cm$, there exist $d \leq d_0$, an integer x_0 and positive integers $x_1, \cdots, x_d$ and $k_1, \cdots, k_d$ such that $k_1 k_2 \cdots k_d \leq Km$ and*

$$A \subset \{x_0 + \sum_{i=1}^{d} a_i x_i : 0 \leq a_i < k_i (i = 1,2,...,d)\}$$

The same is true if $|A - A| \leq Cm$.

This beautiful result can be generalised in many directions. For instance, let $A+n := \{x+n, x \in A\}$ we can ask what structure A must possess if each $A+in \cap A+jn$ is not a small set for most pair (i,j)). As we will see, this will be a crucial step of proving the following Szemerédi's Theorem.

1.1.2 Theorem. *Let k be a positive integer and let $\delta > 0$. There exists a positive integer $N = N(k,\delta)$ such that every subset of the set $\{1,2,...,N\}$ of size at least δN contains an arithmetic progression of length k.*

Another possibility is to consider A as a subset of a (possibly non-abelian) group G. We say that A has bounded doubling if $|A \cdot A| \leq K|A|$ for some $K = O(1)$, where

$A \cdot A = \{ab : a, b \in A\}$ and $|A|$ is the cardinality of A. Then we can try to classify what a set with bounded doubling property looks like. One result of this type is the famous Gromov Theorem, which states as follows.

1.1.3 Theorem. *Every finitely generated group of polynomial growth is virtually nilpotent.*

The goal of this thesis is to present proofs of 1.1.2 and 1.1.3.

1.2 Ideas to prove the Szemerédi's Theorem

We start with Theorem 1.1.2, the first observation is that a random subset $A \subset \{1, 2, ..., N\}$ of cardinality δN contains an arithmetic progression of length k almost surely for all k. There are many ways to define what "random" means mathematically. For example we may assume the set is uniformly generated, let random variable X be the number of k-term arithmetic progressions in A, then

$$\mathbb{E}(X) = \sum_{P} \delta^k$$

where P is any k-term arithmetic progression in $\{1, 2, ..., N\}$. Therefore $\mathbb{E}(X) \to \infty$, so $\mathbb{P}(X > 0) = 1$ almost surely.

The randomness condition is more like an independence relation among different sets. For example, given a probability measure μ on $\{1, 2, ..., N\}$, then for a random generated set we should expect set A and $A + n$ are independent sets, and therefore

$$\mu(A \cap (A + n)) = \mu(A)\mu(A + n).$$

On the other hand, if A is not randomly generated, then those sets must share some correlations from which we may extract arithmetic progressions. Again, there are different ways to make correlations mathematically precise. But to define it properly requires deep tools from other fields because this is where we need to build the structure on a random chosen set.

1.3 The Ergodic Theoretic Approach

Let us say we have a reasonable measure μ on $\mathbb{N}$ and $\mu(A) > 0$, to find k-term arithmetic progressions in A it is natural to look at the set $A_{n_k} = A \cap A + n \cap \cdots \cap A + (k-1)n$. If for some n, $\mu(A_{n_k}) > 0$ then $\exists a \in \mathbb{N}$ s.t $a \in A_{n_k} \neq \emptyset$ and $a, a+n, \cdots, a+(k-1)n$ will be our progression. As we have discussed, if A is randomly generated then each $A+in$ and $A+jn$ should be independent, therefore

$$\mu(A_{n_k}) = \mu(A)\mu(A+n)\cdots\mu(A+(k-1)n) = \mu(A)^k > 0$$

We may expect that $A+in$ and $A+jn$ should have some correlations if they are not independent, but it is possible that $\frac{1}{N}\sum_{n=0}^{N-1}\mu(A\cap A+n) \to \mu(A)^2$. This weaker independence is the notion of randomness we will use. In ergodic theory this notion corresponds to the *weak-mixing system.* Let T denote the shift operator. With tools from ergodic theory (which is probably some generalisation of the law of large numbers), we can show that

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

as desired.

Now we can investigate what if T is not weak-mixing. With some technical work we can assume $\mu(A\cap T^n A)$ converge to zero in density, see 2.1.1. Therefore

$$\frac{1}{N}\sum_{n=0}^{N-1} |\mu(A\cap T^n A) - \mu(A)\mu(T^n A)|^2 \to \delta > 0$$

It is usually better to think of the functional space defined on $\mathbb{N}$ rather than only subsets of it because there are more tools to study the former. For example, in L^2 space we can generalize the notion of inner product in Euclidean space which allow us to discuss different modes convergence and L^2 decomposition. The idea of weakly convergence, which means that $\langle f_n, g\rangle \to \langle f, g\rangle$, is not available if we only look at sets. For this reason we restate the above condition as:

For each function f with positive mean regarding the measure μ, and for all function $g \in L^2$,

$$\frac{1}{N}\sum_{n=0}^{N-1} \left|\langle T^n f, g\rangle\right| - \int f \int g \to \delta > 0 \tag{1.3.1}$$

and we want to prove that

$$\liminf \frac{1}{N}\int_X f \times T^n f \times \cdots \times T^{(k-1)n} f d\mu > 0$$

The exact definitions can be found in chapter 2.

Now the ergodic theory comes in. Equation 1.3.1 will imply that there are many pairs of $\{m, n\}$ s.t $\langle T^m f, T^n f\rangle > 0$ (see lemma 2.1.3). The main result is that there is always an nontrivial almost periodic function g s.t $\langle f, g\rangle > 0$ (see theroem 3.0.1 and

3.0.2). Here almost periodic means that for any $\epsilon > 0$ $\{n : \|T^n f - f\|_{L^2} < \epsilon\}$ has bounded gap. The proof is based on the following idea from ergodic theory: If T is measure preserving($\int Tf d\mu = \int f d\mu$ for all f), for any function f the ergodic average

$$g_N := \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^n f$$

will be more and more T-invariant because

$$\lim_{N\to\infty} \left| Tg_N - g_N \right| = \lim_{N\to\infty} \frac{1}{N} \left| T^N f - f \right| \to 0$$

as $T^n f - f$ is bounded (see theorem 2.2.2). This will help us to find the almost-periodic function g.

The final step returns to a combinatorial argument that loosely speaking extract arithmetic progressions from almost periodic functions.

Finally, we have not specify the measure we are discussing, because the existence of certain measure is already an nontrivial question. First of all, how to actually define a measure on $\mathbb{N}$? We also need the measure μ to be invariant under the shift operator T,(meaning $\mu(A) = \mu(T^{-1}A)$) because the whole idea of ergodic theory is based on measure preserving systems. We will construct such a measure at the beginning of chapter 3.

1.4 The Fourier Analytic Approach

There are also Fourier analytic and the Combinatorial approach to Theorem 1.1.2, however, these appracohes will not be discussed in great details in this thesis. These are certainly interesting in their own right and they provide more quantitative results, which is missing from the Ergodic theory approach. It is worth mentioning that Gowers uses a combination of Fourier analysis and additive combinatorics (the Freiman problem) to give a significant improvement of 1.1.2, which is the best known bound to date.

1.4.1 Theorem (Gowers)**.** *For every positive integer k there is a constant $c = c(k) > 0$ such that every subset of $\{1, 2, ..., N\}$ of size at least $N(loglogN)^{-c}$ contains an arithmetic progression of length k. Moreover, c can be taken to be $2^{-2^{k+9}}$.*

Here we will discuss the special case when $k = 3$, in this case the Fourier transform provide a direct way to compute the number of length 3 arithmetic progressions in a set.

We consider A as subset of $\mathbb{Z}_N$ and write ω for the number $exp(2\pi i/N)$. Recall that the discrete Fourier transform for a function $f : \mathbb{Z}_N \to \mathbb{C}$ is

$$\hat{f}(r) = \sum_{s\in\mathbb{Z}_N} f(s)\omega^{-rs}$$

Let 1_A be the indicator function, then the number of length 3 arithmetic progression in A is just $\sum_{n_1+n_3=2n_2} 1_A(r)1_A(r)1_A(-2r)$, which can be conveniently written in Fourier coefficients because

$$\begin{aligned}
&\sum_{r\in\mathbb{Z}_N} \hat{1}_A(r)\hat{1}_A(r)\hat{1}_A(-2r) \\
=&\sum_{r\in\mathbb{Z}_N}\Big(\sum_{n_1\in\mathbb{Z}_N} 1_A(n_3)\omega^{-rn_1}\Big)\Big(\sum_{n_1\in\mathbb{Z}_N} 1_A(n_3)\omega^{-rn_3}\Big)\Big(\sum_{n_2\in\mathbb{Z}_N} 1_A(n_2)\omega^{2rn_2}\Big) \\
=&\frac{1}{N^3}\sum_{r\in\mathbb{Z}_N}\sum_{n_1,n_2,n_3\in\mathbb{Z}_N} 1_A(n_1)1_A(n_2)1_A(n_3)\omega^{n_1 r}\omega^{n_2(-2r)}\omega^{n_3 r} \\
=&\frac{1}{N^2}\sum_{n_1,n_2,n_3\in\mathbb{Z}_N} 1_A(n_1)1_A(n_2)1_A(n_3)\sum_{r\in\mathbb{Z}_N}\omega^{r(n_1+n_3-2n_2)} \\
=&\frac{1}{N^2}\sum_{n_1+n_3=2n_2} 1_A(n_1)1_A(n_2)1_A(n_3)
\end{aligned} \tag{1.4.1}$$

Observe that

$$\begin{aligned}
\sum_{r\in\mathbb{Z}_N} \hat{1}_A(r)\hat{1}_A(r)\hat{1}_A(-2r) =&\hat{1}_A(0)^3 + \sum_{r\neq 0}\hat{1}_A(r)\hat{1}_A(r)\hat{1}_A(-2r) \\
=&d(A)^3 + \sum_{r\neq 0}\hat{1}_A(r)\hat{1}_A(r)\hat{1}_A(-2r)
\end{aligned}$$

where $d(A)$ is the density of the set A. If $\hat{1}_A(r)$ is very small for all r, then

$$\sum_{r\in\mathbb{Z}_N} \hat{1}_A(r)\hat{1}_A(r)\hat{1}_A(-2r) \approx d(A)^3 > 0$$

which gives positive number of 3-term arithmetic progressions. If there is a r s.t $\hat{1}_A(r)$ is reasonably large, say

$$\left|\hat{1}_A(r)\right| = \left|\sum_{s\in\mathbb{Z}_N} 1_A(s)\omega^{-rs}\right| > \delta > 0$$

Because of the cycling nature trigonometric function, we use a pigeonhole argument to show that $\mathbb{Z}/N\mathbb{Z}$ can be split into the union of arithmetic progressions $P_1 \cup P_2 \cdots$ s.t each P_j has almost same length and ω^{-rs} are almost constant in each P_j. Then

$$\delta < \left|\sum_{s\in\mathbb{Z}_N} 1_A(s)\omega^{-rs}\right| \le \sum_i \left|\sum_{s\in P_i} 1_A(s)\omega^{-rs}\right| \approx \sum_i \left|\sum_{s\in P_i} 1_A(s)\right|$$

we can then show that for some $P_j \in \{1, 2, \cdots, N\}$ the density of set A in P_j is larger than in $\{1, 2, \cdots, N\}$. Apply this argument repeatedly to get $P_j, P_j', \cdots$ s.t the density of

A in P_j^m keeps increasing and eventually we will obtain our 3-term arithmetic progression in one of the P_j^m.

When $k > 3$, there is no formula to count the number of k-term arithmetic progression as equation 1.4.1. We will also encounter many other practical difficulties to apply the density increment argument. For more details, see [5]

1.5 Ideas to prove the Gromov Theorem

Now we move to Theorem 1.1.3. Here too, one hopes to get an insight to structural properties of a group using functions defined on it. Let us recall the following classical result:

1.5.1 Theorem (Choquet, Deny). *Every bounded harmonic function on an abelian group is constant.*

The idea of studying behaviors Harmonic functions largely come from differential Geometry. In [6], Tobias H. Colding and William P. Minicozzi proved the following theorem.

1.5.2 Theorem. *For an open manifold with nonnegative Ricci curvature the space of harmonic functions with polynomial growth of a fixed rate is finite dimensional.*

There is also a classical result claiming that given (M,g) a non-compact and connected complete Riemannian manifold of dimension $n \geq 3$. If the Ricci curvature is non-negative, then the Bishop—Gromov volume comparison tells that the volume growth is at most Euclidean,

$$\lim_{r\to\infty} \frac{Vol(B(o,r))}{r^n} = \rho \neq \infty$$

Although there is no obvious analogous concept of curvature in group theory, we can study the analogous harmonic functions on groups in hope to obtain structural information. The method of Kleiner considers the linear space of Lipschitz harmonic functions on the finite generated group G. For a finite generated sets S, f is harmonic if

$$f(g) = \frac{1}{|S|} \sum_{s\in S} f(gs)$$

for all $g \in G$, and Lipschitz means

$$\left| f(g) - f(gs) \right| < C$$

for all $g \in G$, $s \in S$, and $C < \infty$. He proves that when the group is of polynomial growth the linear space must be non-trivial and finite dimensional. Since the group acts on the harmonic functions by left translation, the finite dimensional space will give a representation of the group, which we can use to derive the nilpotent structure we want.

CHAPTER 2

Functional Theory

2.1 L^2 space, Mode of convergence, and Conditional Expectation

Definition 2.1.1 (Mode of Convergence)**.** Let $v_0, v_1, \dots$ be a sequence in a normed vector space V . Let $v \in V$.

1. (Usual convergence in norm) We say that $lim_{n\to\infty} v_n = v$ if $lim_{n\to\infty} \parallel v_n - v \parallel = 0$.

2. (Convergence in density) We say that v_n converges to v in density, denoted $D-lim_{n\to\infty} v_n = v$, if for any $\epsilon > 0$, the set $\{n \in \mathbb{N} : \parallel v_n - v \parallel \geq \epsilon\}$ has upper density zero.

3. (Cesàro convergence) We say that v_n converges to v in a Cesàro sense, denoted $C - lim_{n\to\infty} v_n = v$, if $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} v_n = v$.

4. (Cesàro supremum) Define $C - sup_{n\to\infty} v_n = \limsup_{n\to\infty} \parallel \frac{1}{N}\sum_{n=0}^{N-1} v_n \parallel$

Here we collect some standard result that connect different mode of convergence.

2.1.2 Proposition. *Let $v_0, v_1, v_2, \cdots$ be a bounded sequence of vectors in a normed vector space V , and let $v \in V$. Then the following are equivalent:*

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

The following lemma says that if v_n does not converge in the Cesàro sense, then vectors within the sequence must share some correlations. In fact, this is the only analytical tool we need to prove 1.1.2.

2.1.3 Lemma (Van der Corput)**.** *Let $v_0, v_1, v_2, \dots$ be a bounded sequence of vectors in a real Hilbert space. If*

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

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

Proof. When $0 < h < H$ we have

$$\frac{1}{N}\sum_{n=0}^{N-1} v_n = \frac{1}{N}\sum_{n=0}^{N-1} v_{n+h} + O(\frac{H}{N})$$

Averaing over h we obtain

$$\frac{1}{N}\sum_{n=0}^{N-1} v_n = \frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{H}\sum_{h=0}^{H-1} v_{n+h} + O(\frac{H}{N})$$

By Cauchy-Schwarz,

$$\begin{aligned}\left\|\frac{1}{N}\sum_{n=0}^{N-1} v_n\right\|^2 &\leq \frac{1}{N}\sum_{n=0}^{N-1}\left\|\frac{1}{H}\sum_{h=0}^{H-1} v_{n+h}\right\|^2 + O(\frac{H}{N}) \\ &= \frac{2}{H}\sum_{h=0}^{H-1}(1-\frac{h}{H})\frac{1}{N}\sum_{n=0}^{N-1}\langle v_n, v_{n+h}\rangle + O(\frac{H}{N})\end{aligned}$$

Therefore

$$\left\|\text{C-sup}_{n\to\infty} v_n\right\|^2 \leq \frac{2}{H}\sum_{h=0}^{H-1}(1-\frac{h}{H})\,\text{C-sup}_{n\to\infty}\langle v_n, v_{n+h}\rangle \leq \frac{2}{H}\sum_{h=0}^{H-1}\text{C-sup}_{n\to\infty}\langle v_n, v_{n+h}\rangle$$

letting $H \to \infty$ gives the result. □

Definition 2.1.4. (Conditional Expectation) For a probability space (X, χ, μ), and a measurable function $f : X \to R$. Let χ' be a sub-σ-algebra of X. Then $L^2(X, \chi', \mu) \subset L^2(X, \chi, \mu)$. The conditional expectation is the orthogonal projection

$$\mathbb{E}(f|\chi') : L^2(X, \chi, \mu) \to L^2(X, \chi', \mu).$$

So in particular, $\mathbb{E}(f|\chi')$ is the unique function in $L^2(X, \chi, \mu)$ such that

$$\langle f, g\rangle = \langle \mathbb{E}(f|\chi'), g\rangle$$

for all $g \in L^2(X, \chi', \mu)$,

Example 1. (Projections from $\mathbb{R}^2$ to $\mathbb{R}$)

Consider a function $f : \mathbb{R}^2 \to R$ with a σ-algebra $X \times Y$, then

$$\mathbb{E}(f|X)(x) = \int_Y f(x, y)dy$$

because for any function $g \in L^2(X)$,

$$\langle f, g\rangle = \int_{X\times Y} fgdydx = \int_X g\left(\int_Y fdy\right)dx = \langle \mathbb{E}(f|X), g\rangle$$

By definition, if $g \in L^2(X, \chi', \mu)$, then for all $f \in L^2(X, \chi, \mu)$

$$\mathbb{E}(fg|\chi') = \mathbb{E}(f|\chi')g$$

This means that we can always take the χ'-measurable component out when computing the projection. In general,

$$\mathbb{E}(cf + dg|Y) = c\mathbb{E}(f|Y) + d\mathbb{E}(g|Y), \forall f, g \in L^2(X|Y), c, d \in L^\infty(Y).$$

To better manipulate the functional space of conditional expectation we can define an inner product by

$$\langle f, g \rangle_{X|Y} := \mathbb{E}(fg|Y)$$

With these tools we can define a hilbert module by treating $L^\infty(Y)$ as constants.

Definition 2.1.5. (Hilbert module) Let $Y = (Y, \gamma, \nu, S)$ be a factor of (X, χ, μ, T).Define the Hilbert module $L^2(X, \chi, \mu|Y, \gamma, \nu)$ (abbreviate as $L^2(X|Y)$) over the commutative von Neumann algebra $L^\infty(Y, \gamma, \nu)$ to be the space of all $f \in L^2(X)$ such that the conditional norm

$$\|f\|_{L^2(X|Y)} := \mathbb{E}(|f|^2|Y)$$

lies in $L^\infty(Y)$

We also have the conditional version of Cauchy-Schwarz inequality.

2.1.6 Lemma. *Let $X \to Y$ be an extension. Then for any $f, g \in L^2(X|Y)$ we have*

$$\left| \langle f, g \rangle_{L(X|Y)} \right| \leq \|f\|_{L^2(X|Y)} \|g\|_{L^2(X|Y)}$$

almost everywhere.

2.2 Measure Preserving Systems

A measure-preserving dynamical system is a system (X, χ, μ, T) with the following structure:

- X is a set.
- χ is a sigma algebra on X.
- $\mu : \chi \to [0, 1]$ is a probability measure so $\mu(X) = 1$ and $\mu(\emptyset) = 0$.

- $\forall A \in \chi, \mu(T^{-1}(A)) = \mu(A)$

The first result of measure preserving system is the following Poincaré Recurrence Principle, which is an infinite version of the Pigeonhole principle. (All proofs of theorems in this and next section can be found in [1])

2.2.1 Theorem. *(Poincaré Recurrence) If T is measure-preserving, and a set $E \in X$ with $\mu(E) \geq 0$, then $T^n(x)$ for almost all $x \in X$ should return infinitely often to E. More precisely, there exists a measurable set $F \subseteq E$ with $\mu(F) = \mu(E)$ with the property that for every $x \in F$ there exist integers $0 < n_1 < n_2 < \dots$ with $T_{n_i}x \in E$ for all $i \geq 1$.*

2.2.2 Theorem. *Von Neumann's mean ergodic theorem(holds in Hilbert spaces)*

Let U be a unitary operator on a Hilbert space H. (U^* is the adjoint of U that satisfies $\langle Ux, y\rangle = \langle x, U^*y\rangle$, U is unitary if $UU^* = I$) Let P be the orthogonal projection onto the invariant subspace $\{g \in H \mid Ug = g\} = ker(I - U)$. Then, $\forall x \in H$, we have:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U^n x = Px$$

where the limit is with respect to the norm on H. In other words, the sequence of averages

$$\frac{1}{N} \sum_{n=0}^{N-1} U^n$$

converges to P in the strong operator topology.

The sum $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U^n x$ is called the ergodic average of the system. Heuristically, the ergodic average becomes more and more invariant because

$$U(\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U^n x) - \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U^n x = \frac{1}{N}(U^n x - x) \to 0$$

This allows us to use the ergodic average to approximate typical behaviors in this system.

2.3 Ergodic, Weak Mixing, and Mixing Systems

Definition 2.3.1. (Ergodicity) A measure-preserving system (X, χ, μ, T) is ergodic if for any $B \in \chi$, $T^{-1}B = B$ implies $\mu(B) = 0$ or $\mu(B) = 1$.

For ergodic systems we can improve the theorem 2.2.2 to the following:

2.3.2 Theorem. *(Birkhoff) Let (X,B,μ,T) be a measure-preserving system. If $f \in L^1(\mu)$, then*

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

converges almost everywhere to a T-invariant function $f^*(x) \in L^1(\mu)$ and

$$\int f \mathrm{d}\mu = \int f^* \mathrm{d}\mu.$$

If T is ergodic, then

$$f = \int f^* \mathrm{d}\mu.$$

almost everywhere.

Example 2. (Bernoulli shift) Consider the the set of possible outcomes of the infinitely repeated toss of a fair coin. The outcome can be described by an infinite sequence $(x_0, x_1, \cdots)$, which lives in the product space $X = \{0,1\}^{\mathbb{N}}$. Denote μ the usual product measure on X. The left shift map $\sigma : X \to X$ defined by

$$\sigma(x_0, x_1, \cdots) = (x_1, x_2, \cdots)$$

preserves μ and is ergodic.

Example 3. The circle rotation $R_\alpha : T \to T$ is ergodic with respect to the Lebesgue measure m_T if and only if α is irrational.

The two examples still differs fundamentally even if both are ergodic. For the second example, we can predict the behavior of the system after knowing enough values of it. But no matter how many values we know about the first system, the next value is still completely arbitrary. This suggests that we need a stronger notion to distinguish between these two examples.

Definition 2.3.3. (Mixing) A measure-preserving system (X,χ,μ,T) is mixing if for any two sets $A, B \in \chi$,

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

For rotations, take A, B as two small intervals and we can always find some large n s.t $\mu(A \cap T^{-n}B) = 0$. Therefore rotations are not mixing. Think the Bernoulli shift as coin flips and A, B as events of infinite tosses that fix finite many values, then A is independent of $T^{-n}B$ when n is large. Thus $\mu(A\cap T^{-n}B)=\mu(A)\mu(B)$ which means the Bernoulli shift is mixing.

Weak mixing system is an intermediate concept between ergodic system and mixing system.

Definition 2.3.4 (Weak mixing system). A measure preserving system (X,χ,μ,T) is weak mixing if

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

for any two sets $A,B\in X$, or equivalently,

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

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

Example 4. (Interval exchange transformation) Let $n>0$ and let π be a permutation on $\{1,\cdots,n\}$. Consider a vector $\lambda=(\lambda_1,\dots,\lambda_n)$ of positive real numbers (the widths of the subintervals), satisfying

$$\sum_{i=1}^{n}\lambda_i=1$$

Let

$$\beta_i=\sum_{1\leq j<i}\lambda_j$$

and

$$\lambda^{\phi}=(\lambda_{\pi^{-1}1},\cdots,\lambda_{\pi^{-1}m})$$

Finally define $T=T_(\lambda,\pi)$, the (λ,π) interval exchange by the formula $T_(\lambda,\pi)x=x-\beta_{i-1}(\lambda)+\beta_{\pi i-1}(\lambda^{\pi})$, then T is weak mixing but not mixing.

Definition 2.3.5. (Weak mixing functions) Let (X,χ,μ,T) be a measure preserving system. A function $f\in L^2(X)$ is weak mixing if D-$\lim_{n\to\infty}\langle T^nf,f\rangle=0$.

It can be shown that the following definition is equivalent to 2.3.4.

Definition 2.3.6. A measure preserving system (X,χ,μ,T) is weak mixing if all functions $f\in L^2(X)$ are weak mixing.

In fact, using the Van der Corput lemma one can show that weak-mixing function mix with any functions.

2.3.7 Theorem. *Let (X,χ,μ,T) be a measure preserving system, and let $f\in L^2(X)$ be weak mixing. Then for any $g\in L^2(X)$ we have*

$$\text{D-}\lim_{n\to\infty}\langle T^nf,g\rangle=0 \tag{2.3.1}$$

Proof. Since $\langle T^n f, g\rangle$ is bounded for all n, say by C, and f is weak-mixing, we have

$$\text{C-lim}_{n\to\infty}\,\text{C-sup}_{n\to\infty}\langle T^n f, g\rangle\langle T^{n+h} f, g\rangle\langle T^n f, T^{n+h} f\rangle \le C^2\cdot\text{C-lim}_{h\to\infty}\left|\langle f, T^n f\rangle\right| = 0. \tag{2.3.2}$$

Equation 2.3.1 is equivalent to the following

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\|\langle T^n f, g\rangle\|^2 = \lim_{N\to\infty}\frac{1}{N}\left\langle\sum_{n=0}^{N-1}\langle T^n f, g\rangle T^n f, g\right\rangle \to 0 \tag{2.3.3}$$

By Cauchy-Schwarz, it suffices to show that $\text{C-lim}_{n\to\infty}\langle T^n f, g\rangle T^n f = 0$. This can be shown by applying the Van der Corput lemma to equation 2.3.2. This completes the proof. □

There are similar results for ergodic systems and mixing systems. In short, for a measure preserving system (X, χ, μ, T) and any function $f, g \in L^2(X)$, the system is

- Mixing if
$$\langle T^n f, g\rangle \to \int f \int g \tag{2.3.4}$$
- Weak Mixing if
$$\frac{1}{N}\sum_{n=0}^{N-1}\left|\langle T^n f, g\rangle\right| \to \int f \int g \tag{2.3.5}$$
- Ergodic if
$$\frac{1}{N}\left|\sum_{n=0}^{N-1}\langle T^n f, g\rangle\right| \to \int f \int g \tag{2.3.6}$$

2.4 Compact Systems

What if a function is not weak-mixing? In other words, what if there exists a $\delta > 0$ s.t

$$\text{D-lim}_{n\to\infty}\langle T^n f, g\rangle \ge \delta \tag{2.4.1}$$

Again by Van der Corput lemma, we have for some constant c_δ,

$$\text{C-sup}_{h\to\infty}\,\text{C-sup}_{n\to\infty}\langle v_n, v_{n+h}\rangle \ge c_\delta \tag{2.4.2}$$

This may make the following definition more natural.

Definition 2.4.1. (Almost periodic functions) Let (X, χ, μ, T) be a system. We say that $f \in L^2(X)$ is almost periodic if its orbit $\{T^n f : n \in Z\}$ is precompact in $L^2(X)$ in the norm topology. Equivalently, f is almost periodic if for every $\epsilon > 0$, the set$\{n \in Z :\| f - T^n f \|\le \epsilon\}$ has bounded gap.

Similar to weak mixing system we can define

Definition 2.4.2. (compact system) The system (X, χ, μ, T) is compact if every $f \in L^2(X, \chi, \mu)$ is almost periodic.

In a general measure preserving system, most functions are neither weak-mixing nor almost periodic, as illustrated by the following example.

Example 5. Consider the skew torus given by $(\mathbb{R}/\mathbb{Z})^2$ with the shift map $T(y,z) = (y+a, z+y)$. So $T^n(y,z) = (y+na, z+ny+\binom{n}{2}a)$. Consider the function $f(y,z) = e^{2\pi iz}$, which will be send to

$$T^n f(y,z) = e^{-2\pi i\binom{n}{2}a} e^{-2\pi iny} f$$

Since when $n \neq m$,

$$\begin{aligned}\langle T^n f(y,z), T^m f(y,z)\rangle &= \int_{\mathbb{T}^2} e^{-2\pi i\binom{n}{2}a} e^{-2\pi iny} f \times \overline{e^{-2\pi i\binom{m}{2}a} e^{-2\pi imy} f} dydz \\ &= \int_{\mathbb{T}} e^{-2\pi i(n-m)y} dy \times \int_{\mathbb{T}} e^{g(n,m,z)} dz \\ &= 0\end{aligned}$$

$T^n f$ and $T^m f$ are always orthogonal and therefore $f(y,z)$ is not almost periodic. It is worth mentioning that $f(y,z)$ is actually a weak-mixing function, which suggests that the subsystem derived by fixing y-coordinate in the skew torus is weak-mixing.

Now we consider another function $g(y,z) = y$, so $T^n g(y,z) = y + na$ and it is easy to show that this function is almost-periodic.

Now take $h(y,z) = f(y,z) + g(y,z)$ then this function will be neither weak-mixing nor almost periodic. Observe that most functions are of this type.

A good example of a compact system is again the circle rotation, which is also an example of a Kronecker system.

Definition 2.4.3. An (abelian) Kronecker system is a measure preserving system of the form (G, B, μ, T_a), where $(G, +)$ is a compact abelian metrisable group, B is its Borel algebra, μ is its Haar measure, $a \in G$, and $T_a x = x + a$.

In general, a compact system needs not to be ergodic, but when it is the system must be equivalent to a Kronecker system.

2.4.4 Theorem. *(Halmos and von Neumann)*

Every ergodic compact system is equivalent to a Kronecker system.

2.5 Weak Mixing and Compact Extensions

As mentioned, if we view f as a function of z for fixed y, then it is almost periodic. To make this precise we need the theory of conditional expectation and extensions of dynamic systems.

Definition 2.5.1. (Factors and extensions) Let (X, χ, μ, T) and $Y = (Y, \gamma, \nu, S)$ be measure preserving systems. An extension (also called a factor map)$\phi : X \to Y$ is a measure preserving map (i.e. if $A \in Y$, then $\phi^{-1}(A) \in X$ and $\mu(\phi^{-1}(A)) = \nu(A)$) that is shift-compatible, i.e., $\phi \circ T = S \circ \phi$.

Definition 2.5.2. (Conditionally weak mixing function). Let $X \to Y$ be an extension of measure preserving systems. A function $f \in L^2(X|Y)$ is conditionally weak mixing relative to Y if D-$\lim_{n\to\infty} \langle T^n f, f\rangle_{L^2(X|Y)} = 0$ in $L^2(Y)$.

Definition 2.5.3. (Conditionally almost periodic function)

A subset E of $L^2(X|Y)$ is said to be conditionally precompact if for every $\epsilon > 0$,we can find $f_1, \ldots, f_d \in L^2(X|Y)$ so that $\forall f \in E, \min_{1\le i\le d} \|(f - f_i)\|_{L^2(X|Y)}(y) \le \epsilon$ a.e. for $y \in Y$.

A function $f \in L^2(X|Y)$ is conditionally almost periodic if its orbit $\{T^n f : n \in \mathbb{Z}\}$ is conditionally precompact in $L^2(X|Y)$, and it is conditionally almost periodic in measure if for every $\epsilon > 0$ there exists a set E in Y with $\mu(E) > 1-\epsilon$ such that f_{1_E} is conditionally almost periodic.

Definition 2.5.4. An extension $X \to Y$ of measure preserving systems is weak mixing if every $f \in L^2(X|Y)$with conditional mean zero (i.e., $E(f|Y) = 0$ a.e.) is conditionally weak mixing. An extension $X \to Y$ is said to be compact if every function in $L^2(X|Y)$ is conditionally almost periodic in measure.

Back to example 5, consider the skew torus $X = ((\mathbb{R}/\mathbb{Z})^2, \gamma\times\gamma, \mu\times\mu, (y, z) \to (y+a, z+y))$ as an extension of $Y = (\mathbb{R}/\mathbb{Z}, \gamma, \mu, y \to y + a)$. Again consider functions $f(y, z) = e^{2\pi i z}$ and

$$T^n f(y, z) = e^{-2\pi i \binom{n}{2} a} e^{-2\pi i n y} f$$

which lies in the zonotope $\{cf : c \in L^\infty(Y), \|c\|_{L^\infty(Y)} \le 1\}$ and hence f is conditionally almost periodic. Now approximate any function $f(y, z)$ on the torus by its vertical Fourier expansions,(i.e. perform the Fourier expansion $f(y, z)$ for fixed y). For each $y \in Y$ take the partial sum of the vertical Fourier expansions, and we obtain a sequence of functions $\{f\}_n$

that is pointwise convergent to f. By the Egorov's theorem, $\{f\}_n$ converge uniformly to f on the torus except on a set of measure arbitrarily small. This shows that f is conditionally almost periodic in measure and the extension $X \to Y$ is compact.

CHAPTER 3

The Proof of the Szemerédi Theorem

In order to use tools from the Ergodic theory, we first need to define a measure preserving action on $\mathbb{N}$. This is constructed in the following way.

Let $X_0 = \{0,1\}^{\mathbb{Z}}$ be the full shift on two symbols with shift map $T : A \to A+1$.Define a point x_A in X_0 by

$$x_A(n) = \begin{cases} 1 & \text{if } n \in A \\ 0 & \text{if } n \notin A \end{cases} \tag{3.0.1}$$

Now let X denote the smallest closed subset of X_0 that is invariant under T and contains the point x_A. Let E denote the cylinder set $\{x \in X | x(0) = 1\}$, which is both closed and open (clopen) in X. Then

$$T^n(x_A) \in E \Longleftrightarrow T^n(x_A)_0 = 1 \Longleftrightarrow n \in E.$$

Let μ_N denote the measure on X given by

$$\mu_N = \frac{1}{2N+1} \sum_{-N}^{N} \delta_{T^n(x)}$$

where δ_b is the point mass at $b \in X$. Then $\mu_N(E)$ is the density of A in $[-N,N]$ and therefore $\mu_N(E) > 0$. By Banach-Alaoglu theorem μ_{N_j} has a weak limit μ and $\mu(E) > 0$. μ is T-invariant because

$$T\mu_n - \mu_n = \frac{1}{2N+1}\left(\delta_{T^n(x)} - \delta_{T^{-n}(x)}\right) \longrightarrow 0$$

Our problem is now transformed to prove that in every measure preserving system (X,χ,μ,T), for every positive integer k and any $E \in X$ with $\mu(E) > 0$, there exists an n s.t

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

If this is true, then because E is clopen, the intersection will contain at least a point $T^{-m}(x_A)$. Therefore the arithmetic progression $(m+in)_{0 \le i \le k-1}$ is contained in A.

For our purpose we will prove the stronger result

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

or equivalently,

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

whenever $f \in L^\infty(X), f > 0$ and $\mathbb{E}(f) > 0$.

It's not hard to show that this result holds for weak-mixing functions and almost periodic functions. So our first step is to show that all functions can more or less reduce to these two types.

3.0.1 Proposition. *(Decomposition Theorem) Let X be a measure preserving system. Then we have*

$$L^2(X) = WM(X) \bigoplus AP(X)$$

as an orthogonal direct sum of Hilbert spaces.

First we prove that if $f \in WM(X)$ and $g \in AP(X)$, then $\langle f, g \rangle = 0$. This direction is relatively easy by 2.3.7.

Then we prove that if f is not weak-mixing, then there is a function $g \in AP(X)$ such that $\langle f, g \rangle \neq 0$. As discussed under 2.3.7, the equation 2.4.2 is our starting point. But we still need to develop a robust process to find the hidden almost periodic functions. Notice that the naive attempt by taking the ergodic average $g = \lim \frac{1}{N} \sum_{n=0}^{N-1} T^n f$ will not work because that limit goes to zero if T is ergodic.

We will going to use the Hilbert-Schmidt operator to find those almost periodic functions. The idea is to first find an operator Φ that sends the bounded image $\{n : T^n f\}$ to a precompact set. If Φ further commutes with T, then $\{n : T^n \Phi f\} = \Phi\{n : T^n f\}$ will be precompact, so Φf will be almost periodic.

Proof. We now turn to the details of the argument. Let us recall that if H is a Hilbert space and $(e_i)_{i\in I}$ is an orthonormal basis for H, a Hilbert-Schmidt operator on H is an operator T for which $\sum_{i\in I} ||Te_i||^2 < \infty$.

We will use the following property of Hilbert-Schmidt operators: If $\phi : H \to H'$ is a Hilbert-Schmidt operator between two Hilbert spaces, then a bounded set in H will be sent to a totally bounded set in H'.

Now we consider an adjusted ergodic average

$$g = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle T^n f, f \rangle T^n f.$$

The limit will be nonzero if f is not weak mixing since

$$\langle g, f\rangle = \lim_{N\to\infty} \sum_{n=0}^{N-1} |\langle T^n f, f\rangle|^2 = \text{C-lim}_{n\to\infty} |T^n f, f|^2 \neq 0$$

If we denote $\Phi_f : \Phi_f g = \langle f, g\rangle f$, then it is a Hilbert-Schmidt operator because

$$\sum_i ||\Phi_f e_i||^2 \ = \sum_i |\langle f, e_i\rangle|^2 \parallel f \parallel^2 = \|f\|^4 < \infty$$

Furthermore, by the Von Neumann's mean ergodic theorem we know that the limit g converge to a function invariant under the operator $U : \Phi_f \to \Phi_{Tf}$, so g commutes with T. □

In view of Proposition 3.0.1, we can always decompose a function $f \in L^2(X)$ as $f = f_{WX} + f_{AP}$, but the established structure is still insufficient to prove the multi-recurrence principle simply because we have no idea how to estimate cross terms like $\int_X f_{WX1} f_{AP1} f_{WX2} \ldots$. This is why we need to define conditionally weak-mixing functions 2.5.2 and almost periodic functions 2.5.3.

With the conditional Cauchy-Schwarz inequality 2.1.6 we have the following analog of the Decomposition Theorem 3.0.1.

3.0.2 Proposition. *As $L^\infty(X)$-modules, we have*

$$L^2(X|Y) = WM(X|Y) \bigoplus AP(X|Y) \tag{3.0.2}$$

So we can always extract a compact factor out from any function. Keep on applying this proposition we will eventually obtain the following tower like structure for any measure-preserving system.

3.0.3 Theorem. *(Furstenberg-Zimmer). Let (X, χ, μ, T) be a measrue preserving system, then there exists a sequence of measure preserving subsystems Y_i of X such that*

$$X \to Y_\alpha \to \cdots \to Y_1 \to Y_0$$

is a chain of extensions where Y_0 is the trivial system. Furthermore, $Y_{i+1} \to Yi$ is compact and $X \to Y_\alpha$ is weak mixing. (α may be infinite).

Now it only remains to show that SZ property lifts through both weak-mixing and compact extensions.

3.0.4 Theorem. *If $(X, \chi, \mu, T) \to (Y, \gamma, \nu, S)$ is a compact extension and Y is SZ, then X is SZ.*

We want to show that $\forall f \in L^\infty(X)$ with $\mathbb{E} > 0$ we have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f T^n f \dots T^{(k-1)n} f d\mu > 0 \tag{3.0.3}$$

Actually we will prove a stronger result states as follows.

3.0.5 Proposition. *Let $X \to Y$ be a weak mixing extension. Let $k \geq 1$. Let $a_1, \cdots, a_k \in \mathbb{Z}$ be distinct non-zero integers, and $f_1, \cdots, f_k \in L^\infty(X)$. Then*

$$\text{C-lim}_{n\to\infty} \left(\int_X T^{a_1 n} f_1 \cdots T^{a_k n} f_k d\mu - \int_Y S^{a_1 n} \mathbb{E}(f_1|Y) \cdots S^{a_k n} \mathbb{E}(f_k|Y) d\nu \right) = 0$$

in $L^2(X)$.

With this proposition it is easy to deduce that X is SZ from Y is SZ,

Proof. First we have

$$\int_X T^{a_1 n} f_1 \cdots T^{a_k n} f_k d\mu = \int_Y \mathbb{E}(T^{a_1 n} f_1 \cdots T^{a_k n} f_k | Y)(y) d\nu$$

WLOG we can assume that $\mathbb{E}(f_i|Y) = 0$ and it is sufficient to show that

$$\int_Y \mathbb{E}(T^{a_1 n} f_1 \cdots T^{a_k n} f_k | Y)(y) d\nu = 0$$

We proceed by using induction on k. When $k = 1$, the Mean Ergodic Theorem (Theorem 2.2.2) gives

$$\text{C-lim}_{n\to\infty} (T^{a_1 n} f_1 | Y) = \mathbb{E}(f_1|Y)$$

Since $X \to Y$ is a weak-mixing extension, for all $i = 1, 2, \cdots, k$

$$\text{D-lim}_{n\to\infty} \langle T^n f_i, f_i \rangle_{L^2 X|Y} = 0$$

We now would like to apply the van der Corput lemma. Denote $v_n = T^{a_1 n} f_1 \cdots T^{a_k n} f_k$, then in $L(X|Y)$, we have

$$\begin{aligned}
\langle v_n, v_{n+h} \rangle_{L^2(X|Y)} &= \langle T^{a_1 n} f_1 \cdots T^{a_k n} f_k, T^{a_1 (n+h)} f_1 \cdots T^{a_k (n+h)} f_k \rangle_{L(X|Y)} \\
&= \mathbb{E}\left(\prod_{i=1}^k (T^{a_i n} f_i T^{a_i (n+h)} f_i) | Y \right) \\
&= \mathbb{E}\left(\prod_{i=1}^k T^{a_i n} (f_i T^{a_i h} f_i) | Y \right)
\end{aligned}$$

So,

$$\begin{aligned}
&\text{C-sup}_{h\to\infty} \text{C-sup}_{n\to\infty} \langle v_n, v_{n+h} \rangle_{L^2(X|Y)} \\
&\qquad = \text{C-sup}_{h\to\infty} \text{C-sup}_{n\to\infty} \mathbb{E}\left(\prod_{i=1}^k T^{a_i n} (f_i T^{a_i h} f_i) | Y \right)
\end{aligned}$$

By the conditional Cauchy-Schwarz inequality (Lemma 2.1.6),

$$\mathbb{E}\bigg(\prod_{i=1}^{k} T^{a_in}(f_iT^{a_ih}f_i)|Y\bigg) \leq \mathbb{E}\bigg(\prod_{i=1}^{k-1}(T^{a_in}(f_iT^{a_ih}f_i))^2|Y\bigg)^{1/2}\mathbb{E}\bigg((T^{a_kn}(f_kT^{a_kh}f_k))^2|Y\bigg)^{1/2}$$

Since the last term is bounded and $\mathbb{E}(f_i) \geq 0$, it is suffice to show that

$$\text{C-sup}_{h\to\infty}\,\text{C-sup}_{n\to\infty}\,\mathbb{E}\bigg(\prod_{i=1}^{k-1} T^{a_in}(f_iT^{a_ih}f_i)|Y\bigg) = 0$$

By induction, taking $f_iT^{a_ih}f_i$ as the new function, we have

$$\text{C-lim}_{n\to\infty}\bigg(\prod_{i=1}^{k-1} T^{a_in}(f_iT^{a_ih}f_i)|Y\bigg) = \prod_{i=1}^{k-1}\mathbb{E}(f_iT^{a_ih}f_i)$$

As each $\mathbb{E}(f_iT^{a_ih}f_i)$ is bounded, say by M, we have

$$\text{C-sup}_{h\to\infty}\prod_{i=1}^{k-1}\mathbb{E}(f_iT^{a_ih}f_i) \leq M^{k-2}\,\text{C-sup}_{h\to\infty}\mathbb{E}(f_1T^{a_1h}f_1|Y)$$

but $\mathbb{E}(f_1T^{a_1h}f_1|Y)_{L^2(X|Y)} = \langle T^{a_1h}f_1, f_1\rangle_{L^2(X|Y)}$ converge to zero in density, so the above equation converge to zero. $\square$

3.0.6 Theorem. *If $(X,\chi,\mu,T) \to (Y,\gamma,\nu,S)$ is a compact extension and Y is SZ, then f is SZ in X.*

Let us elaborate on the statement of the theorem. We have that Y is SZ, so for any set $A \subset Y$ with $\nu(A) > 0$ and all large K,

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\nu\Big(A\cap T^n(A)\cap\cdots\cap T^{(K-1)n}(A)\Big) > 0; \tag{3.0.4}$$

moreover, $\forall f \in L^\infty(X)$ nonnegative with $\mathbb{E}(f) > 0$, we can find $f_1,\ldots,f_d \in L^2(X|Y)$ so that $\forall n \in \mathbb{N}$,

$$\min_{1\leq i\leq d}\|(T^nf - f_i)\|_{L^2(X|Y)}(y) \leq \epsilon \tag{3.0.5}$$

a.e. for $y \in Y$. Our goal is to use this and prove

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int_X fT^nf\ldots T^{(k-1)n}fd\mu > 0 \tag{3.0.6}$$

for all $k \geq 1$.

But equation 3.0.5 is not easy to use. Ideally we want it to hold for a fixed function f_j and the set of n will form an arithmetic progression. This is where we need the following results from combinatorics.

3.0.7 Theorem (Van der Waerden's theorem)**.** *If the natural numbers are written as a disjoint union of finitely many sets,*

$$\mathbb{N} = C_1 \sqcup C_2 \sqcup \cdots \sqcup C_r$$

then there must be one set C_j that contains arbitrarily long arithmetic progressions.

With Van der Waerden's theorem at hand, for each $y \in Y$ we can find an arithmetic progression $P_y = \{a_y, a_y + r_y \cdots a_y + (k-1)r_y\}$ so that

$$\|T^{a_y + ir_y} f - f_j\|_{L(X|Y)}(y) < \epsilon \tag{3.0.7}$$

Our strategy is now as follows:

$$\mathbb{E}(T^{a_y} f T^{a_y + r_y} f \ldots T^{a_y + (k-1)r_y} f | Y)(y) \approx \mathbb{E}(T^{a_y} f^k | Y)(y), \tag{3.0.8}$$

Hence we may hope to conclude that

$$\begin{aligned} \liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f T^n f \ldots T^{(k-1)n} f d\mu &= \liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_Y \mathbb{E}(f T^n f \ldots T^{(k-1)n} f | Y)(y) d\nu \\ &\approx \liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_Y \mathbb{E}(f^k | Y) d\nu > 0 \end{aligned} \tag{3.0.9}$$

However, there is still a major issue in this argument because the arithmetic progression P_y we use in equation 3.0.9 depends on y. To make the above argument work, we need to find a fixed arithmetic progression. To this end, we have to choose our progression more carefully in the first place and then apply the pigeonhole principle.

The argument is explicated below. Before proceeding to the formal proof of Theorem 3.0.6, however, we present a sketch of the proof of Theorem 3.0.7. Even though the proof can be found in the literature, we present the sketch here both for the convenience of the reader and also to highlight the role of combinatorics even in this ergodic theoretic approach.

Sketch of the proof. Let $W(l,k)$ denote the smallest number N s.t any k-colouring of the segment of positive integers $[1,N]$ contains a monochromatic l-term arithmetic progression. So if we consider a very large interval $[1,MN]$, then any coloring will give many l-term arithmetic progressions. Those progressions may form a $l+1$-term arithmetic progression if we choose our M properly. To make this precise, we say that r different l-term arithmetic progressions $A_1, A_2, \cdots, A_r$, where

$$A_i = \{a_i + jd_i : j \in [0, l-1]\}, \quad i \in [1, r],$$

is color-focused on a positive number m if

1) Each A_i is entirely in one color, and A_i and A_j have different colors if $i \neq j$.

2) $a_1 + ld_1 = a_2 + ld_2 = \cdots = a_r + ld_r = m$

The proof is based on a double induction. Note that, for any positive integer k, $W(1,k) = 1$ and $W(2,k) = k+1$, so we may induct on l. We also have that $W(l,1)$ is trivial, but to lift k to any number needs a proper inductive hypothesis.

3.0.8 Lemma. *Fix a $k \geq 2$, for all $r \leq k$ there is an M_r such that any k-colouring of $[1,M]$ contains a monochromatic l-term arithmetic progression or r colour-focused $(l-1)$-term arithmetic progressions together with their focus.*

Sketch of the proof of the lemma. When $r = 1$ simply set $M = 2W(l-1,k)$. When $r \geq 2$ consider $[1, M_{r-1}W(l-1, k^{M_{r-1}})]$. Suppose that c is a k-colouring of this interval that does not contain a monochromatic l-term arithmetic progression. The key argument is: if we split that interval into consecutive blocks B_i with length M_{r-1}, since there are only $k^{M_{r-1}}$ ways to color a block, the coloring c will induce a $k^{M_{r-1}}$ coloring of $[1, W(l-1, k^{M_{r-1}})]$. By the induction hypothesis, The induced colouring contains a monochromatic $(l-1)$-term arithmetic progression. This implies that there are $l-1$ blocks $B_{i_j}, 1 \leq j \leq l-1$, that are identically coloured by c and that are equally spaced between each other. From this it is not hard to find r colour-focused $(l-1)$-term arithmetic progressions and hence complete the induction. This concludes the discussion.

The proof of the theorem can be completed using the lemma. □

Proof of Theorem 3.0.6. Fix $k \geq 1$, and let $f \in L^\infty(X)$ be a nonnegative function with positive mean. We want to show that

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f T^n f \dots T^{(k-1)n} f \, d\mu > 0 \tag{3.0.10}$$

First, we can normalize f so that $\|f\|_{L^\infty} = 1$. Choose $\delta > 0$ and consider the subset

$$A = \{y \in Y : E(f|Y)(y) > \delta\}$$

of Y with $\nu(A) > 0$. Then since the system Y is SZ, we know that for all large $K \in \mathbb{N}$,

$$\nu(A_{n_K}) := \nu(A \cap T^n(A) \cap \cdots \cap T^{(K-1)n}(A)) > 0 \tag{3.0.11}$$

for all $n \in \Xi$ where $\Xi \subset \mathbb{N}$ has positive lower density. The large number K will be chosen later to help us find the fixed arithmetic progression.

For each n with the above property, we want to find a set satisfying (3.0.8). Fix one such n. The extension $X \to Y$ is compact, so there are finite number of functions $f_1, f_2, \dots, f_d$ so that for all $0 \le a \le K$, we will have

$$\min_{1 \le j \le d} ||T^{an} f - f_j||_{L^2(X|Y)}(y) < \epsilon, \qquad \nu\text{-a.e. } y.$$

Now choose K large enough so that for each y, by the Van der Waerden's theorem there exists a length k arithmetic progression $a_y, a_y + r_y \cdots a_y + (k-1)r_y \subset [0, K]$ and some $1 \le j \le d$ so that

$$\|T^{(a_y + ir_y)n} f - f_j\|_{L^2(X|Y)}(y) < \epsilon \tag{3.0.12}$$

The arithmetic progression we obtain depends on y, but since $a_y, r_y \le K$ there must be a set

$$B_n \subset A \cap T^n(A) \cap \cdots \cap T^{(K-1)n}(A)$$

with $\nu(B_n) \ge \frac{\nu(A_{nK})}{K} > 0$ so that a_y and r_y are defined and are constant on B_n.

Now we have $\|T^{(a+ir)n} f - f_j\|_{L^2(X|Y)}(y) < \epsilon$ for $i = 0, 1 \cdots, k-1$. By the triangle inequality,

$$\|T^{(a+ir)n} f - T^{an} f\|_{L^2(X|Y)}(y) < 2\epsilon, \quad \forall y \in B_n, i = 0, 1, \cdots, k-1$$

Since $\|f\|_{L^\infty} = 1$,

$$\|(T^{an} f T^{(a+r)n} f \dots T^{(a+(k-1)r)n} f - (T^{an} f)^k\|_{L^2(X|Y)}(y) = O_k(\epsilon)$$

So,

$$\mathbb{E}(T^{an} f T^{(a+r)n} f \dots T^{(a+(k-1)r)n} f | Y)(y) \ge \mathbb{E}((T^{an} f)^k | Y)(y) - O_k(\epsilon)$$

But we know that $\mathbb{E}(f|Y)(y) > \delta$ when $y \in B_n$, since T is measure preserving we obtain

$$\begin{aligned}
\int_X f T^{rn} f \dots T^{(k-1)rn} f d\mu &= \int_X T^{an} f T^{(a+r)n} f \dots T^{(a+(k-1)r)n} f d\mu \\
&= \int_X \mathbb{E}(T^{an} f T^{(a+r)n} f \dots T^{(a+(k-1)r)n} f | Y)(y) dy \\
&\ge \int_{B_n} \mathbb{E}(T^{an} f T^{(a+r)n} f \dots T^{(a+(k-1)r)n} f | Y)(y) dy \\
&\ge \int_{B_n} [\mathbb{E}((T^{an} f)^k | Y)(y) - O_k(\epsilon)] dy \\
&\ge \nu(B_n)(\delta^k - O_k(\epsilon)) > 0
\end{aligned}$$

The last inequality holds if we choose ϵ small enough. Moreover, the above inequality will hold for all n that 3.0.11 holds. Therefore, 3.0.10 holds. $\square$

CHAPTER 4

Kleiner's proof of the Gromov's Theorem on groups with polynomial growth

Definition 4.0.1. Let G be a finitely generated group, and let $B_G(r) \subset G$ be a ball centered at the identity with respect to a word norm on G. The group G has polynomial growth if for some fixed $d > 0$

$$\limsup_{r\to\infty} \frac{|B_G(r)|}{r^d} < \infty$$

4.0.2 Theorem (Kleiner)**.** *Let G be a group of polynomial growth generated by a finite symmetric set S of generators. Then the vector space V of Lipschitz harmonic functions is finite-dimensional.*

The polynomial growth implies that bounded doubling happens on most scales, or

$$|B_S(2R)| \leq C|B_S(R)|$$

for a fixed C and for most $R > 0$. For simplicity we will assume that bounded doubling is true for all $R > 0$. The full proof can be found in Kleiner's paper [7] with some additional pigeonhole argument.

The reason that harmonic functions reflect the growth condition of the group lies in the following lemma.

4.0.3 Lemma. *Assume bounded doubling for all $R > 0$. Let $\epsilon > 0$ be a small parameter, cover $B_S(4R)$ by balls $\{B_i\}$ of radius ϵR. Suppose that a harmonic function $f : G \to \mathbb{R}$ has mean zero on every such ball. Then one has*

$$\|f\|_{\ell^2(B_S(R))} \ll \epsilon \|f\|_{\ell^2(B_S(4R))} \tag{4.0.1}$$

This lemma shows that harmonic functions defined on a larger ball $B_S(4R)$ can not vary too much on smaller balls $B_S(R)$, which in turn implies the lack of different harmonic functions. The proof of the lemma requires two inequalities which we will state later. For now let's see how the lemma implies Kleiner's theorem.

Lemma 4.0.3 implies Theorem 4.0.2. Consider Lipschitz harmonic functions $\{u_1, \dots, u_n\}$, and let $V = span\{u_1, \cdots, u_n\}$. Since we only care of dimensions we can assume $\{u_1, \dots, u_n\}$ forms an orthogonal basis of V and all functions vanish at identity. Define the quadratic form Q_R on V as

$$Q_R(u_i, u_j) = \sum_{B_S(R)} u_i u_j$$

Our goal is to prove that

$$\lim_{n\to\infty} \det(Q_R(u_i, u_j))_{1\le i,j\le n} \longrightarrow 0$$

which shows that V is finite dimensional as $n \to \infty$.

The Lipschitz condition controls the growth rate of each harmonic function, which gives

$$\det(Q_R(u_i, u_j))_{1\le i,j\le n} = \prod_{i=1}^{n} Q_R(u_i, u_i) \ll R^{2n} \tag{4.0.2}$$

By bounded doubling, we can cover $B_S(4R)$ by $O_\epsilon(1)$ balls of radius ϵR. Now V splits into two types of functions: (1) $u_i \in V$ has mean zero on all balls of radius ϵR, which are bounded by lemma 4.0.3. We denote this subspace W. (2) $u_i \in V$ has nonzero mean on some balls of radius ϵR. These functions will have dimension $O_\epsilon(1)$ in space V. Therefore we obatin

$$\begin{aligned}
\det(Q_R(u_i, u_j))_{1\le i,j\le n} &= \prod_{u_i\in W} Q_R(u_i, u_i) \prod_{u_i\in W^\perp} Q_R(u_i, u_i) \\
&\le \prod_{u_i\in W} O(\epsilon) Q_{4R}(u_i, u_i) \prod_{u_i\in W^\perp} Q_{4R}(u_i, u_i) \qquad (4.0.3)\\
&= O(\epsilon)^{n-O_\epsilon(1)} \det(Q_{4R}(u_i, u_j))_{1\le i,j\le n}
\end{aligned}$$

We can choose small ϵ and as $n \to \infty$, inequality 4.0.2 and 4.0.3 will force

$$\det(Q_R(u_i, u_j))_{1\le i,j\le n} \longrightarrow 0.$$

So we are done. $\square$

Definition 4.0.4. Let $u : G \to \mathbb{R}$ be a function. The gradient $\nabla u : G \to \mathbb{R}^S$ of u is defined by the formula

$$\nabla u(x) := (u(xs) - u(x))_{s\in S}$$

So,

$$|\nabla u(x)| := (\sum_{s\in S} |u(xs) - u(x)|^2)^{1/2}$$

The *Laplacian* of a function $u : G \to \mathbb{R}$ is defined by the formula

$$\Delta u := -\nabla \cdot \nabla u$$

or more explicitly

$$\Delta u(x) = 2|S|u(x) - 2\sum_{s\in S} u(xs).$$

Now we can state the two inequalities required to prove Lemma 4.0.3, the proofs are taken from [3].

4.0.5 Lemma (Poincaré inequality)**.** *Let $f : G \to \mathbb{R}$, $x \in G$, and $r \geq 1$. Let*

$$f_{B(x,r)} := \frac{1}{|B(x,r)|} \int_{B(x,r)} f$$

be the average value of f on $B(x,r)$. Then

$$\|f - f_{B(x,r)}\|_{\ell^2(B(x,r))} \ll r^2 \frac{\left|B_S(2r)\right|}{\left|B_S(r)\right|} \|\nabla f(x)\|_{\ell^2(B(x,3r))} \tag{4.0.4}$$

Proof. By definition of the gradient, we have the pointwise bound

$$|f(ygs) - f(yg)| \leq |\nabla f(yg)| \tag{4.0.5}$$

for all $y, g \in G$ and $s \in S$. Take $g \in B_S(2r)$ and average this in ℓ^2 over all $y \in B(x,r)$, we conclude that

$$\sum_{y\in B(x,r)} |f(ygs) - f(yg)|^2)^{1/2} \leq \|\nabla f\|_{\ell^2(B(x,3r))}. \tag{4.0.6}$$

Telescoping this using the triangle inequality, we conclude that

$$\sum_{y\in B(x,r)} |f(ygs) - f(yg)|^2)^{1/2} \leq 2r\|\nabla f\|_{\ell^2(B(x,3r))} \tag{4.0.7}$$

for all $g \in B_S(2r)$. Summing in g using the triangle inequality, we conclude that

$$(\sum_{y\in B(x,r)} (\sum_{y\in B_S(2r)} |f(ygs) - f(yg)|^2)^{1/2} \leq 2r|B_S(2r)|\|\nabla f\|_{\ell^2(B(x,3r))} \tag{4.0.8}$$

But for any $y \in B(x,r)$, we have

$$|f(y) - f_{B(x,r)}(y)| \leq \frac{1}{|B_S(r)|} \sum_{z\in B(x,r)} |f(z) - f(y)| \leq \frac{1}{|B_S(r)|} \sum_{g\in B_S(2r)} |f(yg) - f(y)|$$

and the claim follows. □

4.0.6 Lemma (Reverse Poincaré inequality)**.**

$$\sum_{y\in B_S(x,R)} \left|\nabla f(y)\right|^2 \ll R^{-2} \sum_{y\in B_S(x,2R)} \left|f(y)\right|^2 \tag{4.0.9}$$

where

$$\left|\nabla f(x)\right|^2 := \sum_S \left|f(x)-f(xs)\right|^2$$

Proof. Let $\psi : G \to R$ be the cutoff function $\psi(y) := \max(1-\frac{dist(x,y)}{2R}, 0)$. Then

$$f(ys)-f(y)\cdot\psi^2(y) = f\cdot\psi^2(ys) - f\cdot\psi^2(y) + f(ys)\cdot(\psi^2(y)-\psi^2(ys)) \tag{4.0.10}$$

We have $f(ys) = O(|f(y)|+|\nabla f(y)|)$, and

$$\psi(ys)-\psi(y) = O(\frac{1}{R}).$$

Multiplying 4.0.10 by $f(ys)-f(y)$ and summing in s, we conclude that

$$|\nabla f|^2\psi^2(y) = \nabla(f\cdot\psi^2)\cdot\nabla f + O(|S|(|f(y)|+|\nabla f(y)|)(\frac{\psi(y)}{R}+\frac{1}{R^2})).$$

Summing by parts we have

$$\begin{aligned}\sum_{y\in G}|\nabla f|^2\psi^2(y) \ll \sum_{y\in G}|f\psi^2(y)||\nabla f(y)| + \frac{|S|}{R}|(\nabla f(y)|+|\nabla f(y)|)\psi(y)\\ +\frac{|S|}{R^2}\sum_{y\in B(x,2R-1)}|\nabla f(y)|(|f(y)|+|\nabla f(y)|).\end{aligned} \tag{4.0.11}$$

The first term is zero since f is harmonic. Applying the Cauchy-Schwarz to the second term we obtain

$$\frac{|S|}{R}|(\nabla f(y)|+|\nabla f(y)|)\psi(y) \le t\sum_{y\in G}|\nabla f|^2\psi^2(y) + \frac{1}{t}\frac{|S|^2}{R^2}\sum_{y\in G}(\nabla f(y)|+|\nabla f(y)|)^2$$

for all $t>0$. Insert this into equation 4.0.11 we have

$$\begin{aligned}(1-t)\sum_{y\in G}|\nabla f|^2\psi^2(y) &\ll \frac{C_{t,|S|}}{R^2}\sum_{y\in G}(\nabla f(y)|+|\nabla f(y)|)^2\\ &\le \frac{C_{t,|S|}}{R^2}(\|f\|_{\ell^2(B(x,2R-1))}+\|\nabla f\|_{\ell^2(B(x,2R-1))})\end{aligned} \tag{4.0.12}$$

Choosing small t we get

$$\|\nabla f\|^2_{\ell^2(B(x,R))} \le \|\nabla f\|_{\ell^2(B(x,2R-1))} \ll \frac{1}{R^2}\|f\|_{\ell^2(B(x,2R-1))}$$

and the lemma follows. $\square$

To prove lemma 4.0.3, observe that (f has mean zero on B_i)

$$\|f\|_{\ell^2(B_S(R))} \leq \sum_i \sum_{y \in B_i} \left(f(y) - f_{B_i}\right)^2$$

Applying lemma 4.0.5, we obtain

$$\|f\|_{\ell^2(B_S(R))} \leq \sum_i R^2 \frac{\left|B_S(2\epsilon R)\right|}{\left|B_S(\epsilon R)\right|} \|\nabla f(x)\|_{\ell^2(3B_i)}$$

Using bounded doubling we can refine the family of balls $\{B_i = B(x_i, \epsilon R)\}$ so that the triples $\{3B_i = B(x_i, 3\epsilon R)\}$ have bounded overlap. This means that for growth rate d we can find constant C_d s.t the intersection multiplicity of $\{3B_i\}$ is less than C_d. Since $3B_i \in B(4R)$, we have

$$\sum_i \|\nabla f(x)\|_{\ell^2(3B_i)} \leq C_d \|\nabla f(x)\|_{\ell^2(B_S(4R))}$$

Applying lemma 4.0.6 for each ball and summing we obtain

$$\begin{aligned} \|f\|_{\ell^2(B_S(R))} &\leq C_d \frac{\left|B_S(2\epsilon R)\right|}{\left|B_S(\epsilon R)\right|} \sum_i (\epsilon R)^2 \|\nabla f(x)\|_{\ell^2(B_S(4R))} \\ &\ll \epsilon \|f(x)\|_{\ell^2(B_S(8R))} \end{aligned} \tag{4.0.13}$$

which completes the proof.

We briefly discuss the remaining proof of the Gromov Theorem, for more details see T. Tao's blog [4]. We will first construct a nontrivial Lipschitz harmonic functions, since group G acts on the space of harmonic functions, we obtain a finite dimensional representation $G \to GL_n$. We can then use this representation to do induction on the growth rate R^d to prove that G is virtually nilpotent for all d.

BIBLIOGRAPHY

[1] Manfred Einsiedler and Thomas Ward, *Ergodic Theory with a view towards Number Theory*. 11

[2] Yufei Zhao (2011) *Szemerédi's Theorem via Ergodic Theory*.

[3] Yehuda Shalom, Terence Tao, *A finitary version of Gromov's polynomial growth theorem*, arXiv:0910.4148 28

[4] https://terrytao.wordpress.com/2010/02/18/a-proof-of-gromovs-theorem/ 30

[5] W.T. Gowers, *A NEW PROOF OF SZEMEREDI'S THEOREM* Geom. funct. anal Vol. 11 (2001) 465 - 588 6

[6] Tobias H. Colding; William P. Minicozzi II, *Harmonic Functions on Manifolds*, The Annals of Mathematics, 2nd Ser., Vol. 146, No. 3. (Nov., 1997), pp. 725-747. 6

[7] Bruce Kleiner, *A NEW PROOF OF GROMOV'S THEOREM ON GROUPS OF POLYNOMIAL GROWTH*, Journal of the American mathematical society Volume 23, Number 3, July 2010, Pages 815–829 26