Ergodic theory lecture notes, winter 2015/16

Pavel Zorin-Kranich, pzorin@uni-bonn.de

May 28, 2018

1 Furstenberg correspondence principle

The main motivation for the theory that will be covered in this course is the following.

Theorem 1.1 ([Sze75]). *Let $E \subset \mathbb{Z}$ be a set with positive* upper density

$$\overline{d}(E) := \limsup_{N\to\infty} \frac{|E\cap[1,N]|}{N} > 0. \tag{1.2}$$

Then for every k there exist $a \in \mathbb{Z}$ and $n > 0$ such that

$$a, a+n, \dots, a+kn \in E.$$

The approach that will be presented here has been started in the seminal article of Furstenberg [Fur77] and has led to a number of generalizations of Theorem 1.1, some of which we may discuss, time permitting.

The starting point of this approach is a more flexible reformulation of the above. Let T be the translation operator

$$Tf(n) = f(n+1)$$

on $\ell^\infty(\mathbb{Z}\to\mathbb{C})$. Consider the smallest T-invariant closed sub-$*$-algebra $\mathfrak{A} \subset \ell^\infty$ containing the characteristic function 1_E. Then $\mathfrak{A}$ is separable and there exists a subsequence (N_k) of $\mathbb{N}$ that realizes the supremum in (1.2) and such that

$$\mu(f) := \lim_{k\to\infty} \frac{1}{N_k} \sum_{n=1}^{N_k} f(n)$$

exists for every $f \in \mathfrak{A}$. In particular, μ is a positive bounded linear form on $\mathfrak{A}$ that is T-invariant.

This gives the following reformulation of Szemerédi's theorem.

Theorem 1.3. *Let $\mathfrak{A}$ be a separable commutative unital C^* algebra, T an automorphism of $\mathfrak{A}$, $\mu : \mathfrak{A} \to \mathbb{C}$ a T-invariant positive linear form, and $f \in \mathfrak{A}$ with $f \geq 0$ and $\mu(f) > 0$. Then for every $k \geq 0$ we have*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(f \cdot T^n f \cdots T^{kn} f) > 0.$$

Taking $f = 1_E$ and $\mathfrak{A}$ as above it is clear that Theorem 1.1 is already implied by

$$\mu(f \cdot T^n f \cdots T^{kn} f) > 0$$

for a single $n > 0$, so Theorem 1.3 is formally substantially stronger (but it can be in fact deduced from Theorem 1.1, this might appear as an exercise once we have the necessary technology).

In this lecture we prove the following result

Theorem 1.4 ([Sár78]). *Let $E \subset \mathbb{Z}$ be a set with positive upper density. Then for every polynomial p with integer coefficients and no constant term there exist $a \in \mathbb{Z}$ and $n > 0$ such that*

$$a, a+p(n) \in E.$$

Passing to the translation invariant algebra spanned by 1_E we see that it suffices to prove the following formally stronger statement.

Theorem 1.5. *Let $\mathfrak{A}$ be a separable commutative unital C^* algebra, T an automorphism of $\mathfrak{A}$, $\mu : \mathfrak{A} \to \mathbb{C}$ a T-invariant positive linear form, and $f \in \mathfrak{A}$ with $\mu(f) > 0$. Then for every polynomial p with integer coefficients and zero constant term we have*

$$\liminf_{M\to\infty} \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(f \cdot T^{p(M!n)} \bar{f}) > 0.$$

This is much easier to prove than Theorem 1.3 because this is a Hilbert space problem in disguise. Without loss of generality assume $\mu(1) = 1$ and $p \not\equiv 0$. Consider the sesquilinear form

$$\langle f, g \rangle := \mu(f\bar{g})$$

on $\mathfrak{A}$. By the positivity assumption on μ this form is positive definite, which makes $\mathfrak{A}$ a pre-Hilbert space and T an invertible isometry. Let H be the Hilbert space completion of $\mathfrak{A}$; T extends to a unitary operator on H. The problem now reduces to the following:

we are given a Hilbert space H with a unitary operator T acting on it and a vector $f \in H$. There is also a distinguished element $1 \in H$ with $T1 = 1$ and $\langle f, 1 \rangle > 0$. We have to show

$$\liminf_{M\to\infty} \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left\langle f, T^{p(M!n)} f \right\rangle > 0.$$

Recall that the Borel functional calculus of a normal operator T is the unique homomorphism of unital $*$-algebras that maps a bounded complex-valued Borel function f on the spectrum $\sigma(T)$ to an operator, denoted by $f(T)$, with the following properties:

1. the function $f(z) = z$ is mapped to the operator $f(T) = T$,
2. if f_k is a uniformly bounded sequence of Borel functions that converges pointwise to a function f, then $f_k(T) \to f(T)$ in the strong operator topology.

The spectrum of the unitary operator T is a subset of the unit circle $\Lambda \subset \mathbb{C}$. Let

$$g_{M,N}(\lambda) := \frac{1}{N} \sum_{n=1}^{N} \lambda^{p(M!n)}.$$

These are bounded Borel functions on $\Lambda \supset \sigma(T)$, and with the Borel functional calculus we have

$$\frac{1}{N} \sum_{n=1}^{N} \left\langle f, T^{p(M!n)} f \right\rangle = \langle f, g_{M,N}(T) f \rangle .$$

By the Borel functional calculus it suffices to understand pointwise behaviour of the functions $g_{M,N}$ as first $N \to \infty$ and then $M \to \infty$.

The first claim is that for all M and all $\lambda \in \Lambda$ that are not roots of unity we have $\lim_{N\to\infty} g_{M,N}(\lambda) = 0$. The easiest way to prove this is to use the van der Corput differencing argument. For future use we formulate a Hilbert space valued version of this argument, in the current application the Hilbert space in question will be $\mathbb{C}$.

Proposition 1.6. *Let V be a Hilbert space and let (v_n) be a bounded sequence in V. Then*

$$\limsup_{N\to\infty} \| \frac{1}{N} \sum_{n=1}^{N} v_n \|^2 \leq \limsup_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \limsup_{N\to\infty} | \frac{1}{N} \sum_{n=1}^{N} \langle v_{n+k}, v_n \rangle |$$

Proof. On the left-hand side we can replace the average by the following double average

$$\frac{1}{N} \sum_{n=1}^{N} v_n = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{N} \sum_{n=1}^{N} v_{n+k} + O(K/N)$$

By triangle and Hölder's inequality

$$\| \frac{1}{K} \sum_{k=1}^{K} \frac{1}{N} \sum_{n=1}^{N} v_{n+k} \|^2 \leq \left(\frac{1}{N} \sum_{n=1}^{N} \| \frac{1}{K} \sum_{k=1}^{K} v_{n+k} \| \right)^2 \leq \frac{1}{N} \sum_{n=1}^{N} \| \frac{1}{K} \sum_{k=1}^{K} v_{n+k} \|^2.$$

This can be written

$$\frac{1}{K^2}\sum_{k_1,k_2=1}^{K}\frac{1}{N}\sum_{n=1}^{N}\langle v_{n+k_1}, v_{n+k_2}\rangle \leq \frac{1}{K^2}\sum_{k_1,k_2=1}^{K}\Big|\frac{1}{N}\sum_{n=1}^{N}\langle v_{n+|k_1-k_2|}, v_n\rangle\Big| + O(K/N),$$

and the conclusion follows from

$$\frac{1}{K^2}\sum_{k_1,k_2=1}^{K}\delta_{|k_1-k_2|} = \frac{1}{K^2}\sum_{K'=1}^{K}2\sum_{k=1}^{K'}\delta_k+O(1/K) = \frac{2}{K^2}\sum_{K'=1}^{K}K'(\frac{1}{K'}\sum_{k=1}^{K'}\delta_k)+O(1/K).$$

□

Corollary 1.7. *Let p be a polynomial with real coefficients, at least one of which is irrational. Then*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}e(p(n)) = 0.$$

Proof. Splitting into progressions modulo the least common denominator of the rational coefficients we may assume that the leading coefficient is irrational. Moreover, we may assume that the constant term of p vanishes.

We induct on the degree of p. If $\deg p = 1$, then $p(n) = \alpha n$, so

$$\frac{1}{N}\sum_{n=1}^{N}e(p(n)) = \frac{1}{N}\sum_{n=1}^{N}e^{(2\pi i)\alpha n} = \frac{1}{N}\frac{e^{(2\pi i)\alpha(N+1)}-1}{e^{(2\pi i)\alpha}-1}\to 0$$

as $N\to\infty$. Suppose now $\deg p > 1$. Then for every $k > 0$ we have

$$\limsup_{N\to\infty}\Big|\frac{1}{N}\sum_{n=1}^{N}e(p(n+k))\overline{e(p(n))}\Big| = \limsup_{N\to\infty}\Big|\frac{1}{N}\sum_{n=1}^{N}e(p_k(n))\Big|,$$

where $p_k(n) = p(n+k)-p(n)$ is a polynomial of lower degree with irrational leading coefficient. Hence by the inductive hypothesis this limit vanishes. The conclusion follows from the van der Corput lemma with the Hilbert space $\mathbb{C}$. □

Let us now return to the proof of Theorem 1.5. We have just proved that $g_{M,N}(\lambda)\to 0$ as $N\to\infty$ for λ that are not roots of unity. On the other hand, if λ is a root of unity, then the sequence $(\lambda^{p(M!n)})$ is periodic, so $\lim_{N\to\infty} g_{M,N}$ exists and equals a complete trigonometric sum. The reason for introducing the parameter M is to avoid further analysis of these sums: for a fixed λ and sufficiently large M we will have $\lambda^{p(M!n)} = 1$ for all n. Thus

$$\lim_{M\to\infty}\lim_{N\to\infty} g_{M,N}(\lambda) = \begin{cases} 1 & \text{if } \lambda \text{ is a root of unity and} \\ 0 & \text{otherwise.}\end{cases}$$

In particular,

$$P := \lim_{M\to\infty}\lim_{N\to\infty} g_{M,N}(T)$$

exists in the strong operator topology and is a projection operator, and we have $P(1) = 1$ since $T(1) = 1$.

Therefore

$$\langle f, Pf\rangle \geq \frac{|\langle f,1\rangle|^2}{\langle 1,1\rangle} > 0,$$

and this concludes the proof of Theorem 1.5.

1.1 C^* algebras

In this section we recall the main structural result about commutative C^*-algebras. Recall their definition.

Definition 1.8. A C^*-algebra is an algebra $\mathfrak{A}$ over $\mathbb{C}$ equipped with a Banach space norm $\|\cdot\|$ and an involution $*:\mathfrak{A}\to\mathfrak{A}$ that satisfy the following axioms for all $a,b\in\mathfrak{A}$ and $\mu,\lambda\in\mathbb{C}$:

1. $(\lambda a+\mu b)^*=\bar{\lambda}a^*+\bar{\mu}b^*$ (antilinear),
2. $\|ab\|\le\|a\|\|b\|$ (Banach algebra),
3. $\|a\|^2=\|a^*a\|$ (C^* property),
4. $(ab)^*=b^*a^*$ (antimultiplicative),
5. $(a^*)^*=a$ (involutive)

The positive elements of a C^* algebra are, by definition, the elements of the form a^*a. It is a non-trivial fact that the sum of two positive elements is again positive. A continuous linear form on $\mathfrak{A}$ is called *positive* if it maps positive elements to positive real numbers.

Theorem 1.9 (Gelfand, Naimark, unital separable case)**.** *Let $\mathfrak{A}$ be a commutative unital separable C^* algebra. Its* Gelfand spectrum *$\hat{\mathfrak{A}}$ is by definition the set of all unital $*$-homomorphisms from $\mathfrak{A}$ to $\mathbb{C}$. Every $*$-homomorphism is continuous, positive, and bounded in norm by* 1*. Hence $\hat{\mathfrak{A}}$ inherits the weak-$*$ topology from the Banach space dual $\mathfrak{A}'$, and with this topology $\hat{\mathfrak{A}}$ is a compact metrizable space. Moreover, the map*

$$\mathfrak{A}\to C(\hat{\mathfrak{A}},\mathbb{C}),\qquad a\mapsto(\varphi\mapsto\varphi(a))$$

is an isomorphism of C^* algebras (the norm on $C(\hat{\mathfrak{A}},\mathbb{C})$ being the supremum norm).

The proof can be found in any of the standard books on C^* algebras, e.g. by Takesaki [Tak02].

In the construction of Furstenberg we obtain, along with a commutative unital C^*-algebra $\mathfrak{A}$, also an algebra automorphism T and a positive continuous linear form $\mu:\mathfrak{A}\to\mathbb{C}$. The former induces a homeomorphism on $\hat{\mathfrak{A}}$, again denoted by T, by the formula $T\varphi=\varphi\circ T$. The latter corresponds to an (outer and inner) regular Borel probability measure on $\hat{\mathfrak{A}}$, again denoted by μ, with the property

$$\int_{\hat{\mathfrak{A}}}a(\varphi)\mathrm{d}\mu(\varphi)=\mu(a)$$

for every $a\in\mathfrak{A}$. The correspondence is given by the Riesz–Markov–Kakutani representation theorem.

2 Ergodicity

2.1 Three perspectives on measure-preserving dynamical systems

In the last lecture we have seen a construction of an "algebraic" measure-preserving dynamical system. Such system consists of the following data:

1. A commutative separable unital $C*$-algebra $\mathfrak{A}$,
2. a positive linear functional $\mu : \mathfrak{A} \to \mathbb{C}$ (without loss of generality we will assume $\mu(1) = 1$ from now on),
3. and an automorphism $T : \mathfrak{A} \to \mathfrak{A}$ that preserves μ in the sense $\mu \circ T = \mu$.

The Gelfand spectrum $\hat{\mathfrak{A}}$ is by definition the set of non-zero algebra homomorphisms $\mathfrak{A} \to \mathbb{C}$. It is a weak-$*$ closed subset of the the Banach space dual $\mathfrak{A}'$ and therefore a compact metrizable space. The map

$$\mathfrak{A} \to C(\hat{\mathfrak{A}}, \mathbb{C}), \qquad a \mapsto (\varphi \mapsto \varphi(a))$$

is a C^*-algebra isomorphism by the Gelfand–Naimark theorem. There is a unique regular Borel measure μ on $\hat{\mathfrak{A}}$ satisfying

$$\int_{\hat{A}} \varphi(a) \mathrm{d}\mu(\varphi) = \mu(a),$$

and the map $T\varphi = \varphi \circ T$ is a homeomorphism of $\hat{A}$ that preserves the measure μ in the sense that

$$\int_{\hat{A}} f \mathrm{d}\mu = \int_{\hat{A}} (f \circ T) \mathrm{d}\mu \tag{2.1}$$

for every $f \in C(\hat{A})$. We will write

$$Tf := f \circ T.$$

This gives a second perspective on measure-preserving dynamics. A "topological" measure-preserving dynamical system (mps) consists of the following data.

1. A compact metric space X,
2. a regular Borel probability measure μ on X,
3. and a homeomorphism $T : X \to X$ that preserves μ.

The correspondence between algebraic and topological mps's is one-to-one. Many important concepts in measure-preserving dynamics are most conveniently defined purely in terms of the measurable structure of X and do not directly involve the topology (the first example being ergodicity, which we will discuss later in this lecture). Let us therefore make the following definition.

Definition 2.2. A *measure-preserving dynamical system (mps)* consists of the following data:

1. A complete separable measure space X,
2. a probability measure μ on X,
3. and a measurable, invertible map $T : X \to X$ that preserves the measure μ in the sense that (2.1) holds for all $f \in \mathcal{X} := L^\infty(X, \mu)$.

From now on we denote the C^* algebra of bounded measurable functions modulo equality almost everywhere by the calligraphic version of the letter that denotes the base space. The full notation for an mps is (X, μ, T), but it may be abbreviated to X or $\mathcal{X}$, context permitting.

Clearly, a "topological" mps induces an mps by forgetting the topological structure. This process is not invertible, because on a given compact metric space there typically exist many other compact metrizable topologies with the same Borel structure. One can nevertheless attempt to invert it by observing that $L^\infty(X,\mu)$ is a C^*-algebra, μ induces a positive linear functional on it, and T a μ-preserving algebra automorphism. This has the downside that the Gelfand spectrum of $L^\infty(X,\mu)$ is in general non-metrizable (unless X is finite), and metrizability is desirable for a number of technical reasons.

The right thing to do is to consider a separable closed T-invariant L^2-dense sub-$*$-algebra $\mathfrak{A} \subset L^\infty(X,\mu)$. The Gelfand spectrum $\hat{\mathfrak{A}}$ (with measure μ and homeomorphism T) is then called a *topological model* of the mps (X,μ,T). From a topological model we can recover the original C^*-algebra $L^\infty(X,\mu)$ as follows. As in the previous lecture, consider the inner product

$$\langle a,b\rangle := \mu(ab^*)$$

on $\mathfrak{A}$. This coincides with the inner product on $L^2(X,\mu)$, and by the density assumption the Hilbert space completion H of $(\mathfrak{A},\langle\cdot,\cdot\rangle)$ is isomorphic to $L^2(X,mu)$. We have an injective C^*-algebra homomorphism $\iota : \mathfrak{A} \to L(H)$, with the operator $\iota(a)$ given by $\iota(a)h = ah$ for $h \in \mathfrak{A}$ and extended to H by continuity. By the von Neumann double commutant theorem, the closure of $\iota(\mathfrak{A})$ in the weak operator topology on $L(H)$ equals the double commutant[1] $\iota(\mathfrak{A})''$.

On the other hand, $L^\infty(X)$ embeds into $L(H)$ as the space of multiplication operators, and this space is weakly closed in $L(H)$. The weak operator topology on this space coincides with the weak-$*$ topology on L^∞ as a dual space of L^1. Hence it suffices to show that the unit ball $B_{\mathfrak{A}}$ of $\mathfrak{A}$ is weak-$*$-dense in the unit ball B_∞ of L^∞ in order to establish that

$$L^\infty(X,\mu) \cong \iota(\mathfrak{A})''$$

as C^* algebras. Using the fact that $\mathfrak{A}$ is an algebra and the Stone–Weierstraß theorem it is not hard to show that $B_{\mathfrak{A}}$ is L^2 dense in B_∞. But on the unit ball B_∞ the L^2 topology is finer that the $\sigma(L^\infty, L^1)$ topology, and we are done.

Running the same argument on $L^\infty(\hat{\mathfrak{A}},\mu)$ (using the fact that $\mathfrak{A} \cong C(\hat{\mathfrak{A}})$ is dense in $L^2(\hat{\mathfrak{A}},\mu)$) we see that

$$L^\infty(\hat{\mathfrak{A}},\mu) \cong L^\infty(X,\mu) \text{ as } C^* \text{ algebras.}$$

In other words, the passage to a topological model preserves the algebra of bounded measurable functions (and also μ and T, which is in a easier to show in the sense that no sophisticated tools such as the double commutant theorem are needed).

Since we will be mostly concerned with results that can be formulated in terms of bounded measurable functions, we will be free to choose topological models for our measure-preserving systems. It will be convenient to choose different models at different stages of our investigations, and the above result gives us the freedom to do so.

2.2 Factors

Definition 2.3. Let (X,μ,T) be an mps. A *factor* of X is a (closed) T-invariant unital sub-$C*$-algebra of $\mathcal{X}$.

[1] If $A \subset L(H)$ is a C^* algebra, then its commutant is defined by $A' := \{b \in L(H) : \forall a \in A\, ab = ba\}$. The double commutant is $A'' = (A')'$. Observe that $A'' \supseteq A$. We will not use the notion of commutant outside of the current argument, and A' will otherwise always stand for the Banach space dual of A.

Example. The set of invariant functions

$$\mathcal{I}(X,T) := \{f \in \mathcal{X} : Tf = f\}$$

is a factor of X, called the *invariant factor.*

If $\mathcal{Y} \subset \mathcal{X}$ is a factor, then every topological model $\mathfrak{B}$ of $\mathcal{Y}$ can be extended to a topological model $\mathfrak{A}$ of $\mathcal{X}$:

$$\begin{tikzcd}
\mathcal{Y} \arrow[r, hook] & \mathcal{X} \\
\mathfrak{B} \arrow[u, hook] \arrow[r, hook] & \mathfrak{A} \arrow[u, hook]
\end{tikzcd}$$

What does this say about the corresponding compact metric spaces? Since $\mathfrak{B} \subset \mathfrak{A}$, we have a natural map $\pi : \hat{\mathfrak{A}} \to \hat{\mathfrak{B}}$ between Gelfand spectra, which maps a C^*-algebra homomorphism defined on $\mathfrak{A}$ to its restriction to $\mathfrak{B}$. The map π is clearly continuous, T-equivariant, and pushes the measure induced from μ on $\hat{\mathfrak{A}}$ to the measure induced from μ on $\hat{\mathfrak{B}}$ (hence it makes sense to denote both these measures by μ).

A less obvious fact is that the map π is surjective. This is most easily seen using an alternative characterization of $\hat{\mathfrak{A}}$ for a commutative C^* algebra $\mathfrak{A}$. Namely,

$$\hat{\mathfrak{A}} = \operatorname{extr} M(\hat{\mathfrak{A}}).$$

Here extr stands for "extremal points" and $M(\hat{\mathfrak{A}})$ is the set of regular Borel probability measures on the compact metric space $\hat{\mathfrak{A}}$. Indeed, by the Riesz–Markov–Kakutani representation theorem we have

$$M(\hat{\mathfrak{A}}) = \{\varphi \in \mathfrak{A}' : \|\varphi\| \leq 1,\ \varphi(1) = 1\},$$

and this is a convex set which is weak-$*$-compact by the Banach–Alaoglu theorem. Its extreme points are the (Dirac δ) point measures.

Now, given $\psi \in M(\hat{\mathfrak{B}})$ consider the set

$$\{\varphi \in M(\hat{\mathfrak{A}}) : \varphi|_{\mathfrak{B}} \equiv \psi\}.$$

This is a weak-$*$-compact convex set, and it is non-empty by the Hahn–Banach theorem. By the Krein–Milman theorem it has an extreme point, and it is not hard to verify that every such extreme point must already be an extreme point of $M(\hat{\mathfrak{A}})$ using extremality of ψ in $M(\hat{\mathfrak{B}})$. Thus we have found a $\psi \in \hat{\mathfrak{A}}$ that maps to ψ under π.

Summarizing, the topological model of a factor has the form of a surjective continuous map

$$\hat{\mathfrak{A}} \to \hat{\mathfrak{B}}$$

which intertwines the maps T on the left and on the right and pushes the measure μ from the left to the right.

This description also makes sense in the measurable category: in the literature a factor is frequently defined as a measurable map $\pi : X \to Y$ between mps (X, μ, T) and mps (Y, ν, S) such that $\pi \circ T = S \circ \pi$ holds almost everywhere and the pushforward measure $\pi_* \mu$ equals ν. I prefer the algebraic definition because invariant algebras of functions are easier to construct than corresponding measure spaces, as is already apparent from the example of the invariant factor.

2.3 Invariant factors in some examples

Example (Rotation on the torus). The simplest measure-preserving system (after the finite ones) is a rotation on the circle. Let $X = \mathbb{T} := \mathbb{R}/\mathbb{Z}$ be the torus with the Lebesgue measure μ and the map $Tx = T_\alpha x := x + \alpha$ with some fixed $\alpha \in \mathbb{T}$. The invariant factor $\mathcal{I}(X)$ clearly depends on α. If α is rational with denominator q in reduced form, then $\mathcal{I}(X)$ consists precisely of the $1/q$-periodic functions.

Suppose now that α is irrational. For any L^2 function f we have

$$\widehat{Tf}(n) = e(n\alpha)\widehat{f}(n),$$

where $\widehat{\cdot}$ denotes the Fourier transform. Therefore, $f \in \operatorname{fix} T$ if and only if all but the 0-th Fourier coefficients vanish. Hence in this case

$$\mathcal{I}(X) = \mathbb{C}1_X. \tag{2.4}$$

An mps for which (2.4) holds is called *ergodic*.

An example of a non-ergodic mps is given by a rational rotation with $\alpha = \frac{1}{q}$ rational. In this example we can write X as a product space $\{0, \frac{1}{q}, \dots, \frac{q-1}{q}\} \times [0, \frac{1}{q})$, and the transformation T factors into a cyclic permutation on the first multiplicand and the identity of the second multiplicand. Hence the overall system is a union of infinitely many copies of $\{0, \frac{1}{q}, \dots, \frac{q-1}{q}\}$, one for each point in $[0, \frac{1}{q})$. This gives a very misleading picture of what a generic non-ergodic system looks like. It is true in general that any non-ergodic system is essentially a union of ergodic systems (this will be proved in the next lecture). However, the ergodic components may vary wildly.

Example. Consider the space $X = \mathbb{T}^2$ with the Lebesgue measure and the transformation

$$T(x, y) := (x, y + x).$$

Let $\mathcal{Y} \subset \mathcal{X}$ be the space of functions that depend only on the first coordinate. Then $\mathcal{I}(X) = \mathcal{Y}$. Indeed, the inclusion $\supseteq$ is clear. To see the converse, take $f \in \mathcal{I}(X)$ and fix an everywhere defined representative for it (recall that L^∞ is defined modulo equality almost everywhere). In order for f to be T-invariant, the following must hold: for almost every x we have

$$f(x, \cdot) \in \mathcal{I}(\mathbb{T}, T_x).$$

On the other hand, almost every x is irrational, and then $\mathcal{I}(\mathbb{T}, T_x) = \mathbb{C}1_{\mathbb{T}}$. Hence $f(x, \cdot)$ is equivalent to a constant for almost every x, and the claim follows using Fubini's theorem.

For your amusement, here is another ergodic mps that plays a role in the theory of continued fractions.

Example (Gauss). Let $X = [0, 1)$ and $Tx := \{1/x\}$ (fractional part of $1/x$). Then the measure $\mathrm{d}\mu(x) = \frac{1}{\log 2} \frac{\mathrm{d}x}{1+x}$ is T-invariant and the mps (X, μ, T) is ergodic.

Ergodicity is substantially harder to prove here than above. A well-known (family of) open problem(s) in thermodynamics involving ergodicity is the *ergodic hypothesis*, which postulates that certain Hamiltonian systems (equipped with the Lioville measure) are ergodic.

2.4 Mean ergodic theorem

Let (X, μ, T) be an mps. In the proof of Sárkőzy's theorem on polynomial differences in sets of positive measure we have observed that, for every function $f \in L^2(X, \mu)$, the limit

$$Pf := \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} T^n f \tag{2.5}$$

exists in $L^2(X, \mu)$. Moreover, P is a projection operator given by the functional calculus of T as the image of the indicator function of $\{1\} \subset \sigma(T)$. Note that $1 \in \sigma(T)$ because the constant function 1_X is an eigenvector of T with this eigenvalue.

What is an explicit description of P? The answer is that its range consists of the T-invariant functions:

$$\operatorname{ran} P = \operatorname{fix} T.$$

This fact, together with the existence of the above limit, is known as the *mean ergodic theorem* (on $L^2(X)$). Let us prove the last inequality. The inclusion $\supseteq$ is clear from (2.5). On the other hand, suppose $g \in \operatorname{ran} P$, so $g = Pf$. Then

$$\begin{aligned} Tg = T \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^n f &= \lim_{N\to\infty} T \frac{1}{N} \sum_{n=1}^{N} T^n f \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^n f + \frac{T^{N+1} f - Tf}{N} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^n f = g \end{aligned}$$

as required. In particular we have

$$\mathcal{I}(X) = \operatorname{ran} P \cap \mathcal{X}.$$

for the invariant factor.

3 Measure disintegration and ergodic decomposition

The next few lectures (up to relatively compact and weakly mixing extensions) will cover classical material which appears most notably in [Fur81], [EW11], [Tao09].

Recall that a measure-preserving dynamical system (X, μ, T) is called *ergodic* if the T-invariant subspace $\mathcal{I}(X) \subset \mathcal{X}$ consists only of the constant functions. It is in fact possible to write any measure-preserving system as a direct integral of ergodic systems, similarly to the example $(x, y) \mapsto (x, y + x)$ on the 2-torus. More precisely, the following holds.

Proposition 3.1 (Ergodic decomposition). *Let (X, μ, T) be an mps. Then, upon passing to a suitable topological model for the invariant factor Y, there exists a continous T-invariant map*

$$\mu_{\cdot} : Y \to M(X), \qquad y \mapsto \mu_y$$

such that

$$\mu = \int_Y \mu_y \mathrm{d}\mu(y) \tag{3.2}$$

and, for μ-almost every y, the measure μ_y is T-invariant and the mps (X, μ_y, T) is ergodic.

One application of this result is the reduction of the multiple recurrence problem to ergodic systems. Recall that one of our goals is to prove Szemerédi's theorem in the following form: let (X, μ, T) be an mps and $0 \leq f \in \mathcal{X}$ a not identically zero function. Then

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int_X f \cdot T^n f \cdots T^{kn} f \mathrm{d}\mu > 0.$$

Suppose that this is known for ergodic systems. In the general case we may write the left-hand side as

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int \int f \cdot T^n f \cdots T^{kn} f \mathrm{d}\mu_x \mathrm{d}\mu,$$

and by Fatou's lemma this is bounded from below by

$$\int \Big(\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int f \cdot T^n f \cdots T^{kn} f \mathrm{d}\mu_x \Big) \mathrm{d}\mu.$$

Since f is not almost everywhere zero, it is not μ_x-a.e. zero for a positive μ-measure set of x, so the quantity in the brackets is positive on a positive measure set by the ergodic case of the multiple recurrence theorem.

A decomposition of a measure of the form (3.2) is called a *measure disintegration*. We will construct measure disintegrations over a general factor and obtain Proposition 3.1 in the case of the invariant factor.

3.1 Conditional expectation

Definition 3.3. Let $\mathcal{Y} \subset \mathcal{X}$ be a factor. The *conditional expectation onto $\mathcal{Y}$* is the orthogonal projection from $L^2(X)$ to $L^2(Y)$. It is denoted by $\mathbb{E}(\cdot|\mathcal{Y})$.

The L^2 spaces in this definition can be thought of as the Hilbert space completions of the respective C^* algebras or the L^2 spaces on Gelfand spectra of some compatible topological models of X and Y. This definition does not make use of the measure-preserving transformation T, nothing changes if e.g. T is replaced by the identity map.

Lemma 3.4. *The conditional expectation has the following properties.*

1. $\mathbb{E}(1|Y) = 1$.
2. $\int \mathbb{E}(f|Y) = \int f$ *for* $f \in L^1(X)$.
3. *Let* $f \in L^2(X)$ *and* $F \subset Y$ *measurable. Then* $\mathbb{E}(f1_F|Y) = \mathbb{E}(f|Y)1_F$.
4. *Conditional expectation maps positive functions in* $L^1(X)$ *to positive functions.*
5. $\mathbb{E} : L^\infty(X) \to L^\infty(Y)$ *is a contraction.*
6. $\mathbb{E} : L^1(X) \to L^1(Y)$ *is a contraction.*
7. *Assume that* $f \in L^1(X), g \in L^0(Y)$ *and either* $fg \in L^1(X)$ *or* $f \geq 0, \mathbb{E}(f|Y)g \in L^1(Y)$. *Then*

$$\mathbb{E}(fg|Y) = \mathbb{E}(f|Y)g, \text{ and both functions are in } L^1(Y).$$

This is of course well-known but it is important to have the weakest possible assumptions in (7).

Proof. (1) holds since $1 \in L^2(Y)$.

(2) holds for $f \in L^2(X)$ since $\int \mathbb{E}(f|Y) = \langle \mathbb{E}(f|Y), 1\rangle = \langle f, \mathbb{E}(1|Y)\rangle = \langle f, 1\rangle = \int f$.

For (3) note first that $\operatorname{supp} \mathbb{E}(f1_F|Y) \subset F$, since otherwise $\mathbb{E}(f1_F|Y)1_F \in L^2(Y)$ would have strictly smaller L^2 distance to $f1_F$, contradicting the fact that $\mathbb{E}$ is an orthogonal projection. Suppose now $\mathbb{E}(f1_F|Y) \neq \mathbb{E}(f|Y)1_F$, then

$$\int_F |\mathbb{E}(f1_F|Y) - f|^2 < \int_F |\mathbb{E}(f|Y) - f|^2.$$

It follows that the function $g := \mathbb{E}(f1_F|Y)1_F + \mathbb{E}(f|Y)1_{F^c} \in L^2(Y)$ has strictly smaller L^2 distance to f than $\mathbb{E}(f|Y)$, a contradiction.

To show (4) let $0 \leq f \in L^2(X)$ and $F = \{\mathbb{E}(f|Y) < 0\}$. Then $\|f1_F - 0\| < \|f1_F - 1_F\mathbb{E}(f|Y)\| = \|f1_F - \mathbb{E}(f1_F|Y)\|$, which is a contradiction unless $F = \emptyset$.

To show (5) note that $\Pi^k(z) = z \cdot \min(1, k/|z|)$ is a contraction on $\mathbb{C}$ for every $k \geq 0$. It follows that

$$\int |f - \mathbb{E}(f|Y)|^2 \geq \int |\Pi^{\|f\|_\infty} \circ f - \Pi^{\|f\|_\infty} \circ \mathbb{E}(f|Y)|^2 = \int |f - \Pi^{\|f\|_\infty} \circ \mathbb{E}(f|Y)|^2$$

with equality if and only if $\|\mathbb{E}(f|Y)\|_\infty \leq \|f\|_\infty$. But strict inequality would contradict the fact that $\mathbb{E}(f|Y)$ is the function in $L^2(Y)$ that has the smallest distance from f.

Since $\mathbb{E}$ is self-adjoint and $L^\infty(X) \subset L^2(X)$ this implies (6). Thus $\mathbb{E}$ can be extended to a contraction $L^1(X) \to L^1(Y)$ by continuity. The properties (2) and (4) continue to hold for $f \in L^1(X)$.

Consider now (7). By linearity we obtain $\mathbb{E}(fg|Y) = \mathbb{E}(f|Y)g$ for $f \in L^2(X)$ and simple functions $g \in L^\infty(Y)$. By density we may weaken the assumption to $g \in L^\infty(Y)$.

Suppose now $fg \in L^1(X)$ and denote the truncation of g at level k by $g^k := \Pi^k g$. By (4), the monotone convergence theorem, (2), and monotone convergence theorem again we see that

$$\begin{aligned}\int |\mathbb{E}(f|Y)g| \leq \int \mathbb{E}(|f||Y)|g| = \lim_k \int \mathbb{E}(|f||Y)|g^k| = \lim_k \int \mathbb{E}(|f||g^k||Y) \\ = \lim_k \int |f||g^k| = \int |fg|,\end{aligned}$$

so that $\mathbb{E}(f|Y)g \in L^1(Y)$. Moreover, the inequality turns into an equality in the case $f \geq 0$, and we obtain the converse implication.

By linearity we may now assume $f, g \geq 0$. Then, by the monotone convergence theorem,

$$\mathbb{E}(fg|Y) = \lim_{k\to\infty} \mathbb{E}(f^k g^k|Y) = \lim_{k\to\infty} \mathbb{E}(f^k|Y)g^k = \mathbb{E}(f|Y)g. \qquad \square$$

3.2 Measure disintegration

Theorem 3.5 (Measure disintegration)**.** *Let (X, μ, T) be an mps and $\mathcal{Y} \subset \mathcal{X}$ a factor. Then, upon passing to a suitable topological model, there exists a continous map*

$$\mu_\cdot : Y \to M(X), \qquad y \mapsto \mu_y$$

such that (3.2) *holds, $\mu_{Ty} = T_*\mu_y$, and for every representative f of every equivalence class (modulo equality a.e.) in $L^1(X)$ we have*

$$\int f \mathrm{d}\mu_y = \mathbb{E}(f|Y)(y) \tag{3.6}$$

pointwise a.e. (in particular, $f \in L^1(\mu_y)$ for a.e. $y \in Y$).

Finally, let $\pi : X \to Y$ be the spatial factor map. Then for μ-a.e. y and μ_y-a.e. $x \in X$ we have

$$\mu_x := \mu_{\pi(x)} = \mu_y. \tag{3.7}$$

Proof. We use (3.6) to define the measures μ_y. In order to do so we first choose a suitable topological model. Let $\mathfrak{B}_0 \subset \mathfrak{A}_0$ be any topological model of $\mathcal{Y} \subset \mathcal{X}$. Define inductively

$$\mathfrak{B}_{n+1} := \mathbb{E}(\mathfrak{A}_n|Y), \qquad \mathfrak{A}_{n+1} := \langle \mathfrak{A}_n, \mathfrak{B}_{n+1} \rangle .$$

This is an increasing sequence of topological models since $\mathbb{E}(\mathfrak{A}_n|Y)$ is separable by L^∞-contractivity of conditional expectation. Let finally

$$\mathfrak{B} := \overline{\cup_{n\in\mathbb{N}} \mathfrak{B}_n}, \qquad \mathfrak{A} := \overline{\cup_{n\in\mathbb{N}} \mathfrak{A}_n}.$$

Then $\mathfrak{B} \subset \mathfrak{A}$ and $\mathbb{E}(\mathfrak{A}|Y) = \mathfrak{B}$. Write $Y := \hat{\mathfrak{B}}$.

For each $y \in Y$ define a linear form on $\mathfrak{A}$ by

$$\mu_y(f) := \mathbb{E}(f|Y)(y).$$

This is a positive linear form and $\|\mu_y\|_{L^\infty\to\mathbb{C}} = 1$ by the properties of conditional expectation. Moreover, by T-invariance of $\mathfrak{B}$ we have

$$\mu_{Ty}(f) = \mathbb{E}(f|Y)(Ty) = \mathbb{E}(f \circ T|Y)(y) = (T_*\mu_y)(f).$$

Now we will show (3.7). Let $0 \leq g \in \mathfrak{B}$, then

$$\begin{aligned}\iint |g(y) - g(\pi(x))|^2 \mathrm{d}\mu_y(x)\mathrm{d}\mu(y) &= \iint g(y)^2 - 2g(y)g(\pi(x)) + g(\pi(x))^2 \mathrm{d}\mu_y(x)\mathrm{d}\mu(y) \\ &= \int g(y)^2 - 2g(y)\mathbb{E}(g|Y)(y) + \mathbb{E}(g^2|Y)(y)\mathrm{d}\mu(y) \\ &= 0,\end{aligned}$$

so for μ-a.e. y we have $g \circ \pi = g(y)$ μ_y-a.e.. It follows that $\pi_*\mu_y = \delta_y$, and in particular (3.7) holds.

It remains to extend (3.6) to $f \in L^1(X)$. Consider first an everywhere defined bounded function f on X. Take a bounded sequence $(f_k) \subset \mathfrak{A}$ such that $f_k \to f$ in $L^2(X)$ and pointwise almost everywhere. Then for a.e. y we have convergence μ_y-pointwise a.e. and also $\mathbb{E}(f_k|Y)(y) \to \mathbb{E}(f|Y)(y)$. The dominated convergence now gives (3.6). Integrable functions are handled similarly, restricting to positive functions and using the monotone convergence theorem. $\square$

Proof of Proposition 3.1. Consider the measure disintegration over the invariant factor constructed in Theorem 3.5. Since $T|_{\mathcal{I}} = \mathrm{id}_{\mathcal{I}}$ we obtain $T_*\mu_y = \mu_{Ty} = \mu_y$.

It remains to show that a.e. measure μ_y is T-ergodic. To this end we consider the alternative description of the invariant factor provided by the mean ergodic theorem. Let $f \in L^\infty(X)$. The mean ergodic theorem tells that

$$\frac{1}{N}\sum_{n=1}^{N} T^n f \to \mathbb{E}(f|\mathcal{I})$$

in L^2 as $N \to \infty$. Passing to a subsequence[2] we may assume convergence pointwise a.e. In particular, for a.e. y we have convergence μ_y-a.e. and, since the sequence is uniformly bounded in L^∞, also in $L^2(\mu_y)$.

On the other hand, by the mean ergodic theorem for the mps (X, μ_y, T) we also have

$$\frac{1}{N}\sum_{n=1}^{N} T^n f \to \mathbb{E}(f|\mathcal{I}(X, \mu_y, T))$$

in $L^2(\mu_y)$. It follows that

$$\mathbb{E}(f|\mathcal{I}(X, \mu_y, T))(z) = \mathbb{E}(f|\mathcal{I})(z) = \int f \mathrm{d}\mu_z$$

for μ_y-a.e. z. By (3.7) this function of z is constant μ_y-a.e., so $\mathbb{E}(f|\mathcal{I}(X, \mu_y, T))$ is a constant function. By separability of $\mathfrak{A}$ it follows that $\mathbb{E}(\mathfrak{A}|\mathcal{I}(X, \mu_y, T)) = \mathbb{C}1$, so that (X, μ_y, T) is ergodic. $\square$

[2]By the *pointwise ergodic theorem* the full sequence already converges poinwise a.e., but we will not need this fact.

4 Kronecker factor

4.1 Weyl equidistribution theorem

We will need the following Fourier analytic fact.

Theorem 4.1 (Weyl equidistribution theorem)**.** *Let $\alpha_1, \dots, \alpha_d \in \mathbb{R} \setminus \mathbb{Q}$ be rationally independent. Then the sequence $n\vec{\alpha}$ is equidistributed modulo $\mathbb{Z}^d$ in the sense that*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(n\vec{\alpha} + \mathbb{Z}^d) = \int f$$

for every continuous function $f \in C(\mathbb{R}^d/\mathbb{Z}^d)$. The integral on the right-hand side is taken with respect to the Lebesgue measure.

In particular, this shows that the sequence $n\vec{\alpha}$ is dense modulo $\mathbb{Z}^d$, which is what will be used in this lecture.

Proof. Approximating f uniformly we may assume that f is smooth, and in particular that its Fourier series converges absolutely. In the latter case it suffices to consider a single Fourier mode, $f(\vec{x}) = e^{2\pi i \sum_i k_i x_i}$. Then

$$\frac{1}{N} \sum_{n=1}^{N} f(n\vec{\alpha} + \mathbb{Z}^d) = \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i n \sum_i k_i \alpha_i}.$$

There are two cases. If $k\vec{k} = \vec{0}$, then this is indentically 1, and the limit is 1 as required. Otherwise the number $\sum_i k_i \alpha_i$ is irrational, and in particular non-integer, by the hypothesis, and the average converges to 0 as required. □

4.2 Eigenfunctions

Throughout the remaining part of the lecture let (X, μ, T) be an ergodic measure-preserving system. Consider an eigenvector $f \in L^2$ of T. Since T is unitary, the corresponding eigenvalue λ has absolute value $|\lambda| = 1$. Moreover, since T comes from a transformation on X, we have

$$T|f| = |Tf| = |\lambda f| = |f|.$$

Hence, by ergodicity assumption, $|f|$ is a constant function. In particular, $f \in L^\infty(X)$. Note also that the constant function 1 is an eigenfunction to eigenvalue 1.

Let now f_1, f_2 be two eigenfunctions with eigenvalues λ_1, λ_2, respectively. We may normalize $|f_1| = |f_2| \equiv 1$. Then, since T is an algebra homomorphism, we have

$$T(f_1 \bar{f}_2) = Tf_1 \cdot \overline{Tf_2} = \lambda_1 f_1 \overline{\lambda_2 f_2}.$$

Hence the set of L^∞-normalized eigenfunctions is a group under multiplication, and the point spectrum $\sigma_d(T)$ is a subgroup of the complex unit circle Λ.

Moreover, if $\lambda_1 = \lambda_2$, then $f_1 \bar{f}_2$ is an eigenfunction to eigenvalue 1, so it is a constant function. It follows that all eigenspaces of T are at most 1-dimensional.

Let $\mathcal{E}$ denote the L^∞ closed linear subspace of $\mathcal{X}$ spanned by the eigenfunctions of X. Since products of eigenfunctions are eigenfunctions, this is a subalgebra. The factor $\mathcal{E}$ is called the *Kronecker factor*. It has a useful spatial description.

Definition 4.2. A group Γ is called *monothetic* if there exists a group element γ such that the orbit $\{\gamma^n, n \in \mathbb{Z}\}$ is dense in Γ.

Theorem 4.3 (Halmos–von Neumann)**.** *There exists a compact metrizable monothetic Abelian group (G, γ) and a homeomorphism $\hat{\mathcal{E}} \cong \Gamma$ that intertwines T with the map $g \mapsto \gamma g$ and pushes the measure μ forward to the Haar measure on G.*

Proof. Define

$$G := \{\varphi : \sigma_d(T) \to \Lambda \text{ homomorphism}\}, \qquad \Lambda = \{z \in \mathbb{C} : |z| = 1\}$$

with pointwise operations and the topology of pointwise convergence (G is the Pontryagin dual of the group $\sigma_d(T)$ equipped with the discrete topology). It is clear that G is compact, metrizable, and Abelian.

Fix any point $a \in \hat{\mathcal{E}}$ and for each eigenvalue $\lambda \in \sigma_d(T)$ fix the (unique) corresponding eigenfunction f_λ with $f_\lambda(a) = 1$. Define the map

$$\Phi : \hat{\mathcal{E}} \to G, \qquad a \mapsto (f_\lambda(a))_\lambda.$$

This is well-defined (in the sense that the right-hand side is an element of G) because $f_{\lambda_1}\overline{f_{\lambda_2}} = f_{\lambda_1\bar{\lambda}_2}$ by construction. The map Φ is clearly continuous from the weak* topology to the topology of pointwise convergence. Moreover,

$$\Phi(Ta) = (f_\lambda(Ta))_\lambda = (\lambda f_\lambda(a))_\lambda = (\lambda)_\lambda (f_\lambda(a))_\lambda,$$

so Φ intertwines T with the translation by the group element $\gamma := (\lambda)_\lambda$.

Now we will show that the orbit of γ is dense in G. By definition this means that for every finite set $F \subset \sigma_d(T)$, every homomorphism $\varphi : \sigma_d(T) \to \Lambda$, and every $\varepsilon > 0$ there is a power of γ that approximates φ on F to within ε.

Consider the subgroup $\langle F\rangle$ of $\sigma_d(T)$ generated by F. By the structure theorem for finitely generated Abelian groups it is isomorphic to $\mathbb{Z}^d \times \prod_i \mathbb{Z}/r_i\mathbb{Z}$. But $\sigma_d(T) \subset \Lambda$, and Λ has only 1 subgroup of order r for each integer $r \geq 1$, so by the Chinese remainder theorem $\langle F\rangle \cong \mathbb{Z}^d \times \mathbb{Z}/r\mathbb{Z}$. Let $\lambda_1, \dots, \lambda_d$ and λ_0 be generators of $\langle F\rangle$. Since φ is a homomorphism, we may assume $F = \{\lambda_0, \dots, \lambda_d\}$.

Since φ is a homomorphism, the order of the value $\varphi(\lambda_0)$ must be divisible by the order of λ_0, so it in fact lies in the subgroup generated by λ_0. Hence, multiplying φ by a power of γ, we may assume $\varphi(\lambda_0) = 1$. It remains to approximate φ on $\{\lambda_1, \dots, \lambda_d\}$ by powers of γ^r. But the values $\lambda_1^r, \dots, \lambda_d^r$ are rationally independent, so this is possible by Weyl's equidistribution theorem.

This shows in particular that $\Phi(\hat{\mathcal{E}})$ is dense in G, and since $\mathcal{E}$ is compact and by continuity the map Φ is surjective. Compactness also implies that Φ is a homeomorphism.

Finally, the pushforward measure $\Phi_*\mu$ is a Borel probability measure on G that is invariant under the shift by the element γ. Since the orbit of γ is dense in G, it is in fact invariant under the action of G on itself. But there is only one such measure, namely the Haar measure. □

The construction of G shows that the structure of the factor $\mathcal{E}$ is uniquely determined by the point spectrum $\sigma_d(T)$, so the point spectrum classifies measure-preserving dynamical systems for which $\mathcal{E}$ is L^2-dense. Such systems are called *compact*.

4.3 Orthogonal complement of the Kronecker factor

We define several subspaces of $L^2(X)$.

- The space spanned by *eigenfunctions* of T:

$$E(X) := \overline{\mathcal{E}} \subset L^2(X).$$

- The space of *almost periodic* functions

$$A(X) := \{f : T^{\mathbb{Z}} f \subset L^2 \text{ totally bounded}\} \subset L^2(X).$$

- The *weakly mixing* space

$$W(X) := \{f : \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} |\langle T^n f, f\rangle|^p = 0\} \subset L^2(X),$$

where $0 < p < \infty$.

Note that the space $W(X)$ does not depend on $0 < p < \infty$ because the sequence $|\langle T^n f, f\rangle|$ is bounded, and for every positive bounded sequence (a_n) and $0 < p < q < \infty$ by Jensen's inequality and a termwise estimate we have

$$\Big(\frac{1}{N}\sum_{n=1}^{N} a_n^p\Big)^{1/p} \leq \Big(\frac{1}{N}\sum_{n=1}^{N} a_n^q\Big)^{1/q} \leq \Big(\frac{1}{N}\sum_{n=1}^{N} a_n^p\Big)^{1/q} \|(a_n)\|_{\ell^\infty}^{1-p/q}.$$

It is clear that $E(X)$ and $A(X)$ are closed linear subspaces of L^2. It is also clear that $W(X)$ is closed in L^2, but the proof that it is a linear subspace requires the following lemma.

Lemma 4.4. *Let $f \in W(X)$ and $g \in L^2(X)$. Then for every $0 < p < \infty$ we have*

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

Proof. It suffices to show this for $p = 2$. In this case the left-hand side of the conclusion can be written

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

By the van der Corput differencing lemma it suffices to show

$$\limsup_{H\to\infty} \frac{1}{H}\sum_{h=1}^{H} \limsup_{n\to\infty} \Big| \frac{1}{N}\sum_{n=1}^{N} \Big\langle \overline{\langle T^{n+h} f, g\rangle} T^{n+h} f, \overline{\langle T^n f, g\rangle} T^n f \Big\rangle \Big| = 0.$$

By T-invariance of the inner product the left-hand side can be written as

$$\begin{aligned}
\limsup_{H\to\infty} \frac{1}{H}\sum_{h=1}^{H} \limsup_{N\to\infty} \Big| \frac{1}{N}\sum_{n=1}^{N} \Big\langle \langle T^n f, g\rangle\, \overline{\langle T^{n+h} f, g\rangle} T^h f, f \Big\rangle \Big| \\
\leq \|f\|^2 \|g\|^2 \limsup_{H\to\infty} \frac{1}{H}\sum_{h=1}^{H} \Big| \Big\langle T^h f, f \Big\rangle \Big| \\
= \|f\|^2 \|g\|^2 \limsup_{H\to\infty} \frac{1}{H}\sum_{h=1}^{H} \Big| \Big\langle T^h f, f \Big\rangle \Big|,
\end{aligned}$$

and this vanishes by the assumption. $\square$

Now we consider the relations between these spaces. In the remaining part of the lecture we will show

$$E(X) = A(X) = W(X)^\perp.$$

The inclusion $E(X) \subset A(X)$ is clear. The next two lemmas show the inclusions $A(X) \subset W(X)^\perp$ and $W(X)^\perp \cap E(X)^\perp = \{0\}$, from which the conclusion follows.

Lemma 4.5. *$W(X) \perp A(X)$.*

Proof. Let $f \in W(X)$, $g \in A(X)$, and $\varepsilon > 0$. By the assumption there exist $g_1, \dots, g_k \in L^2(X)$ such that for every n there exists $i(n)$ with $\|T^n g - g_{i(n)}\|_2 < \varepsilon$. It follows that

$$\begin{aligned} |\langle f, g\rangle| &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\langle T^n f, T^n g\rangle| \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\langle T^n f, g_{i(n)}\rangle| + O(\|f\|_2 \varepsilon) \\ &= \sum_{i=1}^{k} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\langle T^n f, g_i\rangle| + O(\|f\|_2 \varepsilon). \end{aligned}$$

By the assumption $f \in W(X)$ the limits in the last line vanish, so that $\langle f, g\rangle = O(\|f\|_2 \varepsilon)$. Since ε was arbitrary, this implies $\langle f, g\rangle = 0$ as claimed. □

Lemma 4.6. *Let $f \in L^2(X) \setminus W(X)$. Then $f \not\perp E(X)$.*

Proof. We need to construct an eigenfunction of T that correlates with f. By the hypothesis we know

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\langle T^n f, f\rangle|^2 \neq 0.$$

This can be written as

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle (T \times T)^n (f \otimes \bar{f}), f \otimes \bar{f}\rangle \neq 0,$$

the inner product now being taken in $L^2(X \times X, \mu \times \mu)$. By the mean ergodic theorem applied to this product space the left-hand side equals $\langle H, f \otimes \bar{f}\rangle$ with a (non-zero) function $H \in L^2(X \times X)$.

Consider the integral operator

$$Sg(x) := \int H(x, x') g(x') \mathrm{d}\mu(x).$$

The operator S is self-adjoint because $H(x, x') = \overline{H(x', x)}$. Moreover, it is a Hilbert–Schmidt operator, and in particular compact.

By the spectral theorem for compact operators there exists a finite-dimensional eigenspace $V \subset L^2(X)$ to a non-zero eigenvalue λ. Since the integral kernel H is T-invariant, the operator S commutes with T, so the space V is T-invariant. But T is unitary, so there exists a $0 \neq g \in V$ that is an eigenvalue of both S and T.

By construction we have

$$0 \neq \lambda \|g\|^2 = \langle Sg, g\rangle = \iint \overline{g(x)} H(x, x') g(x') \mathrm{d}\mu(x') \mathrm{d}\mu(x).$$

By definition of H it follows that there exists $n \in \mathbb{N}$ such that

$$0 \neq \iint \overline{g(x)} T^n f(x) T^n \bar{f}(x') g(x') \mathrm{d}\mu(x') \mathrm{d}\mu(x) = |\langle T^n f, g\rangle|^2 = |\langle f, T^{-n} g\rangle|^2.$$

Thus g is an eigenfunction of T with the required property. □

5 Roth's theorem

Theorem 5.1 (Roth, [Rot53]). *Let $E \subset \mathbb{Z}$ be a set with positive upper density. Then there exist $a \in \mathbb{Z}$ and $n > 0$ such that $a, a+n, a+2n \in E$.*

Roth's theorem has the following ergodic-theoretic formulation.

Theorem 5.2. *Let (X, μ, T) be an ergodic measure-preserving system and $f \in \mathcal{X}$ non-negative with $\int f > 0$. Then*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f \cdot T^n f \cdot T^{2n} f \mathrm{d}\mu > 0. \tag{5.3}$$

The proof consists of two steps: reduction to the Kronecker factor and an application of Weyl's equidistribution theorem to eigenvalues of T.

Lemma 5.4. *Let (X, μ, T) be an ergodic mps, $f_0, f_1, f_2 \in \mathcal{X}$ and suppose that $f_i \in W(X)$ for some $i \in \{0, 1, 2\}$. Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0 \cdot T^n f_1 \cdot T^{2n} f_2 \mathrm{d}\mu = 0.$$

Proof. Suppose first either $f_1 \in W(X)$ or $f_2 \in W(X)$. By the van der Corput differencing lemma applied to the vectors $u_n = T^n f_1 \cdot T^{2n} f_2$ it suffices to show

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

The left-hand side equals

$$\begin{aligned}
\limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \limsup_{N\to\infty} \Big| \frac{1}{N} \sum_{n=1}^{N} \int T^n f_1 T^{2n} f_2 T^{n+h} \bar{f}_1 T^{2n+2h} \bar{f}_2 \Big| \\
= \limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \limsup_{N\to\infty} \Big| \int \frac{1}{N} \sum_{n=1}^{N} f_1 T^n f_2 T^h \bar{f}_1 T^{n+2h} \bar{f}_2 \Big|.
\end{aligned}$$

By the mean ergodic theorem applied to the average over n this equals

$$\limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \Big| \int f_1 T^h \bar{f}_1 \mathbb{E}(f_2 T^{2h} \bar{f}_2 | \mathcal{I}) \Big|.$$

By the ergodicity assumption the conditional expectation onto the invariant factor equals the integral of the function, so this equals

$$\limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \Big| \int f_1 T^h \bar{f}_1 \Big| \cdot \Big| \int f_2 T^{2h} \bar{f}_2 \Big|.$$

By Cauchy–Schwarz in the summation over h this is bounded by

$$\limsup_{H\to\infty} \Big(\frac{1}{H} \sum_{h=1}^{H} |\big\langle f_1, T^h f_1 \big\rangle|^2 \Big)^{1/2} \Big(\frac{1}{H} \sum_{h=1}^{H} |\big\langle f_2, T^{2h} f_2 \big\rangle|^2 \Big)^{1/2}.$$

By the assumption one of the factors goes to 0 as $H \to \infty$, while the other is certainly bounded.

It remains to consider the case $f_0 \in W(X)$. In this case use the fact that T is a homomorphism to write

$$\frac{1}{N} \sum_{n=1}^{N} \int f_0 \cdot T^n f_1 \cdot T^{2n} f_2 \mathrm{d}\mu = \frac{1}{N} \sum_{n=1}^{N} \int (T^{-1})^{2n} f_0 \cdot (T^{-1})^n f_1 \cdot f_2 \mathrm{d}\mu$$

and apply the above reasoning, noting that $W(X)$ does not change upon replacing T by T^{-1}. □

By multilinearity (splitting $f_i = \mathbb{E}(f_i|\mathcal{E}) + f_i^\perp$ with $f_i^\perp \in W(X)$) it follows that

$$\frac{1}{N}\sum_{n=1}^{N}\int f_0 \cdot T^n f_1 \cdot T^{2n} f_2 \mathrm{d}\mu - \frac{1}{N}\sum_{n=1}^{N}\int \mathbb{E}(f_0|\mathcal{E}) \cdot T^n \mathbb{E}(f_1|\mathcal{E}) \cdot T^{2n}\mathbb{E}(f_2|\mathcal{E})\mathrm{d}\mu \to 0 \quad (5.5)$$

as $N \to \infty$ for any functions $f_0, f_1, f_2 \in \mathcal{X}$. The property (5.5) is described by saying that the Kronecker factor is *characteristic* for the ergodic averages (5.3). There are also other characterisitc factors, for instance $\mathcal{X}$ itself is characteristic for any kind of ergodic averages. The point here is that the Kronecker factor has an explicit algebraic description.

Proof of Theorem 5.2. Note that $\int \mathbb{E}(f|\mathcal{E}) = \int f > 0$, so by Lemma 5.4 we may assume $f \in E(X) \cap \mathcal{X}$. This means that f can be approximated in L^2 by finite linear combinations of eigenfunctions of T. Let $\varepsilon > 0$ and write

$$f = \sum_{i=1}^{r} a_i f_{\lambda_i} + O(\varepsilon)$$

accordingly, where λ_i are distinct eigenvalues of T and f_{λ_i} are corresponding (orthogonal) L^2 normalized eigenfunctions. Then

$$T^n f = \sum_{i=1}^{r} \lambda_i^n a_i f_{\lambda_i} + O(\varepsilon).$$

By Weyl's equidistribution theorem the sequence $((\lambda_i^n)_i)_n$ is equidistributed in a subgroup H of the torus Λ^r. Let $\varphi \in C(\Lambda^r)$ be a non-zero positive function supported in a δ-neighborhood of the identity and bounded by 1. Then $\varphi((\lambda_i^n)_i) \neq 0$ implies

$$T^n f = \sum_{i=1}^{r}(1 + O(\delta))a_i f_{\lambda_i} + O(\varepsilon) = f + O(\varepsilon) + O(\delta).$$

It follows that

$$\begin{aligned}\int f \cdot T^n f \cdot T^{2n} f \mathrm{d}\mu &\geq \varphi((\lambda_i^n)_i) \int f \cdot T^n f \cdot T^{2n} f \mathrm{d}\mu \\ &= \varphi((\lambda_i^n)_i) \int f \cdot (f + O(\varepsilon) + O(\delta)) \cdot (f + O(\varepsilon) + O(\delta))\mathrm{d}\mu \\ &= \varphi((\lambda_i^n)_i)\Big(\int f^3 \mathrm{d}\mu + O(\varepsilon) + O(\delta)\Big)\end{aligned}$$

If both ε and δ are small enough, this is

$$\geq \varphi((\lambda_i^n)_i)c, \qquad \text{where } c = \frac{1}{2}\int f^3 \mathrm{d}\mu > 0.$$

Averaging in n, taking the limit, and using equidistribution we obtain the lower bound $c\int_H \varphi > 0$ for (5.3). □

5.1 Uniformity seminorms

Let us now introduce higher order analogues of the weakly mixing space W, which are going to have a properties analogous to Lemma 5.4 for "longer" ergodic averages.

Let (X, μ, T) be an mps. We introduce the following sequence of functionals on $L^\infty(X, \mu, T)$:

$$\|f\|_{[0],X,\mu,T} := \int_X f \mathrm{d}\mu, \qquad \|f\|_{[k+1],X,\mu,T}^{2^{k+1}} := \limsup_{N\to\infty} \Big|\frac{1}{N}\sum_{n=1}^{N} \|f\overline{T^n f}\|_{[k]}^{2^k}\Big|.$$

We will usually omit the subscripts X, μ, T if they are clear from the context. The functional $\|\cdot\|_{[k]}$, $k > 1$, is called the *k-th uniformity seminorm* (or Gowers–Host–Kra seminorm). Other common notations in the literature include $\|\cdot\|_{[k]} = |||\cdot|||_k = \|\cdot\|_{U^k(X,\mu,T)}$. Bibliographical remark: currently the only reference for the structural theory of these seminorms is the original article of Host and Kra [HK05]. A more axiomatic treatment of the surrounding issues is being prepared by Gutman, Manners, and Varjú.

At this point subadditivity of the uniformity seminorms is not clear; we shall prove it when a diffirent characterization becomes available. We shall also see that the limit in the above definition actually exists. Moreover, note that the absolute value in the definition of the $k+1$-th seminorm, which we included to make the lim sup a priori well-defined, is unnecessary: for $k > 0$ this is clear because the previous seminorm is already positive. For $k = 0$ note

$$\frac{1}{N}\sum_{n=1}^{N} \|f\overline{T^n f}\|_{[0]} = \frac{1}{N}\sum_{n=1}^{N} \int f\overline{T^n f}\mathrm{d}\mu \to \int f\overline{\mathbb{E}(f|\mathcal{I})}\mathrm{d}\mu = \|\mathbb{E}(f|\mathcal{I})\|_2^2 > 0$$

by the mean ergodic theorem, so $\|f\|_{[1]} = \|\mathbb{E}(f|\mathcal{I})\|_2$. This shows in particular that $\|f\|_{[1]} = 0 \iff f \perp \mathcal{I}$.

The uniformity seminorm of order 2 recovers the weakly mixing space, but only for ergodic systems. Indeed, assume that X is ergodic, then the projection onto the invariant factor equals the integral of a function, and by the above calculation we obtain

$$\|f\|_{[1]} = \Big|\int f\mathrm{d}\mu\Big|.$$

Hence, by definition,

$$\begin{aligned}
\|f\|_{[2]}^4 &= \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \|f\overline{T^n f}\|_{[1]}^2 \\
&= \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \Big|\int f\overline{T^n f}\mathrm{d}\mu\Big|^2 \\
&= \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} |\langle f, T^n f\rangle|^2,
\end{aligned}$$

and the right-hand side provides one of the equivalent descriptions of $W(X)$.

Next we will show that uniformity seminorms control ergodic averages.

Lemma 5.6. *Let $f_1, \dots, f_k \in \mathcal{X}$. Then*

$$\limsup_{N\to\infty} \Big\| \frac{1}{N}\sum_{n=1}^{N}\prod_{j=1}^{k} T^{jn} f_j \Big\|_2 \lesssim_k \min_l \|f_l\|_{[k]} \prod_{j\neq l} \|f_j\|_{2^k}.$$

Here and later $A \lesssim_k B$ means $A \le C_k B$ with a constant C_k that depends only on k.

Proof. By induction on k. In the case $k = 1$ we have

$$\frac{1}{N}\sum_{n=1}^{N}\prod_{j=1}^{k} T^{jn} f_j = \frac{1}{N}\sum_{n=1}^{N} T^n f_1 \to \mathbb{E}(f_1|\mathcal{I})$$

in L^2, and the conclusion follows since, as observed above, $\|f_1\|_{[1]} = \|\mathbb{E}(f_1|\mathcal{I})\|_2$.

Suppose now that the conclusion is known for some $k \geq 1$. Applying the van der Corput differencing lemma with $v_n = \prod_{j=1}^{k+1} T^{jn} f_j$, we see that it suffices to obtain the estimate

$$\limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \limsup_{N\to\infty} \Big| \frac{1}{N} \sum_{n=1}^{N} \langle v_n, v_{n+h} \rangle \Big| \lesssim \min_l \|f_l\|_{[k+1]}^2 \prod_{j\neq l} \|f_j\|_{2^{k+1}}^2.$$

The left-hand side can be written

$$\limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \limsup_{N\to\infty} \Big| \frac{1}{N} \sum_{n=1}^{N} \int \prod_{j=1}^{k+1} T^{jn} f_j \overline{T^{j(n+h)} f_j} \mathrm{d}\mu \Big|.$$

Suppose first that the minimum is assumed for some $l \geq 2$. Then, by T-inavriance of μ, this can be written as

$$\limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \limsup_{N\to\infty} \Big| \frac{1}{N} \sum_{n=1}^{N} \int f_0 \overline{T^h f_0} \prod_{j=2}^{k+1} T^{(j-1)n} (f_j \overline{T^{jh} f_j}) \mathrm{d}\mu \Big|.$$

By the Cauchy–Schwarz inequality this is bounded by

$$\|f_0 \overline{T^h f_0}\|_2 \limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \limsup_{N\to\infty} \Big\| \frac{1}{N} \sum_{n=1}^{N} \prod_{j=2}^{k+1} T^{(j-1)n} (f_j \overline{T^{jh} f_j}) \Big\|_2,$$

and by the inductive hypothesis this is bounded by

$$\|f_0\|_4^2 \limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \|f_l \overline{T^{lh} f_l}\|_{[k]} \prod_{j\geq 2, j\neq l} \|f_j \overline{T^{jh} f_j}\|_{2^k},$$

and this is bounded by

$$\prod_{j\neq l} \|f_j\|_{2^{k+1}}^2 \cdot \limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \|f_l \overline{T^{lh} f_l}\|_{[k]}.$$

By positivity of the k-th uniformity seminorm the latter $\limsup$ is bounded by

$$l \limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \|f_l \overline{T^h f_l}\|_{[k]},$$

and by Jensen's inequality in the average over h this is bounded by

$$\Big(\limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \|f_l \overline{T^h f_l}\|_{[k]}^{2^k} \Big)^{2^{-k}} = \|f_l\|_{[k+1]}^2,$$

as required.

In the case $l = 1$ we can write the expression obtained from the van der Corput differencing lemma as

$$\limsup_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \limsup_{N\to\infty} \Big| \frac{1}{N} \sum_{n=1}^{N} \int f_{k+1} \overline{T^{(k+1)h} f_{k+1}} \prod_{j=1}^{k} T^{(j-k-1)n} (f_j \overline{T^{jh} f_j}) \mathrm{d}\mu \Big|,$$

and use the same argument as before with T replaced by T^{-1} (note that the uniformity seminorms $\|\cdot\|_{[k],T}$ and $\|\cdot\|_{[k],T^{-1}}$ coincide). $\square$

6 Cube spaces

6.1 Joinings

Definition 6.1. A *joining* of measure-preserving systems (Y_i, μ_i, T_i), $i = 1, \dots, r$, is a measure-preserving system (X, μ, T), where $X = Y_1 \times \cdots \times Y_r$ (the product of topological model spaces), $T = T_1 \times \cdots \times T_r$, and the marginal of μ on each Y_i equals μ_i.

Example. The product measure $\mu = \mu_1 \times \cdots \times \mu_r$ defines a joining for any tuple of systems. This joining is called the (cartesian) product.

Example. Suppose $Y_1 = \cdots = Y_r$. Then the diagonal measure

$$\int_X F(y_1, \dots, y_r) \mathrm{d}\mu(y_1, \dots, y_r) = \int_{Y_1} F(y, \dots, y) \mathrm{d}\mu_1(y)$$

defines a joining.

Let Y_i, $i = 1, \dots, r$, be measure-preserving systems and $\pi_i : Y_i \to Z_i$ factors. Then any joining (X, μ, T) of the Y_i's restricts to a joining $(\tilde{X}, \tilde{\mu}, T)$ of the Z_i's with the measure

$$\tilde{\mu} = (\pi_1 \times \cdots \times \pi_r)_* \mu.$$

We write $\tilde{X} = Z_1 \vee \cdots \vee Z_r$ if the ambient joining X is understood.

Any joining of the Z_i's admits at least one extension to a joining of the Y_i's.

Definition 6.2. Let $\pi_i : Y_i \to Z_i$, $i = 1, \dots, r$, be factors, and $\tilde{X}$ a joining of the Z_i's. The *relatively independent joining* of Y_i's over $\tilde{X}$ is defined by the measure

$$\int_{\tilde{X}} \mu_{1,z_1} \times \cdots \times \mu_{r,z_r} \mathrm{d}\tilde{\mu}(z_1, \dots, z_r),$$

where $\mu_i = \int_{Z_i} \mu_{i,z} \mathrm{d}\mu(z)$ are the disintegrations of the measures on Y_i over Z_i.

It is not hard to verify that the relatively independent joining is in fact a joining. It is the unique joining that satisfies

$$\int_X f_1(y_1) \cdots f_r(y_r) \mathrm{d}\mu(y_1, \dots, y_r) = \int_{\tilde{X}} \mathbb{E}(f_1|Z_1)(z_1) \cdots \mathbb{E}(f_r|Z_r)(z_r) \mathrm{d}\tilde{\mu}(z_1, \dots, z_r)$$

for all $f_i \in \mathcal{Y}_i$.

An important special case occurs when $Z_1 = \cdots = Z_r$.

Definition 6.3. Let Y_i, $i = 1, \dots, r$ be measure-preserving systems that share a common factor Z. The *relatively independent joining of Y_i's over Z*, denoted by $Y_1 \times_Z \cdots \times_Z Y_r$, is the relatively independent joining of Y_i's over the diagonal joining of the common factors Z.

In other words, it is given by the measure

$$\mu_1 \times_Z \cdots \times_Z \mu_r = \int_Z \mu_{1,z} \times \cdots \times \mu_{r,z} \mathrm{d}\mu(z),$$

where $\mu_i = \int_Z \mu_{i,z} \mathrm{d}\mu(z)$ are the respective disintegrations.

The relatively independent joining over the common factor Z is the unique joining with the property

$$\int_X f_1(y_1) \cdots f_r(y_r) \mathrm{d}\mu = \int_Z \mathbb{E}(f_1|Z) \cdots \mathbb{E}(f_r|Z) \mathrm{d}\mu.$$

6.2 Cube spaces

Definition 6.4. Let (X, μ, T) be a measure-preserving system. We define a sequence of measure-preserving systems $X^{[k]} = (X^{[k]}, \mu^{[k]}, T^{[k]})$ inductively starting with $X^{[0]} = X$. Once $X^{[k]}$ is defined, let $\mathcal{I}^{[k]}$ be the invariant factor of $X^{[k]}$ and set

$$X^{[k+1]} := X^{[k]} \times_{\mathcal{I}^{[k]}} X^{[k]}.$$

The measure $\mu^{[k]}$ is called the *cube measure* and the measure space $(X^{[k]}, \mu^{[k]})$ the *cube space* of the mps (X, μ, T). The mps $X^{[k]}$ is a joining of 2^k copies of X, which will be indexed by the cube $V_k = \{0,1\}^k$ in such a way that

$$X^{[k+1]} = X^{V_{k+1}} = X^{V_k \times \{0\}} \times X^{V_k \times \{1\}} = X^{[k]} \times X^{[k]}.$$

We write[3] points of $X^{[k+1]}$ as $\vec{x} = (x_\varepsilon)_{\varepsilon \in V_k}$ and coordinate projections as $\pi_\varepsilon \vec{x} = x_\varepsilon$.

Cube spaces and disintegrations

Lemma 6.5. *Let X be a compact metric space, $T : X \to X$ a homeomorphism, Ω a probability space, and $\omega \mapsto \mu_\omega$ a measurable map from Ω to the space of T-invariant regular probability measures on X. Then for every k we have*

$$\int_\Omega \mu_\omega^{[k]} \mathrm{d}\omega = \mu^{[k]}, \quad \textit{where} \quad \mu = \int_\Omega \mu_\omega \mathrm{d}\omega.$$

Proof. It suffices to consider $k = 1$, all other cases follow from the identity $\mu^{[k+1]} = (\mu^{[k]})^{[1]}$. It suffices to test both measures in the conclusion of the lemma on tensor products $f_0 \otimes f_1$ with $f_0, f_1 \in C(X)$. We have

$$\int_{X^2} f_0 \otimes f_1 \mathrm{d}\mu_\omega^{[1]} = \int_{X^2} f_0 \otimes f_1 \mathrm{d}(\mu_\omega \times_{\mathcal{I}} \mu_\omega) = \int_X \mathbb{E}(f_0 | \mathcal{I}(X, \mu_\omega)) \mathbb{E}(f_0 | \mathcal{I}(X, \mu_\omega)) \mathrm{d}\mu_\omega.$$

By the mean ergodic theorem the right-hand side equals

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X f_0 T^n f_1 \mathrm{d}\mu_\omega.$$

Integrating over Ω and using the dominated convergence theorem we obtain

$$\int_\Omega \int_{X^2} f_0 \otimes f_1 \mathrm{d}\mu_\omega^{[1]} \mathrm{d}\omega = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_\Omega \int_X f_0 T^n f_1 \mathrm{d}\mu_\omega \mathrm{d}\omega.$$

By the mean ergodic theorem on the system (X, μ, T) this equals

$$\int_X \mathbb{E}(f_0 | \mathcal{I}(X, \mu)) \mathbb{E}(f_0 | \mathcal{I}(X, \mu)) \mathrm{d}\mu,$$

and this by definition is $\int_{X^2} f_0 \otimes f_1 \mathrm{d}\mu^{[1]}$. □

6.3 The space $X^{[2]}$

Let X be an ergodic mps. Then $X^{[1]}$ is just the cartesian product of two copies of X. In order to construct $X^{[2]}$ we need a description of the invariant factor of $X^{[1]}$. We begin with the following seminorm estimate.

[3]It would be nice to denote elements of V_k by v, but ε is the usual convention in this topic. I also like to write $V_k = 2^k$ because $2 = \{0, 1\}$, but this is probably even more confusing.

Lemma 6.6. *Let (X,μ,T) and (Y,ν,S) be measure-preserving systems. Then for every $k\geq 0$ and every $f\in\mathcal{X}$, $g\in\mathcal{Y}$ we have*

$$|\|f\otimes g\|_{[k],X\times Y}|\leq\|f\|_{[k+1],X}\|g\|_{[k+1],Y}.$$

Proof. We induct on k. For $k=0$ we can use the explicit description of both sides:

$$|\|f\otimes g\|_{[k],X\times Y}|=|\int_X f|\cdot|\int_Y g|\leq\|\mathbb{E}(f|\mathcal{I}(X))\|_2\|\mathbb{E}(g|\mathcal{I}(Y))\|_2=\|f\|_{[k+1],X}\|g\|_{[k+1],Y}.$$

Suppose now that the claim is known for some k and let us prove it for $k+1$. On the left-hand side we have

$$\begin{aligned}\|f\otimes g\|_{[k+1],X\times Y}^{2^{k+1}}&=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\|f\otimes g\cdot(T^n\bar{f}\otimes S^n\bar{g})\|_{[k],X\times Y}^{2^k}\\&\leq\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\|fT^n\bar{f}\|_{[k+1],X}^{2^k}\|gS^n\bar{g}\|_{[k+1],Y}^{2^k}.\end{aligned}$$

By Cauchy–Schwarz in the summation over n this is bounded by

$$\limsup_{N\to\infty}\Big(\frac{1}{N}\sum_{n=1}^{N}\|fT^n\bar{f}\|_{[k+1],X}^{2^{k+1}}\Big)^{1/2}\Big(\frac{1}{N}\sum_{n=1}^{N}\|gS^n\bar{g}\|_{[k+1],Y}^{2^{k+1}}\Big)^{1/2}=\|f\|_{[k+2],X}^{2^{k+1}}\|g\|_{[k+2],Y}^{2^{k+1}},$$

as required. □

An immediate corollary is that the sequence of uniformity seminorms increases monotonically (take Y to be the trivial system and $g=1$ in the above lemma).

In the case $k=1$ we know that $\|f\|_{[2]}=0$ if and only if f is orthogonal to the Kronecker factor. Hence, whenever $f\perp\mathcal{K}(X)$ or $g\perp\mathcal{K}(Y)$, we have $\|f\otimes g\|_{[1]}=0$, which means that $f\otimes g\perp\mathcal{I}(X\times Y)$. In other words, the invariant factor $\mathcal{I}(X\times Y)$ is contained in the joining of Kronecker factors $\mathcal{K}(X)\vee\mathcal{K}(Y)$.

Now return to the case $X=Y$, X ergodic, and let us compute the invariant factor of the square of the Kronecker factor. Recall that by the Halmos–von Neumann theorem the Kronecker factor has a topogical model that is a rotation on a compact monothetic group (Z,γ) (this group is commutative and the group operation will be written additively). On the product space $Z\times Z$ the diagonally invariant functions (i.e. functions constant on each coset of the diagonal subgroup $\{(z,z),z\in G\}$) are certainly invariant under translation by (γ,γ). On the other hand, the translation by (γ,γ) is ergodic with respect to the Haar measure on every coset of the diagonal group, because it is isomorphic to (Z,γ). Hence any invariant function is almost everywhere constant on almost every coset of the diagonal subgroup. It follows that the invariant factor of $X\times X$ consists of the diagonally invariant functions on the product of Kronecker factors:

$$\mathcal{I}(X\times X)=\{f(z_1-z_2)\}.$$

We use this information to give an explicit formula for the cube measure $\mu^{[2]}$. By definition

$$\int_{X^{[2]}}f_{00}\otimes f_{10}\otimes f_{01}\otimes f_{11}\mathrm{d}\mu^{[2]}=\int_{X^{[1]}}\mathbb{E}(f_{00}\otimes f_{10}|\mathcal{I}^{[1]})\mathbb{E}(f_{01}\otimes f_{11}|\mathcal{I}^{[1]})\mathrm{d}\mu^{[2]}.$$

Since $\mathcal{I}^{[1]}\subset\mathcal{K}(X)\times\mathcal{K}(X)$, we may replace the functions f_ε by their respective projections onto the Kronecker factor $\tilde{f}_\varepsilon$. The projection onto the invariant factor then equals the average along the cosets of the diagonal subgroup, so we obtain

$$\int_{Z^2}(\int_Z\tilde{f}_{00}(z_1+z_3)\tilde{f}_{10}(z_2+z_3)\mathrm{d}z_3)(\int_Z\tilde{f}_{01}(z_1+z_4)\tilde{f}_{11}(z_2+z_4)\mathrm{d}z_4)\mathrm{d}z_1\mathrm{d}z_2.$$

It is unsurprising that this is symmetric under exchanging f_{i0} and f_{i1}. Note however that this is also symmetric under exchanging f_{ij} and f_{ij}. The higher oder measures $\mu^{[k]}$ also have similar symmetries.

6.4 Symmetries of the cube measures

Let α be a permutation of the cube V_k and let $\alpha_*\mu^{[k]}$ be the pushforward of $\mu^{[k]}$ under the coordinate permutation map $(x_\varepsilon) \mapsto (x_{\alpha(\varepsilon)})$. We are interested in determining those α leaving the cube measure invariant: $\alpha_*\mu^{[k]} = \mu^{[k]}$. It is clear from definition that this holds for the *reflection* in the last coordinate $\alpha(\varepsilon', j) = (\varepsilon', 1-j)$.

We will now prove that the *digit permutations* $\alpha(\varepsilon_1, \dots, \varepsilon_k) = (\varepsilon_{\sigma(1)}, \dots, \varepsilon_{\sigma(k)})$, where σ is a permutation on $\{1, \dots, k\}$, also leave $\mu^{[k]}$ invariant (we write $\sigma_*\mu^{[k]} = \alpha_*\mu^{[k]}$ in this case). For $k=1$ there is nothing to show, and for $k=2$ the claim has been verified above for ergodic systems X, and for non-ergodic systems it follows from Lemma 6.5.

Suppose that the claim is known for some $k \geq 2$. The group of permutations of $\{1, \dots, k+1\}$ is spanned by the permutations that leave $k+1$ invariant and the transposition $(k, k+1)$, which we consider separately. Let σ be a permutation of $\{1, \dots, k\}$. Then by construction of $\mu^{[k+1]}$ we have

$$\sigma_*\mu^{[k+1]} = (\sigma_*\mu^{[k]})^{[1]},$$

and the claim follows by the inductive hypothesis. On the other hand, for the permutation $\sigma = (k, k+1)$ we have

$$\sigma_*\mu^{[k+1]} = \sigma_*(\mu^{[k-1]})^{[2]},$$

and the claim follows from the case $k=2$.

It follows that cube measures are invariant under the group of symmetries generated by digit permutations and reflections in the last coordinate, which includes also reflections in any other coordinates.

Remark. The last section contains the original proof by Host and Kra that cube measures are invariant under digit permutations and reflections. From the current point of view this fact can also be seen as an easy consequence of norm convergence of multiple ergodic averages associated to commuting actions of $\mathbb{Z}^k$, first proved in [Aus10].

7 Host–Kra–Ziegler factors

Uniformity seminorms can be written in terms of cube spaces:

$$\|f\|_{[k]}^{2^k} = \int_{X^{[k]}} \prod_{\varepsilon \in V_k} C^{|\varepsilon|} f(x_\varepsilon) \mathrm{d}\mu^{[k]}(\vec{x}), \tag{7.1}$$

where C denotes complex conjugation annd $|\varepsilon|$ is the number of 1's in ε. Indeed, for $k=0$ this is immediate. Suppose this is known for some k and consider the $k+1$-th uniformity seminorm. By definition

$$\|f\|_{U^{k+1}}^{2^{k+1}} = \limsup_{N\to\infty} \Big| \frac{1}{N} \sum_{n=1}^{N} \|f T^n C f\|_{U^k}^{2^k} \Big|.$$

By the inductive hypothesis this equals

$$\limsup_{N\to\infty} \Big| \frac{1}{N} \sum_{n=1}^{N} \int_{X^{[k]}} \prod_{\varepsilon \in V_k} C^{|\varepsilon|} f(x_\varepsilon) \cdot C \prod_{\varepsilon \in V_k} C^{|\varepsilon|} T^n f(x_\varepsilon) \mathrm{d}\mu^{[k]} \Big|.$$

By the mean ergodic theorem on the system $X^{[k]}$ the average inside the absolute value converges to

$$\int_{X^{[k]}} \prod_{\varepsilon \in V_k} C^{|\varepsilon|} f \circ \pi_\varepsilon \cdot \mathbb{E}(C \prod_{\varepsilon \in V_k} C^{|\varepsilon|} f \circ \pi_\varepsilon | \mathcal{I}^{[k]}) \mathrm{d}\mu^{[k]},$$

and this gives the claim. This argument shows in particular that the limit superior in the definition of uniformity seminorms is in fact a limit.

7.1 Cauchy–Schwarz–Gowers inequality

The 2^k-linear form

$$(f_\varepsilon)_{\varepsilon \in V_k} \mapsto \int_{X^{[k]}} \prod_{\varepsilon \in V_k} C^{|\varepsilon|} f_\varepsilon(x_\varepsilon) \mathrm{d}\mu^{[k]}(\vec{x})$$

cab be thought of as an "inner product". Indeed, in the case $k=1$, X ergodic, this is just the inner product in $L^2(X)$. In this case the triangle inequality for the L^2 norm follows from the Cauchy–Schwarz inequality for the inner product. Similarly, the triangle inequality for the uniformity seminorms follows from a multilinear version of the Cauchy–Schwarz inequality.

Proposition 7.2. *Let X be an mps annd $k \geq 1$. Then*

$$\Big| \int_{X^{[k]}} \prod_{\varepsilon \in V_k} C^{|\varepsilon|} f_\varepsilon(x_\varepsilon) \mathrm{d}\mu^{[k]}(\vec{x}) \Big| \leq \prod_{\varepsilon \in V_k} \|f_\varepsilon\|_{[k]}$$

Proof. By definition $\mu^{[k]} = \mu^{[k-1]} \times_{\mathcal{I}^{[k-1]}} \mu^{[k-1]}$. Hence

$$\begin{aligned} &\int_{X^{[k]}} \prod_{\varepsilon \in V_k} C^{|\varepsilon|} f_\varepsilon(x_\varepsilon) \mathrm{d}\mu^{[k]}(\vec{x}) \\ &= \int_{X^{[k-1]}} \mathbb{E}(\prod_{\varepsilon' \in V_{k-1}} C^{|\varepsilon'|} f_{\varepsilon' 0} \circ \pi_{\varepsilon'} | \mathcal{I}^{[k-1]}) \mathbb{E}(\prod_{\varepsilon' \in V_{k-1}} C^{|\varepsilon'|+1} f_{\varepsilon' 1} \circ \pi_{\varepsilon'} | \mathcal{I}^{[k-1]}) \mathrm{d}\mu^{[k-1]}. \end{aligned}$$

Applying the usual Cauchy–Schwarz inequality in the integral over $X^{[k-1]}$ we reduce to the case $f_{\varepsilon' 0} = f_{\varepsilon' 1}$ for all $\varepsilon' \in V_{k-1}$, that is, f_ε does not depend on the last digit of ε. Recalling the permutation symmetry of $\mu^{[k]}$ we may as well assume that f_ε does not depend on the first digit of ε.

Repeating the above argument k times we reduce to the case when f_ε does not depend on any digit of ε. But then the left-hand side and the right-hand side of the claim coincide. □

Corollary 7.3. *The uniformity seminorms satisfy the triangle inequality*

$$\|f+g\|_{[k]} \leq \|f\|_{[k]} + \|g\|_{[k]}, \quad k \geq 1.$$

Proof. Use the expression (7.1) for $\|f+g\|_{[k]}^{2^k}$ and expand anto 2^k terms by multilinearity. Estimate each of the terms using Proposition 7.2 and notice that the estimates sum precisely to $(\|f\|_{[k]} + \|g\|_{[k]})^{2^k}$. □

7.2 Characteristic factors

We have seen that uniformity seminorms control multilinear ergodic averages, and now we also know that the space N_k of functions with zero k-th uniformity seminorm is linear. Thus it suffices to consider ergodic averages on the orthogonal complement of N_k. This orthogonal complement turns out to describe a factor, which we will now construct.

Definition 7.4. Let X be an mps. The factor $\mathcal{Z}_k$ of X consists of those functions $f \in \mathcal{X}$ for which the function $f \circ \pi_{\vec{0}}$ on $X^{[k+1]}$ coincides almost everywhere with a function in $\vee_{\varepsilon \in V_{k+1} \setminus \{\vec{0}\}} \mathcal{X} \circ \pi_\varepsilon$, that is, a function that does not depend on the variable $x_{\vec{0}}$.

It is counterintuitive to speak of functions of the variable $x_{\vec{0}}$ that coincide with functions that does not depend on $x_{\vec{0}}$, and indeed, the only such functions are the constants. However, here we are talking about equality almost everywhere, which changes things. Imagine for instance the diagonal joining of two copies of a measure space. Then every function of the first coordinate coincides almost everywhere with a function of only the second coordinate (just take the same function).

The objective is now to obtain the orthogonal splitting $L^2(X) = \overline{\mathcal{Z}_k} + \overline{N_{k+1}}$. This follows from the equivalence $f \in N_{k+1} \iff f \perp \mathcal{Z}_k$. We will prove this in the contrapositive form $\|f\|_{[k+1]} \neq 0 \iff f \not\perp \mathcal{Z}_k$. The direction $\Longleftarrow$ is not hard:

Lemma 7.5. $\|f\|_{[k+1]} \neq 0 \Longleftarrow f \not\perp \mathcal{Z}_k$

Proof. By the hypothesis there exists a function in $\mathcal{Z}_k$ that correlates with f. By definition of $\mathcal{Z}_k$ this means

$$\int_{X^{[k+1]}} (f \circ \pi_{\vec{0}}) F \mathrm{d}\mu^{[k+1]} \neq 0$$

for some function F that does not depend on the $\vec{0}$-th coordinate. Approximating F by tensor products it follows that

$$\int_{X^{[k+1]}} (f \circ \pi_{\vec{0}}) \prod_{\varepsilon \in V_{k+1} \setminus \{0\}} f_\varepsilon \circ \pi_\varepsilon \mathrm{d}\mu^{[k+1]} \neq 0$$

for some functions f_ε. The conclusion follows from the Cauchy–Schwarz-Gowers inequality. □

The converse direction is slgihtly more difficult and requires some additional information about the measures $\mu^{[k]}$. A *side* of the cube V_k is a set of the form $\alpha = \{\varepsilon : \varepsilon_i = j\}$ with $i \in \{1, \dots, k\}$ and $j \in \{0, 1\}$. A *side transformation* is a map on $X^{[k]}$ of the form

$$(T_\alpha \vec{x})_{\varepsilon \in V_k} = \begin{cases} T x_\varepsilon, & \varepsilon \in \alpha, \\ x_\varepsilon, & \varepsilon \notin \alpha. \end{cases}$$

The side transformations preserve the measure $\mu^{[k]}$. Indeed, by the previously established symmetries of $\mu^{[k]}$ it suffices to establish this for the side $\alpha = \{\varepsilon : \varepsilon_k = 1\}$.

In this case we have

$$\int_{X^{[k]}} \otimes_{\varepsilon'\in V_{k-1}} f_{\varepsilon' 0} \otimes \otimes_{\varepsilon'\in V_{k-1}} T f_{\varepsilon' 1} \mathrm{d}\mu^{[k]}$$
$$= \int_{X^{[k]}} \mathbb{E}(\otimes_{\varepsilon'\in V_{k-1}} f_{\varepsilon' 0} | \mathcal{I}^{[k-1]}) \mathbb{E}(\otimes_{\varepsilon'\in V_{k-1}} T f_{\varepsilon' 1} | \mathcal{I}^{[k-1]}) \mathrm{d}\mu^{[k]},$$

and the claim follows by since $T^{[k-1]}$ is the identity on $\mathcal{I}^{[k-1]}$.

Let $\mathcal{J}^{[k]}$ denote the factor of $X^{[k]}$ consisting of the functions invariant under all side transformations for the sides not containing $\vec{0} \in V_k$. This algebra is indeed a factor: invariance follows from the fact that all side transformations commute with $T^{[k]}$.

Lemma 7.6. *The factor $\mathcal{J}^{[k]}$ coincides with the algebra of functions depending only on the $\vec{0}$-th variable.*

Proof. It is clear that any function depending only on the $\vec{0}$-th variable is invariant under side transformations that act trivially on the $\vec{0}$-th variable.

The converse is proved by induction on k. In the case $k = 0$ there is nothing to prove. Suppose now $F \in \mathcal{J}^{[k+1]}$. Consider the side $\alpha = \{\varepsilon \in V_{k+1} : \varepsilon_{k+1} = 1\}$. Let

$$\mu^{[k]} = \int_{\mathcal{I}^{[k]}} \mu_\omega \mathrm{d}\omega$$

be ergodic decomposition of the measure $\mu^{[k]}$. Then

$$\mu^{[k+1]} = \int_\Omega \int_{X^{[k]}} \delta_x \times \mu_\omega \mathrm{d}\mu_\omega(x) \mathrm{d}\omega = \int_{X^{[k]}} \delta_x \times \mu_{\pi(x)} \mathrm{d}\mu^{[k]}(x)$$

(in the last line we have used that $\mu_{\omega(x)} = \mu_{\omega_0}$ holds for μ_{ω_0}-a.e. x), and this is in fact the ergodic decomposition with respect to the side transformation T_α. Hence F is $\delta_x \times \mu_\omega$-a.e. constant for $\mu^{[k]}$-a.e. $x \in X^{[k]}$, and it follows that F is a.e. equal to a function that does not depend on the coordinates in α.

But then F comes from a function in $\mathcal{J}^{[k]}$, and we can use the inductive hypothesis. □

With this knowledge at hand, we are ready to prove the implication $\implies$ mentioned above.

Lemma 7.7. $\|f\|_{[k+1]} \neq 0 \implies f \not\perp \mathcal{Z}_k$.

Proof. Let

$$F := 1 \circ \pi_{\vec{0}} \otimes \otimes_{\varepsilon \in V_{k+1}\setminus\{0\}} f,$$

so that by the assumption $f \circ \pi_{\vec{0}}$ correlates with F with respecto to the measure $\mu^{[k+1]}$. Now project F successively onto the invariant factors of all side transformations corresponding to sides that do not contain $\vec{0}$.

By the mean ergodic theorem, each such projection is given by the limit of certain ergodic averages. It follows that each such projection still does not depend on the coordinate $\vec{0}$. Moreover, after projecting onto all factors we obtain a function in $\mathcal{J}^{[k+1]}$. By the previous lemma it coincides with a function of the form $g \circ \pi_{\vec{0}}$ (depending only on the $\vec{0}$-th variable), and by definition the function g is contained in $\mathcal{Z}_k$. Moreover, by construction f correlates with g. □

8 Conditional weak mixing and almost periodicity

Let X be an ergodic mps. For every k, the factor $\mathcal{Z}_{k+1}(X)$ is an *extension* of $\mathcal{Z}_k(X)$ (which means that the latter factor is contained in the former). In the case $k = 0$ we have seen that the factor $\mathcal{Z}_0$ is trivial, and $\mathcal{Z}_1$ is the Kronecker factor, which is spanned by eigenfunctions. In this lecture we will see that $\mathcal{Z}_{k+1}$ is also spanned by (suitably generalized) eigenfunctions.

Let X be an mps and Y a factor of X. The *conditional scalar product* is defined by

$$\langle f, g \rangle_{L^2(X|Y)} := \mathbb{E}(f\bar{g}|Y) \in L^1(Y), \quad f, g \in L^2(X)$$

and the conditional norm by

$$\|f\|_{L^2(X|Y)} := \langle f, f \rangle_{L^2(X|Y)}^{1/2} = \mathbb{E}(|f|^2|Y)^{1/2} \in L^2(Y), \quad f \in L^2(X).$$

The space $L^2(X|Y)$ consists of $f \in L^2(X)$ such that $\|\|f\|_{L^2(X|Y)}\|_{L^\infty(Y)}$ is finite. Using Cauchy–Schwarz in each fiber of a measure disintegration of X over Y we obtain the conditional Cauchy–Schwarz inequality

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

The space $L^2(X|Y)$ is a module over the algebra $L^\infty(Y)$. A *finitely generated module zonotope of* $L^2(X|Y)$ is a set of the form $f_1 B + \dots + f_r B$, where B is the unit ball of $L^\infty(Y)$ and $f_i \in L^2(X|Y)$.

1. A function $f \in L^2(X|Y)$ is called a *conditional eigenfunction* (or a *generalized eigenfunction*) if its orbit $T^{\mathbb{Z}} f$ is contained in a finitely generated T-invariant sub-$L^\infty(Y)$-module. The space of conditional eigenfunctions is denoted by $E(X|Y)$.

2. A function $f \in L^2(X|Y)$ is called *conditionally almost periodic* (cap) if for every $\varepsilon > 0$ there exists a finitely generated module zonotope C such that the orbit $T^{\mathbb{Z}} f$ is contained in an ε-neighborhood of C. The space of cap functions is denoted by $A(X|Y)$.

3. A function $f \in L^2(X)$ is called *conditionally weakly mixing* (cwm) if

$$\text{C-}\lim_n \| \langle T^n f, f \rangle_{L^2(X|Y)} \|_{L^1(Y)}^p = 0$$

for some/all $0 < p < \infty$. Here C-lim stands for the Cesàro limit, i.e. $\text{C-}\lim_n a_n = \lim_N \frac{1}{N} \sum_{n=1}^N a_n$. The space of cwm functions is denoted by $W(X|Y)$.

In the case of the trivial factor Y the definition of conditional weak mixing and conditional almost periodicity coincide with their non-conditional counterparts. The definition of a conditional eigenfunction is different from an eigenfunction because we do not ask for rank 1 submodules.

As in the non-conditional case the space of cwm functions is in fact a closed linear subspace of $L^2(X)$, and we have

$$\overline{E(X|Y)} = \overline{A(X|Y)} = W(X|Y)^\perp.$$

We begin by showing that any conditional eigenfunction f is conditionally almost periodic. To this end we employ the following version of the Gram–Schmidt process: given any sequence of functions $(f_i)_i \subset L^2(X)$ define

$$f_i' := f_i - \sum_{j<i} \langle f_i, f_j'' \rangle_{L^2(X|Y)} f_j'', \quad f_i'' := f_i' / \|f_i'\|_{L^2(X|Y)}.$$

Here we set $0/0 := 0$ and note that division of a non-zero number by zero can occur only on a set of measure 0. This process coincides with the usual Gram–Schmidt process on each fiber of the measure disintegration, and the main additional feature is that it produces measurable functions on X. Now, applying the above conditional Gram–Schmidt process to the finite set of generators of the module containing a conditional eigenfunction we obtain a normalized set of generators f_i''. Then we can write

$$T^n f = \sum_i \langle T^n f, f_i'' \rangle_{L^2(X|Y)} f_i''.$$

This equality holds in $L^2(X)$ because it holds in $L^2(\mu_y)$ for almost every fiber measure μ_y in the disintegration of X over Y. Now, $| \langle T^n f, f_i'' \rangle_{L^2(X|Y)} | \leq \|T^n f\|_{L^2(X|Y)}$ by conditional Cauchy–Schwarz, and since the latter function is bounded we see that the orbit $T^{\mathbb{Z}} f$ is in fact contained in a finitely generated module zonotope.

The remaining inclusions are separated in a sequence of lemmas that we have already seen in the non-conditional case.

Lemma 8.1. *Let $f \in W(X|Y)$ and $g \in L^2(X)$. Then*

$$\operatorname*{C-lim}_n \| \langle T^n f, g \rangle_{L^2(X|Y)} \|_{L^1(Y)}^2 = 0.$$

This shows in particular that $W(X|Y) \subset L^2(X)$ is a linear subspace.

Proof. Consider first the case $f, g \in L^2(X|Y)$. In this case the conditional Cauchy–Schwarz inequality implies that $\langle g, T^n f \rangle_{L^2(X|Y)}$ is uniformly bounded in $L^\infty(Y)$, say by C. Write

$$\begin{aligned}
\frac{1}{N} \sum_{n=1}^{N} \| \langle T^n f, g \rangle_{L^2(X|Y)} \|_{L^2(Y)}^2 &= \frac{1}{N} \sum_{n=1}^{N} \int_Y \langle g, T^n f \rangle_{L^2(X|Y)} \langle T^n f, g \rangle_{L^2(X|Y)} \\
&= \frac{1}{N} \sum_{n=1}^{N} \int_Y \left\langle \langle g, T^n f \rangle_{L^2(X|Y)} T^n f, g \right\rangle_{L^2(X|Y)} \\
&= \left\langle \frac{1}{N} \sum_{n=1}^{N} \langle g, T^n f \rangle_{L^2(X|Y)} T^n f, g \right\rangle_{L^2(X)}.
\end{aligned}$$

By the van der Corput differencing lemma it suffices to show that

$$\operatorname*{C-lim}_h \limsup_n \frac{1}{N} \sum_{n=1}^{N} | \left\langle \langle g, T^n f \rangle_{L^2(X|Y)} T^n f, \left\langle g, T^{n+h} f \right\rangle_{L^2(X|Y)} T^{n+h} f \right\rangle_{L^2(X)} | = 0.$$

We estimate the scalar product inside the absolute value as follows:

$$\begin{aligned}
&| \int_X \langle g, T^n f \rangle_{L^2(X|Y)} T^n f \overline{\langle g, T^{n+h} f \rangle_{L^2(X|Y)}} T^{n+h} \bar{f} | \\
&= | \int_Y \langle g, T^n f \rangle_{L^2(X|Y)} \overline{\langle g, T^{n+h} f \rangle_{L^2(X|Y)}} \mathbb{E}(T^n f T^{n+h} \bar{f} | Y) | \\
&\leq C^2 \int_Y |\mathbb{E}(T^n f T^{n+h} \bar{f} | Y)| \\
&= C^2 \int_Y |\mathbb{E}(f T^h \bar{f} | Y)| \\
&\leq C^2 \|\mathbb{E}(f T^h \bar{f} | Y)\|_{L^2(Y)} \\
&= C^2 \| \left\langle f, T^h f \right\rangle_{L^2(X|Y)} \|_{L^2(Y)},
\end{aligned}$$

and this converges to 0 in the Cesàro sense by the hypothesis $f \in W(X|Y)$.

To pass to the general case note

$$\| \langle f, g\rangle_{L^2(X|Y)} \|_{L^1(Y)} = \|E(f\bar{g}|Y)\|_{L^1(Y)} \leq \|f\bar{g}\|_{L^1(X)} \leq \|f\|_{L^2(X)}\|g\|_{L^2(X)}.$$

The conclusion follows using the approximations $g1_{|g|<a} \to g$ and $f1_{|\mathbb{E}(f|Y)|<a} \to f$ in $L^2(X)$ as $a \to \infty$. Note that the second approximation is chosen inside $W(X|Y) \cap L^2(X|Y)$. ☐

Lemma 8.2. *Let $f \in W(X|Y)$ and $g \in A(X|Y)$. Then $\langle f, g\rangle_{L^2(X|Y)} = 0$.*

Proof. Let $\varepsilon > 0$ and choose $g_1, \dots, g_r \in L^2(X|Y)$ such that $T^{\mathbb{Z}}g$ in contained in an ε-neighborhood of the zonotope generated by the g_i's:

$$T^n g = \sum_{i=1}^{r} b_{n,i}g_i + r_n, \quad \|b_{n,i}\|_{L^\infty(Y)} \leq 1, \quad \|r_n\|_{L^2(X|Y)} \leq \varepsilon.$$

Then

$$\begin{aligned}
&\| \langle f, g\rangle_{L^2(X|Y)} \|_{L^2(Y)} \\
&\quad = \text{C-}\lim_n \| \langle T^n f, T^n g\rangle_{L^2(X|Y)} \|_{L^2(Y)} \\
&\quad \leq \limsup_N \frac{1}{N} \sum_{n=1}^{N} \Big(\sum_{i=1}^{r} \|\bar{b}_{n,i} \langle T^n f, g_i\rangle_{L^2(X|Y)} \|_{L^2(Y)} + \| \langle T^n f, r_n\rangle_{L^2(X|Y)} \|_{L^2(Y)}\Big) \\
&\quad \leq \limsup_N \frac{1}{N} \sum_{n=1}^{N} \Big(\sum_{i=1}^{r} \| \langle T^n f, g_i\rangle_{L^2(X|Y)} \|_{L^2(Y)} + \|\|T^n f\|_{L^2(X|Y)}\|r_n\|_{L^2(X|Y)}\|_{L^2(Y)}\Big) \\
&\quad \leq \sum_{i=1}^{r} \limsup_N \frac{1}{N} \sum_{n=1}^{N} \| \langle T^n f, g_i\rangle_{L^2(X|Y)} \|_{L^2(Y)} + \varepsilon\|\|T^n f\|_{L^2(X|Y)}\|_{L^2(Y)}.
\end{aligned}$$

The first summand is zero by Lemma 8.1 and the second summand is arbitrarily small. ☐

Lemma 8.3. *If $f \in L^2(X) \setminus W(X|Y)$, then $f \not\perp_{L^2(X)} E(X|Y)$.*

Proof. Consider first the case $f \in L^2(X|Y)$. Fix a disintegration $\mu = \int_Y \mu_y$ of the measure on X over Y. Applying the mean ergodic theorem to the relatively independent product $X \times_Y X$ we obtain

$$\frac{1}{N}\sum_{n=1}^{N} T^n \bar{f} \otimes T^n f \to H \quad \text{in } L^2(X \times_Y X),$$

where H is a $T \times T$-invariant function. By the hypothesis we have

$$\begin{aligned}
\limsup_{N\to\infty} \left\langle \frac{1}{N}\sum_{n=1}^{N} T^n \bar{f} \otimes T^n f, \bar{f} \otimes f \right\rangle_{L^2(X\times_Y X)} &= \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int_{X\times_Y X} fT^n\bar{f}\otimes\bar{f}T^n f \\
= \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int_Y \mathbb{E}(fT^n\bar{f}|Y)\mathbb{E}(\bar{f}T^n f|Y) &= \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \| \langle T^n f, f\rangle_{L^2(X|Y)} \|^2_{L^2(Y)} > 0,
\end{aligned}$$

where we have used Jensen's inequality on Y in the last passage. It follows that $H \neq 0$. Moreover, $M := \|\|H\|_{L^2(X\times_Y X|Y)}\|_{L^\infty(Y)} \leq \|\|f\|^2_{L^2(X|Y)}\|_{L^\infty(Y)}$.

Fix a representative for H. Passing to a subsequence of N's we may assume that

$$\frac{1}{N}\sum_{n=1}^{N} T^n \bar{f} \otimes T^n f \to H \quad \text{in } L^2(\mu_y \times \mu_y) \tag{8.4}$$

for almost every $y \in Y$. For almost every y we can define an integral operator

$$S_y g(x) := \int H(x', x) g(x') \mathrm{d}\mu_y(x') \quad \text{on } L^2(X, \mu_y).$$

Its Hilbert–Schmidt norm is bounded by M for a.e. y. The direct integral of the operators S_y is the operator

$$Sg(x) = S_{\pi(x)} g(x) \quad \text{on } L^2(X).$$

The properties of measure disintegration imply that $Sg = S_y g$ in $L^2(X, \mu_y)$ for a.e. y, and moreover the norm of S is bounded by M.

Using T-invariance of H one can verify that the operator S commutes with T:

$$\begin{aligned} STg(x) &= S_{\pi(x)} Tg(x) \\ &= \int H(x', x) Tg(x') \mathrm{d}\mu_{\pi(x)}(x') \\ &= \int H(Tx', Tx) g(Tx') \mathrm{d}\mu_{\pi(x)}(x') \\ &= \int H(x'', Tx) g(x'') \mathrm{d}\mu_{\pi(Tx)}(x'') \\ &= Sg(Tx). \end{aligned}$$

Note also that

$$\begin{aligned} \langle Sf, f \rangle_{L^2(X)} &= \int \int H(x', x) f(x') \bar{f}(x) \mathrm{d}\mu_{\pi(x)}(x') \mathrm{d}\mu(x) \\ &= \int \int \int H(x', x) f(x') \bar{f}(x) \mathrm{d}\mu_{\pi(x)}(x') \mathrm{d}\mu_y(x) \mathrm{d}\nu(y) \\ &= \lim_N \int \int \int \frac{1}{N} \sum_{n=1}^{N} T^n \bar{f}(x') T^n f(x) f(x') \bar{f}(x) \mathrm{d}\mu_{\pi(x)}(x') \mathrm{d}\mu_y(x) \mathrm{d}\nu(y) \\ &= \lim_N \frac{1}{N} \sum_{n=1}^{N} \int |\langle T^n f, f \rangle_{L^2(X|Y)}(y)|^2 \mathrm{d}\nu(y) \\ &> 0. \end{aligned}$$

The operators S and S_y are self-adjoint by construction. By the measurable functional calculus there exists a constant $a > 0$ such that $\langle p(S) Sf, f \rangle_{L^2(X)} \neq 0$, where $p = \chi_{[-a,a]^\complement}$. We claim that the function $p(S)Sf$ is a generalized eigenfunction.

Let p_n be a sequence of polynomials such that $p_n(-a) = 0 = p_n(a)$ and $p_n \to p$ pointwise and boundedly on $[-M, M]$ and uniformly on $[-M, -a-\varepsilon] \cup [-a, a] \cup [a + \varepsilon, M]$ for every ε. Since S_y are self-adjoint Hilbert–Schmidt operators on Hilbert spaces, $\sigma(S_y) \setminus \{0\}$ is discrete. By the continuous functional calculus $p_n(S_y)$ converges in the operator norm topology to the projection onto the linear span of the eigenspaces of S_y with eigenvalues outside $[-a, a]$.

Recall that the Hilbert-Schmidt norm of S_y is uniformly bounded. Therefore the number of eigenspaces to eigenvalues with absolute value at least a is also uniformly bounded. Therefore the rank of $p(S_y)$ is uniformly bounded. Moreover $p_n(S) \to p(S)$ in the strong operator topology by the measurable functional calculus.

Let $g \in L^2(X)$. For a.e. y and every n we have $p_n(S)g = p_n(S_y)g$ in $L^2(X, \mu_y)$. Here the right-hand side converges in $L^2(X, \mu_y)$. The left-hand side converges in $L^2(X)$, so we can pass to a subsequence such that the convergence is pointwise μ-almost everywhere, hence also pointwise μ_y-a.e. for a.e. y. Therefore the two limits coincide μ_y-a.e. for a.e. y, i.e. $p(S)g = p(S_y)g$ in $L^2(X, \mu_y)$.

It follows that $p(S)Sf \in L^2(X|Y)$. Since T commutes with S, we have $T^{\mathbb{Z}} p(S)Sf = p(S)ST^{\mathbb{Z}} f$. The above reasoning shows that the latter is a bounded sequence in

$L^2(X|Y)$. Applying the Gram–Schmidt procedure with the conditional inner product we obtain a conditionally orthogonal generating set for the module spanned by $p(S)ST^{\mathbb{Z}}f$. However, for each $y \in Y$ there can be only boundedly many functions in this basis that are not zero μ_y-a.e., since $p(S_y)$ has uniformly bounded rank. It is therefore possible to construct a finite generating set for the above module.

Consider now the case $f \notin L^2(X|Y)$. Let $F = \{|\mathbb{E}(f|Y)| \leq a\}$ be a sublevel set with a so large that $1_F f \notin W(X|Y)$ (such a exists because $1_F f \to f$ as $a \to \infty$ in $L^2(X)$ and because the expression defining $W(X|Y)$ is $L^2(X)$-continuous). Since $1_F f \in L^2(X|Y)$, by the above case we know that $1_F f$ correlates with a conditional eigenfunction g. But then $1_F g$ is also a non-zero conditional eigenfunction, and it correlates with f. $\square$

9 Compact extensions

Let X be an mps and $\pi : X \to Y$ a factor. In the last lecture we have proved the splitting

$$L^2(X) = \overline{E(X|Y)} \oplus W(X|Y)^{\perp}.$$

The space of conditional eigenfunctions $E(X|Y)$ is spanned by the finitely generated T-invariant sub-$L^\infty(Y)$-modules of $L^2(X|Y)$. Now we would like to show that this space defines a factor and extend the Halmos–von Neumann theorem to this setting, that is, write the extension[4] X in terms of the factor Y and a compact group. For simplicity we assume throughout that X is ergodic.

9.1 Conditional eigenfunctions are bounded

The main obstacle to showing that conditional eigenfunctions define a factor is that we cannot a priori multiply them (and stay in a reasonable space). As we shall presently see, there is in fact no problem with this because they are bounded.

Consider a finitely generated T-invariant $L^\infty(Y)$-submodule of $L^\infty(X)$ and construct a relatively orthonormal generating set $f_1, \dots, f_r$ for it. By the assumption of T-invariance we have

$$Tf_i = \sum_j a_{i,j} f_j,$$

where $a_{i,j} \in L^\infty(Y)$ with the convention $a_{i,j} \equiv 0$ on the set $\{\|f_j\|_{L^2(X|Y)} = 0\}$. Then

$$T \langle f_i, f_{i'} \rangle_{L^2(X|Y)} = \sum_{j,j'} a_{i,j} \overline{a_{i',j'}} \langle f_j, f_{j'} \rangle_{L^2(X|Y)} = \sum_j a_{i,j} \overline{a_{i',j}},$$

so the non-zero blocks of the matrices $(a_{i,j})$ are isometric.

Consider now the vector-valued function $\vec{f}(x) = (f_1(x), \dots, f_r(x))$ and the matrix-valued function $A(x) = (a_{i,j}(x))$. Then $T\vec{f} = A\vec{f}$, and the matrices A are partial isometries. Taking ℓ^2 norms on both sides we obtain

$$T \sum_i |f_i|^2 \leq \sum_i |f_i|^2$$

pointwise a.e. Integrating both sides and using the fact that T is measure-preserving we see that equality holds a.e. Hence $\sum_i |f_i|^2$ is an invariant function, so it it constant by the ergodicity assumption. In particular, the functions f_i are bounded, and this shows that the product of two finite rank submodules is again a finite rank submodule. Hence $E(X|Y)$ defines a factor of X (which equals $A(X|Y)$. The usual notation is $A(X|Y)$).

Moreover, taking the conditional expectation with respect to Y, we obtain

$$\sum_i \|f_i\|^2_{L^2(X|Y)} = \text{const} =: R^2.$$

This means that the functions f_i span an R-dimensional subspace of almost every fiber $L^2(\mu_y)$. It follows that we can rearrange them into a generating set consisting of R relatively orthogonal functions. In order to preserve measurability we do so by the following procedure:

$$f_{i,0} := 0, \quad f_{i,j+1} := f_{i,j} + (1 - \|f_{i,j}\|_{L^2(X|Y)}) f_{j+1}.$$

Then the functions $f_{i,r}$ span the same sub-$L^\infty(Y)$-module of $L^\infty(X)$ as the f_i's and we have

$$\|f_{1,1}\|_{L^2(X|Y)} = \dots = \|f_{i,R}\|_{L^2(X|Y)} = 1, \quad \|f_{i,R+1}\|_{L^2(X|Y)} = \dots = \|f_{i,r}\|_{L^2(X|Y)} = 0.$$

[4] In case I forgot to say this: if Y is a factor of X, then X is called an extension of Y.

9.2 Group extensions

Definition 9.1. Let (Y, T) be a dynamical system (without a measure), G a compact group, $H \le G$ a closed subgroup, and $a : Y \to G$ a function. We define

$$S(y, gH) = (Ty, a(y)gH) \tag{9.2}$$

and denote the dynamical system (X, S) by $Y \ltimes_a G/H$.

A *compact extension* of an mps (Y, μ, T) is an extension of the form $(Y \ltimes_a G/H, \mu \times m_{G/H})$, where a is measurable and $m_{G/H}$ is the Haar measure. A *group extension* is a compact extension with $H = \{\mathrm{id}_G\}$.

The generalization of the Halmos–von Neumann theorem to almost periodic extensions tells that such extensions are compact. The converse is also true.

Exercise 9.3. Show that a compact extension is generated by generalized eigenfunctions (use the Peter–Weyl theorem).

We will need a classification of invariant measures for the map (9.2). We state it first in the case $H = \{e_G\}$.

Lemma 9.4. *Let (Y, μ, T) be an ergodic mps, G a compact group, and $a : Y \to G$ a measurable function. Then there exists a closed subgroup $K \le G$ and a measurable map $\gamma : Y \to G$ such that the following holds.*

1. *The function $a'(y) = \gamma(Ty)^{-1}a(y)\gamma(y)$ takes values in K almost surely.*
2. *Every invariant ergodic measure on $Y \ltimes_{a'} G$ that projects onto μ on the first coordinate has the form $\mu \times m_{Kg_0}$, where m_{Kg_0} is the Haar measure on a coset Kg_0.*

In other words, there is a change of variables of the form $(y, g) \mapsto (y, \gamma(y)^{-1}g)$ that decomposes the transformation (9.2) into a disjoint union of invariant sets of the form $Y \times Kg_0$, each of which admits only one invariant measure extending μ, namely the product measure.

The group K is called the *Mackey group* of the cocycle[5] $a : Y \to G$. It is unique up to conjugation.

Proof. The measure $\mu \times m_G$ on $Y \times G$ is S-invariant, and by ergodic decomposition we can find an ergodic invariant measure ν on $Y \times G$ that extends μ.

The group G acts on $Y \times G$ on the right via

$$r_h(y, g) = (y, g)h = (y, gh).$$

This action commutes with the transformation S. Let $K \le G$ denote the subgroup whose right action leaves ν invariant.

By the mean ergodic theorem we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} S^n f = \int f \mathrm{d}\nu$$

in $L^2(\nu)$ for every $f \in C(Y \times G)$. Using separability of the space of continuous functions and passing to a subsequence of N's we may assume that for ν-a.e. point (y, g) is *generic*, that is, convergence holds at that point for all continuous functions f.[6]

[5] The actual cocycle here, in the sense of group cohomology, is the map $a(y, n) = a(T^{n-1}y) \cdots a(T^0 y)$. For us "cocycle" is just a convenient shortcut to designate a function taking values in a compact group.

[6] In view of the pointwise ergodic theorem there is no need to pass to a subsequence of N's here, but will not use this fact.

The set of generic points is invariant under the right action of K. Moreover, if (y, g_1) and (y, g_2) are generic, then it is easy to see that $g_1^{-1}g_2 \in K$. Hence, for every y the set G_y of g such that (y, g) is generic is either a coset of K or empty. The latter (empty) possibility can occur only for a zero measure set of y's.

Since the set of generic points is measurable, we can find a measurable function $\gamma : Y \to G$ such that $\gamma(y) \in G_y$ whenever $G_y \neq \emptyset$. This is not obvious; existence of such functions is guaranteed by so-called measurable selector theorems; the one that is most convenient in ergodic theory is due to Arsenin and Kunugui, see [Kec95, Theorem 18.18]. Under the coordinate change $(y, g) \mapsto (y, \gamma(y)^{-1}g)$ the measure ν is mapped to a measure supported on $Y \times K$, and since this measure is invariant under the right action of K, it in fact equals $\mu \times m_K$, where m_K stands for the Haar measure on K.

Under this change of variabels the measure-preserving transformation S is intertwined with the transformation

$$(y, g) \mapsto (Ty, a'(y)g).$$

Since this map preserves the measure $\mu \times m_K$, the function $\gamma(Ty)^{-1}a(y)\gamma(y)$ has to take values in K almost surely.

It remains to show that all other ergodic S'-invariant measures that extend μ have the claimed form. Given any ergodic S'-invariant measure ν' that extends μ, any pushforward measure $(r_g)_*\nu'$ also extends μ. Moreover, the measure $\int_G (r_g)_*\nu' \mathrm{d}g$ is invariant under the right G action and extends μ, so it equals $\mu \times m_G$. The claim follows from essential uniqueness of measure disintegration. □

Lemma 9.5 ([FW96, Lemma 7.3])**.** *Let $X \xrightarrow{\omega} W \xrightarrow{\pi} Y$ be a chain of factors of an ergodic system and assume that $X \to Y$ is a group extension. Then $X \to W$ is also a group extension.*

Proof. By the hypothesis we have $X = Y \ltimes_a G$. Consider the map

$$\iota : X = Y \times_a G \to W \times_{a\circ\pi} G, \quad x = (y, g) \mapsto (\omega(x), g).$$

This map is injective: a left inverse is given by $(w, g) \mapsto (\pi(w), g)$. Moreover, it intertwines the transformations on X and $W \ltimes_{a\circ\pi} G$. The pushforward measure $\iota_*(\mu_X)$ is ergodic, so by Lemma 9.4, up to a change of coordinates, it has the form $\mu_W \times m_K$, where K is the Mackey group of the cocycle $a \circ \pi : W \to G$. □

9.3 Compact extensions

Return now to the setting of Definition 9.1.

Lemma 9.6. *Let (Y, μ, T) be an ergodic mps, G a compact group, $H \leq G$ a closed subgroup, and $a : Y \to G$ measurable. Let ν be an invariant ergodic measure on $X := Y \ltimes_a G/H$ that extends μ. Then (X, ν, S) is measurably isomorphic to a compact extension of Y.*

In other words, an extension of the form (9.2) cannot carry invariant measures which are not of product type.

Proof. After a change of variable we may assume $a = a'$ in Lemma 9.4. The measure ν lifts to an S- and H- invariant measure on $Y \ltimes_a G$. By ergodic decomposition this lift can be written as an integral of ergodic measures, and each such measure has the product form $\mu \times m_{Kg_0}$ by Lemma 9.4.

The projection of such measures onto $Y \times G/H$ has the form $\mu \times m_{Kg_0/H}$. These measures are ergodic and provide an ergodic decomposition of ν. By essential uniqueness of ergodic decomposition ν itself has this form. □

9.4 Construction of homogeneous spaces

We have seen that every finite rank T-invariant sub-$L^\infty(Y)$-module of $L^\infty(X)$ admits a set of generators $f_1, \dots, f_R$ that satisfy

$$\|f_1\|_{L^2(X|Y)} = \dots = \|f_R\|_{L^2(X|Y)} = 1, \quad \sum_i \|f_i\|^2_{L^2(X|Y)} = R^2, \quad Tf_i = \sum_j a_{i,j} f_j,$$

with functions $a_{i,j} \in L^\infty(Y)$ such that the matrices $A = (a_{i,j})$ are unitary almost everywhere. Passing to a suitable topological model we may assume that $f_i, a_{i,j}$ are continuous.

The vector-valued function $\vec{f}$ takes values in the R-dimensional sphere S_R of radius R. The unitary group $U(R)$ acts on this sphere transitively, so the sphere is homeomorphic to a quotient of the unitary group, namely the quotient $U(R)/U(R-1)$. We have a map from X to $Y \times S_R \cong Y \times U(R)/U(R-1)$ given by $(\pi, \vec{f})$. Moreover, on $Y \times U(R)/U(R-1)$ we have the continuous map

$$S(y, gU(R-1)) = (Ty, A(y)gU(R-1))$$

that satisfies

$$S(\pi(x), \vec{f}(x)) = (T\pi(x), A(\pi(x))\vec{f}(x)) = (T\pi(x), (T\vec{f})(x)) = (\pi(Tx), (\vec{f})(Tx)).$$

Equipping $Y \times U(R)/U(R-1)$ with the pushforward measure we thus obtain a measure-preserving system that is a topological model of the factor generated by Y and the f_i's (note that the space of continuous functions on $Y \times S_R$ is generated by $C(X)$ and the coordinate functions on S_R).

Using a countable family of finite rank submodules that span $\overline{A(X|Y)}$ we obtain a topological model for the factor $A(X|Y)$ of the form

$$Y \times G/H, \quad S(y, gH) = (Ty, a(y)gH),$$

where G is compact group, $H \leq G$ a closed subgroup, $a : Y \to G$ is a continuous map, with some S-invariant ergodic measure ν. By Lemma 9.6 this extension is measurably isomorphic to a compact extension.

10 HKZ factors are almost periodic

Let X be an ergodic mps. By monotonicity of uniformity seminorms we know that $Z_k(X) \supset Z_{k-1}(X)$. Our next objective is to show that this extension is almost periodic. This is contained in the following lemma.

Lemma 10.1. *Suppose $f \in W(X|Z_{k-1})$. Then $\|f\|_{[k+1]} = 0$.*

Proof. We may assume that f is real-valued, this will simplify notation. Recall that

$$\|f\|_{[k+1]}^{2^{k+1}} = \int_{X^{[k]}} |\mathbb{E}(\otimes_{V_k} f|\mathcal{I}^{[k]})|^2 \mathrm{d}\mu^{[k]}.$$

Hence it suffices to show $\otimes_{V_k} f \perp \mathcal{I}^{[k]}$. We will prove the stronger statement

$$\otimes_{V_k} f \in W(X^{[k]}|Z_{k-1} \circ \pi_{\vec{0}} \vee X^{[k]*}),$$

where $X^{[k]*}$ is the subalgebra of functions that do not depend on the $\vec{0}$-th coordinate. This is indeed stronger because the space of invariant functions is spanned by one-dimensional invariant subspaces and is contained in the almost periodic subspace over any factor.

To this end it suffices to show that, for every $f \in L^1(X)$, we have

$$\mathbb{E}(f \circ \pi_{\vec{0}}|Z_{k-1} \circ \pi_{\vec{0}} \vee X^{[k]*}) = \mathbb{E}(f|Z_{k-1}) \circ \pi_{\vec{0}}.$$

Assuming this, we can conclude using only the definition of relative weak mixing. By L^1 continuity of conditional expectation we may assume $f \in \mathcal{X}$. Splitting $f = \mathbb{E}(f|Z_{k-1}) + (f - \mathbb{E}(f|Z_{k-1}))$ we may consider two cases separately:

1. $f \in Z_{k-1}$. In this case both sides clearly are equal to $f \circ \pi_{\vec{0}}$.

2. $\mathbb{E}(f|Z_{k-1}) = 0$. In this case we will show that $f \perp \otimes_{\varepsilon \in V_k} f_\varepsilon$ for any $f_{\vec{0}} \in Z_{k-1}$ and $f_\varepsilon \in \mathcal{X}$ for $\varepsilon \in V_k^* := \{0,1\}^k \setminus \{\vec{0}\}$. Indeed, note that $\mathbb{E}(ff_{\vec{0}}|Z_{k-1}) = f_{\vec{0}}\mathbb{E}(f|Z_{k-1}) = 0$, and hence $\|ff_{\vec{0}}\|_{[k]} = 0$. By the Cauchy–Schwarz–Gowers inequality we obtain

$$\int_{X^{[k]}} (ff_{\vec{0}}) \otimes \otimes_{\varepsilon \in V_k^*} f_\varepsilon \mathrm{d}\mu^{[k]} = 0,$$

as required.

□

10.1 Monotone approximation by almost periodic functions

Let $X \to Y$ be a factor and $f \in \overline{A(X|Y)}$ (L^2 closure). Then in particular $f \in \overline{E(X|Y)}$, so let $(f_n) \subset E(X|Y)$ be a sequence such that $f_n \to f$ in L^2. One way to write this is

$$\int_Y \mathbb{E}(|f - f_n|^2|Y) \to 0.$$

By Egorov's theorem we may pass to a subsequence of f_n's such that for every $\varepsilon > 0$ convergence is uniform outside a subset F_ε of Y of measure $\leq \varepsilon$. It follows that the functions $f1_{Y\setminus F_\varepsilon}$ are conditionally almost periodic over Y.

If f is positive, then $0 \leq f1_{Y\setminus F_\varepsilon} \leq f$ and $f1_{Y\setminus F_\varepsilon} \to f$ as $\varepsilon > 0$.

10.2 Product systems of compact extensions

Notation: Let $X = Y \ltimes_a G/H$ be a compact extension. Wer have a left G-action on X given by

$$l_g(y, g_0H) = g(y, g_0H) = (y, gg_0H).$$

In general, this action does not commute with the measure-preserving transformation (T, a).

Lemma 10.2. *Let $X = Y \ltimes_a G$ be an ergodic group extension. Then the left and right G-actions coincide on the Kronecker factor $K(X)$ and vanish on the commutator subgroup $[G, G]$.*

Proof. Let f be a non-zero eigenfunction on X with eigenvalue λ. Then $f \circ r_g$ is also an eigenfunction with eigenvalue λ, so by ergodicity $f \circ r_g = \pi(g)f$ with $\pi(g) \in \mathbb{C}$. Since $\|f\|_2 = \|f \circ r_g\|_2$, we have $|\pi(g)| = 1$. Moreover, it is easy to verify that π is a group homomorphism. Since the range of π is an abelian group, it vanishes on the commutator subgroup $[G, G]$.

Finally, this shows that f comes from a function on $Y \times G/[G, G]$. On the latter space the left and the right G-actions coincide. □

Lemma 10.3. *Let $X = Y \ltimes_a G$ be an ergodic group extension. Then for every $g \in G$ the left action of the element $(g, g) \in G^2$ on the invariant factor $I(X^2)$ is trivial.*

Proof. Recall that $\mathcal{I}(X^2) \subset \mathcal{K}(X)^2$. Moreover, $\mathcal{I}(X^2)$ is spanned by functions of the form $f \otimes \bar{f}$, where f is an eigenfunction on X. We have just seen that

$$f \circ l_g = \pi(g)f,$$

where $\pi : G \to \{z \in \mathbb{C}, |z| = 1\}$ is a homomorphism. Hence

$$f \circ l_g \otimes \overline{f \circ l_g} = \pi(g)f \otimes \overline{\pi(g)f} = f \otimes \overline{f}$$

as required. □

10.3 HKZ factors are abelian group extensions

Errata: in [FW96, Lemma 8.4] replace $\hat{Z}$ by Z_2. The proof of [HK05, Proposition 6.3(1)] does not work as stated because the ergodicity hypothesis in [HK05, Lemma 6.1] is not satisfied. Solution: pass to a "normal" extension in the sense of [FW96]. This introduces additional complications in the inductive scheme that proves the structure theorem for Z_k factors, see [Zie07] for details. Since I have failed to account for this problem, we will probably have to stick to $k = 2$ in this course.

Lemma 10.4. *Let $X \to Y$ be a factor of an ergodic mps. Then $\mathcal{Z}_k(Y) = \mathcal{Z}_k(X) \cap \mathcal{Y}$.*

Proof. To see the inclusion $\subseteq$ recall the definition of $\mathcal{Z}_k(Y)$: it consists of the functions f such that the function $f \circ \pi_{\vec{0}}$ on $Y^{[k+1]}$ coinsides $\mu^{[k+1]}$-a.e. with a function F that does not depend on the $\vec{0}$-th coordinate. Then F lifts to a function on $X^{[k+1]}$ that does not depend on the $\vec{0}$-th coordinate, and this shows that $f \in \mathcal{Z}_k(X)$.

To see the inclusion $\supseteq$ note

$$\mathcal{Z}_k(X) \cap \mathcal{Y} = N_{k+1}(X)^{\perp} \cap \mathcal{Y} \subseteq N_{k+1}(Y)^{\perp} \cap \mathcal{Y} = \mathcal{Z}_k(Y).$$

□

Lemma 10.5. *Let Y be an ergodic mps. Then for every $k \geq 0$ there exists an ergodic extension $X \to Z_{k+1}(Y)$ such that $X \to Z_k(X)$ is a group extension.*

Proof. We know that $Z_{k+1}(Y) \to Z_k(Y)$ is a compact extension, hence $Z_{k+1}(Y) = Z_k(Y) \ltimes_a G/H$, and we can choose a, G, H so that the system $X := Z_k(Y) \ltimes_a G$ is ergodic.

By Lemma 10.4 we have $\mathcal{Z}_k(X) \supseteq \mathcal{Z}_k(Y)$. Since X is a group extension of $Z_k(Y)$, a lemma from the previous lecture implies that X is a group extension of $Z_k(X)$. $\square$

Lemma 10.6. *Let X be an ergodic mps, $k \geq 2$, and suppose that $X = Z_{k-1}(X) \ltimes_a G$ is a group extension. Then*

1. *For every $g \in G$ and every edge $\alpha \subset V_k$, the transformation*
$$(g_\alpha \vec{x})_\varepsilon = \begin{cases} g x_\varepsilon, & \varepsilon \in \alpha \\ x_\varepsilon, & \varepsilon \notin \alpha \end{cases}$$
acts trivially on $\mathcal{I}^{[k]}$.

2. *For every $g \in G$ and every edge $\alpha \subset V_{k+1}$, the transformation g_α preserves $\mu^{[k+1]}$.*

3. *For every $g \in [G, G]$ the transformation g acts trivially on $\mathcal{Z}_k(X)$.*

4. *$\mathcal{Z}_k(X)$ is an abelian group extension of $\mathcal{Z}_{k-1}(X)$.*

Proof. I only write down the proof of the first claim and refer to [HK05, Proposition 6.3] for the remaining claims. By symmetry it suffices to prove the first claim for any fixed edge $\alpha \subset V_k$, say $\alpha = \{0 \dots 00, 0 \dots 01\}$.

We claim
$$\mathbb{E}(F|\mathcal{I}^{[k-1]}) = \mathbb{E}(\tilde{F}|\mathcal{I}^{[k-1]}), \tag{10.7}$$
for every bounded function F on $X^{[k-1]}$, where
$$\tilde{F}((y_\varepsilon, g_\varepsilon)_{\varepsilon \in V_{k-1}}) = \int_{G^{2^{k-1}}} F((y_\varepsilon, g'_\varepsilon)_{\varepsilon \in V_{k-1}}) \mathrm{d}(g'_{\vec{0}}, \dots, g'_{\vec{1}}).$$

It suffices to verify this for the dense subspace of tensor products $F = \otimes_{\varepsilon \in V_{k-1}} f_\varepsilon$. In this case we can write $f_\varepsilon = f_{\varepsilon,k-1} + f_{\varepsilon,\perp}$ with $f_{\varepsilon,k-1} = \mathbb{E}(f_\varepsilon | \mathcal{Z}_{k-1})$. The expectation on the left-hand side splits into 2^{k-1} terms. All but one of them (the one without $\perp$ functions) vanish in view of the Cauchy–Schwarz–Gowers inequality. This allows to conclude the proof of (10.7).

Let
$$\mu_{k-1}^{[k-1]} = \int \mu_{k-1,\omega}^{[k-1]} \mathrm{d}\omega$$
be an ergodic decomposition of the measure on $Z_{k-1}(X)^{[k-1]}$. Then by (10.7) and the definition of measure disintegration
$$\mu^{[k-1]} = \int \mu_\omega^{[k-1]} \times m_G^{2^{k-1}} \mathrm{d}\omega$$
is an ergodic decomposition of the measure on $X^{[k-1]}$. Thus by definition we have
$$\mu^{[k]} = \int \mu_{k-1,\omega}^{[k-1]} \times m_G^{2^{k-1}} \times \mu_{k-1,\omega}^{[k-1]} \times m_G^{2^{k-1}} \mathrm{d}\omega.$$

A bounded function on $X^{[k]}$ is $\mu^{[k]}$-a.e. invariant under $T^{[k]}$ iff it is a.e. invariant under $T^{[k]}$ with respect to ω-a.e. measure
$$(\mu_{k-1,\omega}^{[k-1]} \times m_{G^{2^{k-1}}})^2.$$

The first claim of the Lemma follows from Lemma 10.3 applied with the ergodic group extension $(Z_{k-1}^{[k-1]} \ltimes_{a^{\otimes 2^{k-1}}} G^{2^{k-1}}, \mu_{k-1,\omega}^{[k-1]} \times m_{G^{2^{k-1}}})$ and the group element $(g, \mathrm{id}_G, \dots, \mathrm{id}_G) \in G^{2^{k-1}}$. $\square$

11 The Conze–Lesigne equation

11.1 Nilpotent groups acting on HKZ factors

Lemma 11.1. *Let (X, μ, T) be an mps, $k \geq 1$, and $g : X \to X$. Then the following conditions are equivalent.*

1. *$g^{[k]}$ preserves $\mu^{[k]}$ and acts trivially on $\mathcal{I}^{[k]}$,*
2. *for every face $\alpha \subset V_{k+1}$ the transformation g_α preserves $\mu^{[k+1]}$,*
3. *for every face $\alpha \subset V_k$ the transformation g_α preserves $\mu^{[k]}$ and maps $\mathcal{I}^{[k]}$ to itself.*

We denote the set of transformations satisfying the above equivalent conditions by G_k.

Proof. (1) $\implies$ (2): By symmetry we may consider the side $\alpha = V_k \times \{0\}$. Then

$$\int (F \otimes \tilde{F}) \circ g_\alpha \mathrm{d}\mu^{[k+1]} = \int F \circ g^{[k]} \otimes \tilde{F} \mathrm{d}\mu^{[k+1]} = \int \mathbb{E}(F \circ g^{[k]} | \mathcal{I}^{[k]}) \tilde{F} \mathrm{d}\mu^{[k]},$$

and one can remove $g^{[k]}$ by the hypothesis (1).

(3) $\implies$ (2): By symmetry we may consider a side of the form $\alpha = \alpha' \times \{0, 1\}$, where $\alpha' \subset V_k$ is a side. Then

$$\int (F \otimes \tilde{F}) \circ g_\alpha \mathrm{d}\mu^{[k+1]} = \int F \circ g_{\alpha'} \otimes \tilde{F} \circ g_{\alpha'} \mathrm{d}\mu^{[k+1]} = \int \mathbb{E}(F \circ g_{\alpha'} | \mathcal{I}^{[k]}) \tilde{F} \circ g_{\alpha'} \mathrm{d}\mu^{[k]}.$$

By the hypothesis that $g_{\alpha'}$ maps $\mathcal{I}^{[k]}$ to itself we may pull it out of the conditional expectation, and by the hypothesis that $g_{\alpha'}$ preserves $\mu^{[k]}$ we may remove it.

(2) $\implies$ (3): Let $\alpha \subset V_k$ be a face, then $\alpha' = \alpha \times \{0, 1\} \subset V_{k+1}$ is also a face. Invariance of $\mu^{[k]}$ under g_α follows by projection from invariance of $\mu^{[k+1]}$ under $g_{\alpha'}$. The algebra $\mathcal{I}^{[k]}$ is mapped to itself because

$$\|\mathbb{E}(F \circ g_\alpha | \mathcal{I}^{[k]})\|^2_{L^2(\mu^{[k]})} = \int F \circ g_\alpha \otimes \overline{F \circ g_\alpha} \mathrm{d}\mu^{[k+1]} = \int (F \otimes \overline{F}) \circ g_{\alpha'} \mathrm{d}\mu^{[k+1]},$$

and by the hypothesis g can be replaced by identity.

(2) $\implies$ (1): We have already proved (3), so g_α preserves $\mu^{[k]}$ for every side $\alpha \subset V_k$; a fortiori $g^{[k]}$ preserves $\mu^{[k]}$. Let now $F \in L^2(\mathcal{I}^{[k]})$, then

$$\begin{aligned}\int (F \circ g^{[k]}) \bar{F} \mathrm{d}\mu^{[k]} &= \int \mathbb{E}(F \circ g^{[k]} | \mathcal{I}^{[k]}) \bar{F} \mathrm{d}\mu^{[k]} \\ &= \int (F \circ g^{[k]}) \otimes \bar{F} \mathrm{d}\mu^{[k+1]} = \int (F \otimes \bar{F}) \circ g_\alpha \mathrm{d}\mu^{[k+1]},\end{aligned}$$

where $\alpha = V_k \times \{0\} \subset V_{k+1}$ is a side. By the hypothesis we may remove g. □

Observations:

1. Using face projections we see $G_{k+1} \subseteq G_k$.
2. The transformation T is contained in every G_k.
3. Each set G_k is a group.

Lemma 11.2. *Suppose that $X = Z_k(X)$. Then G_k is a nilpotent group of step k.*

Proof. It suffices to show that each k-fold iterated commutator $g = [\dots[g_1, g_2], \dots, g_{k+1}]$ with $g_i \in G_k$ acts trivially on X. To this end it suffices to show that for each bounded real-valued function f on X one has

$$0 = \|f - f \circ g\|_{[k+1]}^{2^{k+1}} = \int \otimes_{V_{k+1}} (f - f \circ g) \mathrm{d}\mu^{[k+1]}.$$

Expanding the multilinear expression on the right-hand side we obtain signed integrals that cancel out precisely provided that each map g_α, $\alpha \subset V_{k+1}$, preserves the measure $\mu^{[k+1]}$, where

$$(g_\alpha \vec{x})_\varepsilon = \begin{cases} g x_\varepsilon, & \varepsilon \in \alpha \\ x_\varepsilon, & \varepsilon \notin \alpha. \end{cases}$$

It suffices to consider singletons $\alpha = \{\varepsilon\}$. In this case we can write $\alpha = \cap_{i=1}^{k+1} \alpha_i$, where $\alpha_i \subset V_{k+1}$ are sides. Then $g_\alpha = [\dots[(g_1)_{\alpha_1}, (g_2)_{\alpha_2}], \dots, (g_{k+1})_{\alpha_{k+1}}]$, and the claim follows because each map $(g_i)_{\alpha_i}$ preserves $\mu^{[k+1]}$. $\square$

It is interesting to know when the group G_k acts transitively on Z_k. We will address this question only in a special case. An mps X is said to have *order 2* if $X = Z_2(X)$ and $Z_2(X)$ is an abelian group extension of $Z_1(X)$. Recall that Z_1 is a compact abelian group on which T acts by translation by a group element $t \in Z_1$. Our objective is to obtain information on the cocycle defining the group extension X from the hypothesis that X is of order 2.

For simplicity we make the standing assumption that the group in the extension $Z_2 = Z_1 \ltimes_\rho S^1$ is the circle group $S^1 = \{z \in \mathbb{C} : |z| = 1\}$.

11.2 Conze–Lesigne equation

Let (X, μ, T) be an ergodic mps and let U be a group of automorphisms that acts freely on X. Let also $\rho : X \to S^1$. The *Conze–Lesigne equation* is

$$\rho \circ u / \rho = c f \circ T / f, \tag{11.3}$$

where $u \in U$, $f : X \to S^1$, $c \in S^1$.

Let (u, f, c) and (u', f', c') be two solutions to the CL equation. Then

$$\rho \circ (uu')/\rho = (\rho \circ u'/\rho) \circ u \cdot (\rho \circ u/\rho) = (c' f' T/f') \circ u \cdot (c f T/f) = c' c f'' T/f'',$$

where $f'' = f' \circ u \cdot f$. Hence $(uu', f' \circ u \cdot f, cc')$ is also a solution. Therefore the set of solutions of the CL equation is a (closed) subgroup of $U \ltimes C(X, S^1) \times \mathbb{T}$.

Denote this group by $\mathcal{H}$. It follows from ergodicity that the commutator subgroup of $\mathcal{H}$ is contained in $\mathcal{H}_2 := \{\mathrm{id}_U\} \times C_{\mathrm{const}}(X, \mathbb{T}) \times \{1\}$, where $C_{\mathrm{const}}(X, \mathbb{T}) \cong \mathbb{T}$ is the set of constant maps. Then $\mathcal{K} := \mathcal{H}/\mathcal{H}_2$ is a locally compact abelian group. The projection onto the last coordinate defines a character φ on $\mathcal{K}$. Denote the projection onto the first coordinate by $q : \mathcal{K} \to U$.

By the structure theorem for locally compact abelian groups, $\mathcal{K}$ admits an open subgroup $\mathcal{L} \cong K \times \mathbb{R}^d$, where K is a compact abelian group. Set $K_0 := K \cap \ker \varphi$, then $U_0 := q(K_0)$ is a closed subgroup of U. Claim: U/U_0 is a compact Lie group. Consider the short exact sequence

$$0 \to q(\mathcal{L})/U_0 \to U/U_0 \to U/q(\mathcal{L}) \to 0.$$

The last group in this sequence is finite because $q(\mathcal{L})$ is an open subgroup of U, because q is an open map.

$$0 \to q(K)/q(K_0) \to q(\mathcal{L})/U_0 = q(\mathcal{L})/q(K_0) \to q(\mathcal{L})/q(K) \to 0$$

The first group in this sequence is a quotient of $K/K_0 \cong \varphi(K)$, which is subgroup of the torus. The last group in this sequence is a compact quotient of $\mathcal{L}/K \cong \mathbb{R}^d$, hence a torus.

Let U_1 be the connected component of the identity of U_0, then $\mathcal{H} \cap U_1 \ltimes C(X, \mathbb{T}) \times \{1\}$ is an abelian group (this follows from compactness of U_0). In other words, if $(u, f, 1)$ and $(u', f', 1)$ are two solutions with $u, u' \in U_1$, then $f' \circ u \cdot f = f \circ u' \cdot f'$.

11.3 Solvability of the Conze–Lesigne equation

The main structural result will be that in the case $X = U = Z_1$ the set of solutions of (11.3) has full projection onto the coordinate U, that is, for every $s \in S^1$ there is a solution (s, f, c)..

It is easy to see that the set of s for which (11.3) has a solution is a group. Moreover, for $s = t$ it has the solution $f = \rho$, $c = 1$. Thus it suffices to consider s in a sufficiently small neighborhood of the identity.

For any mps X, compact abelian group U and map $\rho : X \to U$ denote

$$\Delta^k \rho : X^{[k]} \to U, \quad (x_\varepsilon)_{\varepsilon \in V_k} \mapsto \prod_{\varepsilon \in V_k} C^{|\varepsilon|} \rho(x_\varepsilon),$$

where C denotes inversion in U.

Lemma 11.4. *Let $\rho : Z_{k-1} \to S^1$ be a cocycle that defines a system $X = Z_{k-1} \ltimes_\rho S^1$ of order k. Then the cocycle $\Delta^k \rho$ is a coboundary, that is, there exists a function F on $Z_{k-1}^{[k]}$ such that*

$$T^{[k]} F = F \Delta^k \rho \tag{11.5}$$

and $|F| \equiv 1$.

Proof. Consider the function

$$\psi : X = Z_{k-1} \times S^1 \to \mathbb{C}, \quad (x, u) \mapsto u.$$

By the hypothesis $X = Z_k(X)$ we have $\|\psi\|_{[k+1]} \neq 0$. By definition of cube seminorms we have

$$\|\psi\|_{[k+1]}^{2^{k+1}} = \int |\mathbb{E}(\Psi|\mathcal{I}^{[k]})|^2 \mathrm{d}\mu^{[k]},$$

where $\Psi = \otimes_{\varepsilon \in V_k} C^{|\varepsilon|} \psi$, so $\mathbb{E}(\Psi|\mathcal{I}^{[k]}) \neq 0$.

Note

$$T^{[k]} \Psi = \Delta^k \rho \cdot \Psi,$$

and consequently

$$(T^{[k]})^n \Psi = \Psi \cdot \prod_{m=0}^{n-1} (T^{[k]})^m \Delta^k \rho.$$

By the mean ergodic theorem we obtain

$$\mathbb{E}(\Psi|\mathcal{I}^{[k]}) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} (T^{[k]})^n \Psi = \bar{F} \Psi$$

with

$$\bar{F} := \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \prod_{m=0}^{n-1} (T^{[k]})^m \Delta^2 \rho \in L^\infty(Z_{k-1}^{[k]}).$$

This function is not identically zero because $\mathbb{E}(\Psi|\mathcal{I}^{[k]}) \neq 0$ and we have

$$T^{[k]} \bar{F} = T^{[k]}(\bar{F} \Psi \bar{\Psi}) = \bar{F} \Psi T^{[k]}(\bar{\Psi}) = \bar{F} \Psi \bar{\Psi} \overline{\Delta^k \rho} = \bar{F} \overline{\Delta^k \rho}.$$

Thus the function F satisfies (11.5). It remains to ensure $|F| \equiv 1$. Consider the function

$$\Pi : \mathbb{C} \to \mathbb{C}, \quad z \mapsto \begin{cases} |z|^{-1}z, & z \neq 0, \\ 0, & z = 0. \end{cases}$$

Replacing F by $\Pi \circ F$ we may assume that $|F|$ is $\{0,1\}$-valued.

Let now F be a function that satisfies (11.5) and let $\alpha \subset V_k$ be a side. Then

$$T^{[k]}(T_\alpha F \cdot \overline{\Delta_\alpha \rho}) = T_\alpha(F \cdot \Delta^k \rho) \cdot \overline{T_\alpha \Delta_\alpha \rho} = T_\alpha F \cdot \Delta_{V_k \setminus \alpha} \rho = (T_\alpha F \cdot \overline{\Delta_\alpha \rho}) \Delta^{[k]} \rho,$$

so that the function $T_\alpha F \cdot \overline{\Delta_\alpha \rho}$ also satisfies (11.5). It follows that for every element $\tilde{T}$ of the side transformation group $T^{[k]}_{k-1}$ there is a unimodular function $u_{\tilde{T}}$ such that $\tilde{T}F \cdot u_{\tilde{T}}$ satisfies (11.5). Let $(\tilde{T}_n)_{n=1}^\infty$ be an enumeration of $T^{[k]}_{k-1}$ and define

$$\tilde{F} := \sum_{n=0}^{\infty} 3^{-n} \tilde{T}_n F \cdot u_{\tilde{T}_n}.$$

By lacunarity of the coefficients 3^{-n} and since the group $T^{[k]}_{k-1}$ acts ergodically on $Z^{[k]}_{k-1}$, this function is non-zero $\mu^{[k]}$-a.e., and $\Pi \circ \tilde{F}$ satisfies the conclusion of the lemma. □

Corollary 11.6. *Let (X, μ, T) be an ergodic mps, ρ a cocycle of type k, and U a compact abelian group that acts on X freely by automorphisms such that the corresponding edge transformations preserve $\mu^{[k]}$ and act weakly L^2 continuously. Let $s \in U$ be in a sufficiently small neighborhood of the identity. Then there exists a non-zero bounded function F on $X^{[1]}$ such that*

$$T^{[1]} F = F \Delta^1(\rho \circ s \cdot \bar{\rho}).$$

In fact one need not restrict to s in a small neighborhood of the identity and one can also take $|F| \equiv 1$, see [HK05, Corollary 7.5(1)], but this is substantially harder to prove.

Proof. Let $\alpha \subset V_k$ be the first side and $\xi_\alpha : X^{[k]} \to X^{[1]}$ the corresponding coordinate projection. Then

$$\Delta^1(\rho \circ s \cdot \bar{\rho}) \circ \xi_\alpha = \Delta^k \rho \circ s_\alpha \cdot \overline{\Delta^k \rho}.$$

By the hypothesis

$$T^{[k]} F = F \Delta^k \rho$$

for some measurable $F : X^{[2]} \to \mathbb{T}$. Hence

$$\Delta^1(\rho \circ s \cdot \bar{\rho}) \circ \xi_\alpha = (\bar{F} T^{[k]} F) \circ s_\alpha \cdot \overline{\bar{F} T^{[k]} F}$$

Since the transformation s_α on $X^{[k]}$ commutes with $T^{[k]}$, we obtain

$$T^{[k]} \tilde{F} = \tilde{F} \cdot \Delta^1(\rho \circ s \cdot \bar{\rho}) \circ \xi_\alpha$$

with the function $\tilde{F}_s := F \circ s_\alpha \cdot \overline{F}$. Projection onto the first side yields

$$T^{[k]} \mathbb{E}(\tilde{F}_s | \operatorname{im} \xi_\alpha) = \mathbb{E}(\tilde{F}_s | \operatorname{im} \xi_\alpha) \Delta^1(r_s \rho \cdot \bar{\rho}) \circ \xi_\alpha.$$

It remains to show that $\mathbb{E}(\tilde{F}_s | \operatorname{im} \xi_\alpha) \neq 0$. But by weak continuity of the edge action of U we have $\int \tilde{F}_s \neq 0$ for s in a sufficiently small neighborhood of the identity, and the integral is preserved under conditional expectation. □

Lemma 11.7. *Let (X, μ) be an ergodic mps and $\rho : X \to S^1$. Then there exists $\lambda \in S^1$ such that $(X \ltimes_{\lambda\rho} S^1, \mu \times m_{S^1})$ is ergodic.*

Proof. Let λ be rationally independent from all eigenvalues of T on $L^2(X)$ and suppose that both $X \ltimes_\rho S^1$ and $X \ltimes_{\lambda\rho} S^1$ are not ergodic.

Then by the Mackey group construction ρ and $\lambda\rho$ are cohomologous to cocycles taking values in proper closed subgroups of S^1. The only such subgroups are the finite subgroups, and the two finite subgroups above are contained in a common finite subgroup, say K. The quotient S^1/K is again isomorphic to S^1 and under this isomorphy the congruence class of λ is not an eigenvalue of T.

Hence we may assume that ρ and $\lambda\rho$ are both cohomologous to the constant zero cocycle:

$$\rho(x) = f_1(Tx)/f_2(x), \quad \lambda\rho(x) = f_2(Tx)/f_2(x).$$

But then

$$\lambda = (f_2(Tx)/f_2(x))/(f_1(Tx)/f_2(x)) = (f_2/f_1)(Tx)/(f_2/f_1)(x),$$

so λ is an eigenvalue of T, a contradiction. □

This fills a gap in the proof of the following lemma.

Lemma 11.8 ([FW96, Lemma 10.3]). *Let X be an ergodic mps and $\rho : X \to \mathbb{T}$ be such that*

$$T^{[1]}F = F\Delta^1\rho$$

for some non-zero $F \in L^\infty(X^{[1]})$. Then ρ is a quasi-coboundary, that is, $\rho(x) = \lambda f(Tx)\bar{f}(x)$ for some $f : X \to S^1$ and some constant $\lambda \in S^1$.

Proof. Since for any constant λ we have $\Delta^1\rho = \Delta^1(\lambda\rho)$ and by Lemma 11.7 we may assume without loss of generality that $\tilde{X} := X \ltimes_\rho S^1$ is ergodic. On $\tilde{X}^2$ we have the invariant function

$$((x_1, u_1), (x_1, u_2)) \mapsto u_1\bar{u}_2F(x_1, x_2).$$

This can be written in terms of eigenfunctions on $\tilde{X}$ as

$$\sum_\lambda c_\lambda\varphi_\lambda \otimes \overline{\varphi_\lambda}.$$

By Fourier expansion in the S^1 coordinate $\varphi_\lambda(x, u) = \sum_m \varphi_{\lambda,m}(x)u^m$ we obtain

$$u_1\bar{u}_2F(x_1, x_2) = \sum_{\lambda,m_1,m_2} c_\lambda\varphi_{\lambda,m_1}(x_1)u_1^{m_1}\overline{\varphi_{\lambda,m_2}(x_2)u_2^{m_2}} = \sum_\lambda c_\lambda\varphi_{\lambda,1}(x_1)u_1\overline{\varphi_{\lambda,1}(x_2)u_2},$$

so that $\varphi_{\lambda,1} \not\equiv 0$ for some λ. On the other hand,

$$\lambda\sum_m \varphi_{\lambda,m}(x)u^m = (T, \rho)\varphi_n(x, u) = \sum_m \varphi_{\lambda,m}(Tx)(\rho(x)u)^m,$$

and by uniqueness of Fourier series

$$\lambda\varphi_{\lambda,1}(x) = \varphi_{\lambda,1}(Tx)\rho(x).$$

Taking absolute values on both sides we obtain

$$|\varphi_{\lambda,1}(x)| = |\varphi_{\lambda,1}(Tx)|,$$

so by ergodicity of X we may normalize $|\varphi_{\lambda,1}| \equiv 1$, and this gives the claim with $f = \varphi_{\lambda,1}$. □

Lemma 11.9 ([Zie07, Theorem 3.6]). *Let Y be an ergodic mps, $W_i = Y \ltimes_{\rho_i} H_i$ be ergodic abelian group extensions, $\sigma_i : W_i \to \mathbb{T}$, $i = 1, 2$, and suppose that $\sigma_1\sigma_2$ is a coboundary on $W_1 \times_Y W_2$. Then σ_i are cohomologous to functions on Y.*

Proof. By Lemma 11.7 we may assume that the systems $X_i := W_i \ltimes_{\sigma_i} \mathbb{T}$ are ergodic. By the hypothesis we have

$$TF = F\sigma_1\sigma_2$$

for a function $F : W = W_1 \times_Y W_2 \to \mathbb{T}$. It follows that the function

$$\tilde{F} := F u_1^{-1} u_2^{-1}$$

on $X := X_1 \times_Y X_2$ is invariant, so

$$\tilde{F} = \sum_{j_1, j_2} \vec{\psi}_{j_1} A_{j_1,j_2} \vec{\psi}_{j_2},$$

where the summation indices j_i run over a complete orthogonal sets of irreducible finite rank submodules of $L^2(X_i|Y)$, each submodule j_i has rank d_{j_i} and is spanned by the components of the bounded vector-valued function $\psi_{j_i} : X_i \to \mathbb{C}^{d_{j_i}}$ satisfying

$$T\psi_{j_i} = U_{j_i}\psi_{j_i}$$

with a function $U_{j_i} : Y \to \mathbb{U}(d_{j_i})$, and where $A_{j_1,j_2} : Y \to \mathrm{Mat}(d_{j_1} \times d_{j_2})$ are measurable functions. With the Fourier expansion in the H_i and $\mathbb{T}$ coordinates

$$\psi_{j_i}(y, h_i, u) = \sum_{m \in \mathbb{Z}, \chi \in \widehat{H_i}} \psi_{j_i,m,\chi}(y)\chi(h_i)u^m$$

we obtain $\psi_{j_i,-1,\chi_i} \not\equiv 0$ for each i and some $\chi_i \in \widehat{H_i}$.

On the other hand, we have

$$U_{j_i}(y)\psi_{j_i}(y, h_i, u) = T\psi_{j_i}(y, h_i, u) = \psi_{j_i}(Ty, \rho_i(y)h_i, \sigma_i(y, h_i)u) = \sum_{m,\chi} T\psi_{j_i,m,\chi}(y)\chi(\rho_i(y)h$$

and comparing the Fourier coefficients we obtain

$$U_{j_i}(y)\psi_{j_i,-1}(y, h_i) = T\psi_{j_i,-1}(y, h_i)\sigma_i(y, h_i)^{-1}.$$

□

11.4 Transitivity of G_2

Lemma 11.10. *Let $s \in S^1$ and f, c be a solution to the Conze–Lesigne equation* (11.3)*. Then the transformation*

$$S_{s,f}(x, u) = (sx, f(x)u)$$

belongs to G_2.

Proof. Recall that in course of the reduction to abelian group extensions we have proved

$$\mu^{[2]} = \mu_1^{[2]} \times m_{S^1}^4.$$

From here it is easy to see that $(S_{s,f})_\alpha$ fixes $\mu^{[2]}$ for every side $\alpha \subset V_2$, since it suffices to do so over the factor Z_1.

We claim that the transformation $(S_{s,f})_\alpha$ also leaves the algebra $\mathcal{I}^{[2]}$ invariant. To this end we compute

$$S_{s,f}T(x, u) = S_{s,f}(tx, \rho(x)u) = (stx, f(tx)\rho(x)u) = (stx, \rho(sx)f(x)u/c) = T(sx, f(x)u/c)$$

Let $F \in \mathcal{I}^{[2]}$. The above calculation shows

$$F \circ (S_{s,f})_\alpha \circ T^{[2]} = F \circ T^{[2]} \circ (S_{s,f})_\alpha \circ (c^{-1})_\alpha = F \circ (S_{s,f})_\alpha \circ (c^{-1})_\alpha.$$

Since c^{-1} commutes with $S_{s,f}$, it remains to show $F = F \circ c_\alpha^{-1}$. But this has been proved in the construction of the abelian group extension. □

Note that if s is the identity of the group Z_1, then every constant function f solves the Conze–Lesigne equation.

12 Polynomials in nilpotent groups

12.1 Commutators and filtrations

We use the convention $[a,b]=a^{-1}b^{-1}ab$ for commutators and $a^b=b^{-1}ab$ for conjugation. We will frequently use the following group identities, first used by Hall [Hal33]:

$$[a,bc]=[a,c][a,b]^c \tag{12.1}$$

$$[ab,c]=[a,c]^b[b,c] \tag{12.2}$$

$$[[a,b],c^a][[c,a],b^c][[b,c],a^b]=\mathrm{id} \tag{12.3}$$

Theorem 12.4 (see e.g. [MKS66, Theorem 5.2])**.** *Let G be a group and $A,B,C\trianglelefteq G$ be normal subgroups. Then*

$$[[A,B],C]\leq[[C,A],B][[B,C],A].$$

Proof. In view of (12.2) it suffices to show that for every $a\in A$, $b\in B$, and $c\in C$ the commutator $[[a,b],c]$ is contained in the group on the right. Since $C^a=C$, this follows from (12.3). □

Definition 12.5. Let G be a group. The *lower central series* of G is the sequence of subgroups G_i, $i\in\mathbb{N}$, defined by $G_0=G_1:=G$ and $G_{i+1}:=[G_i,G]$ for $i\geq 1$. The group G is called *nilpotent (of nilpotency class d)* if $G_{d+1}=\{\mathrm{id}\}$.

A *prefiltration* $G_\bullet$ is a sequence of nested groups

$$G_0\geq G_1\geq G_2\geq\dots\quad\text{such that}\quad[G_i,G_j]\subset G_{i+j}\quad\text{for any } i,j\in\mathbb{N}. \tag{12.6}$$

A *filtration* (on a group G) is a prefiltration in which $G_0=G_1$ (and $G_0=G$).

We will frequently write G instead of G_0. Conversely, most groups G that we consider are endowed with a prefiltration $G_\bullet$ such that $G_0=G$. A group may admit several prefiltrations, and we usually fix one of them even if we do not refer to it explicitly.

A prefiltration is said to have *length* $d\in\mathbb{N}$ if G_{d+1} is the trivial group and length $-\infty$ if G_0 is the trivial group. Arithmetic for lengths is defined in the same way as conventionally done for degrees of polynomials, i.e. $d-t=-\infty$ if $d<t$.

Lemma 12.7 (see e.g. [MKS66, Theorem 5.3])**.** *Let G be a group. Then the lower central series $G_\bullet$ is a filtration.*

Proof. The fact that

$$[G_0,G_i]=[G_i,G_0]\subset G_i$$

is equivalent to G_i being normal in G, and this is quickly established by induction on i. This also shows that $G_{i+1}\subseteq G_i$ for all i.

It remains to show that

$$[G_i,G_j]\subseteq G_{i+j}\quad\text{for } i,j\geq 1.$$

To this end use induction on j. For $j=1$ this follows by definition of G_{i+1}, so suppose that the above statement is known for j. Then we have

$$\begin{aligned}[G_i,G_{j+1}]=[G_i,[G_j,G_1]]\subset[[G_1,G_i],G_j][[G_j,G_i],G_i]\\ =[G_{i+1},G_j][G_{i+j},G_1]\subset G_{i+1+j}\end{aligned}$$

by Theorem 12.4 and two applications of the inductive hypothesis. □

Definition 12.8. Let G be a group and $H \subset G$. We write

$$\sqrt[r]{H} := \{g \in G : g^r \in H\} \quad \text{and} \quad \sqrt{H} := \bigcup_{r \in \mathbb{N}_{>0}} \sqrt[r]{H}.$$

The set $\sqrt{H}$ is called the *closure of H* in [BL02].

Lemma 12.9. *Let G be a nilpotent group, $H \leq G$ a finitely generated subgroup, and suppose that G is generated by H and F, where $F \subset \sqrt{H}$ is finite. Then $[G : H] < \infty$.*

Proof. We induct on the nilpotency step d. If $d = 1$, then G is commutative, so H is normal, and we may factor out H. Hence G is generated by finitely many torsion elements, so it is a finite commutative group.

Suppose that the claim is known for groups with nilpotency step $\leq d$ and consider a group G of step $d+1$. Let $G_\bullet$ be the lower central series of G. Then the commutator maps

$$[\cdot, \cdot] : G_1 \times G_d \to G_{d+1},$$

and the image of this map generated G_{d+1}. Since the conjugation action of G on G_{d+1} is trivial, the identities (12.1) and (12.2) and the fact that $G_\bullet$ is a filtration show that the commutator map factors through a bihomomorphism

$$B : G_1/G_2 \times G_d/G_{d+1} \to G_{d+1}.$$

By the inductive hypothesis $H/G_{d+1} \leq G/G_{d+1}$ is a finite index subgroup. In particular, $H/G_2 \leq G_1/G_2$ and $(H \cap G_d)/G_{d+1} \leq G_d/G_{d+1}$ are finite index subgroups, with index a, b, say. Since B is a bihomomorphism, it follows that the power ab of every element in the image of B is contained in $H \cap G_{d+1}$. On the other hand, the image of B generates the group G_{d+1}, and since it has a finite generating subset an by the $d = 1$ case of the lemma the subgroup $H \cap G_{d+1} \leq G_{d+1}$ has finite index.

Since G_{d+1} is central in G, it follows that the group H has finite index in the group H' generated by H and G_{d+1}. Replacing H by H' we can factor out the subgroup G_{d+1} and reduce to step d. □

Corollary 12.10. *Let G be a nilpotent group and $H \leq G$. Then $\sqrt{H}$ is a subgroup of G.*

12.2 Polynomial mappings

In this section we set up the algebraic framework for dealing with polynomials with values in a nilpotent group.

Let $G_\bullet$ be a prefiltration of length d and let $t \in \mathbb{N}$ be arbitrary. We denote by $G_{\bullet+t}$ the prefiltration of length $d - t$ given by $(G_{\bullet+t})_i = G_{i+t}$ and by $G_{\bullet/t}$ the prefiltration of length $\min(d, t-1)$ given by $G_{i/t} = G_i/G_t$ (this is understood to be the trivial group for $i \geq t$; note that G_t is normal in each G_i for $i \leq t$ by (12.6)). These two operations on prefiltrations can be combined: we denote by $G_{\bullet/t+s}$ the prefiltration given by $G_{i/t+s} = G_{i+s}/G_t$, it can be obtained applying first the operation $/t$ and then the operation $+s$ (hence the notation).

We define $G_\bullet$-polynomial maps by induction on the length of the prefiltration.

Definition 12.11. Let $G_\bullet$ be a prefiltration of length $d \in \{-\infty\} \cup \mathbb{N}$. A map $g \colon \mathbb{Z}^r \to G_0$ is called *$G_\bullet$-polynomial* if either $d = -\infty$ (so that g identically equals the identity) or for every $a \in \mathbb{Z}^r$ the map

$$D_a g(n) = g(n)^{-1} T_a g(n) := g(n)^{-1} g(n+a) \tag{12.12}$$

is $G_{\bullet+1}$-polynomial. We write $P(\mathbb{Z}^r, G_\bullet)$ for the set of $G_\bullet$-polynomial maps.

Informally, a map $g : \mathbb{Z}^r \to G_0$ is polynomial if every discrete derivative Dg is polynomial "of lower degree" (the "degree" of a $G_\bullet$-polynomial map would be the length of the prefiltration $G_\bullet$, but we prefer not to use this notion since it is necessary to keep track of the prefiltration $G_\bullet$ anyway).

Note that if a map g is $G_\bullet$-polynomial then the map gG_t is $G_{\bullet/t}$-polynomial for any $t \in \mathbb{N}$ (but not conversely). We abuse the notation by saying that g is $G_{\bullet/t}$-polynomial if gG_t is $G_{\bullet/t}$-polynomial. In assertions that hold for all $a \in \mathbb{Z}^r$ we omit the subscript in D_a, T_a.

The next theorem is the basic result about $G_\bullet$-polynomials.

Theorem 12.13. *For every prefiltration $G_\bullet$ of length $d \in \{-\infty\} \cup \mathbb{N}$ the following holds.*

1. *Let $t_i \in \mathbb{N}$ and $g_i \colon \mathbb{Z}^r \to G$ be maps such that g_i is $G_{\bullet/(d+1-t_{1-i})+t_i}$-polynomial for $i = 0, 1$. Then the commutator $[g_0, g_1]$ is $G_{\bullet+t_0+t_1}$-polynomial.*

2. *Let $g_0, g_1 \colon \mathbb{Z}^r \to G$ be $G_\bullet$-polynomial maps. Then the product $g_0 g_1$ is also $G_\bullet$-polynomial.*

3. *Let $g \colon \mathbb{Z}^r \to G$ be a $G_\bullet$-polynomial map. Then its pointwise inverse g^{-1} is also $G_\bullet$-polynomial.*

Proof. We use induction on d. If $d = -\infty$, then the group G_0 is trivial and the conclusion hold trivially. Let $d \geq 0$ and assume that the conclusion holds for all smaller values of d.

We prove part (1) using descending induction on $t = t_0 + t_1$. We clearly have $[g_0, g_1] \subset G_t$. If $t \geq d + 1$, there is nothing left to show. Otherwise it remains to show that $D[g_0, g_1]$ is $G_{\bullet+t+1}$-polynomial. To this end we use the commutator identity

$$D[g_0, g_1] = [g_0, Dg_1] \cdot [[g_0, Dg_1], [g_0, g_1]] \\ \cdot [[g_0, g_1], Dg_1] \cdot [[g_0, g_1 Dg_1], Dg_0] \cdot [Dg_0, g_1 Dg_1]. \quad (12.14)$$

We will show that the second to last term is $G_{\bullet+t+1}$-polynomial, the argument for the other terms is similar. Note that Dg_0 is $G_{\bullet/(d+1-t_1)+t_0+1}$-polynomial. By the inner induction hypothesis it suffices to show that $[g_0, g_1 Dg_1]$ is $G_{\bullet/(d-t_0)+t_1}$-polynomial. But the prefiltration $G_{\bullet/(d-t_0)}$ has smaller length than $G_\bullet$, and by the outer induction hypothesis we can conclude that $g_1 Dg_1$ is $G_{\bullet/(d-t_0)+t_1}$-polynomial. Moreover, g_0 is clearly $G_{\bullet/(d-t_0-t_1)}$-polynomial, and by the outer induction hypothesis its commutator with $g_1 Dg_1$ is $G_{\bullet/(d-t_0)+t_1}$-polynomial as required.

Provided that each multiplicand in (12.14) is $G_{\bullet+t+1}$-polynomial, we can conclude that $D[g_0, g_1]$ is $G_{\bullet+t+1}$-polynomial by the outer induction hypothesis.

Part (2) follows immediately by the Leibniz rule

$$D(g_0 g_1) = Dg_0 [Dg_0, g_1] Dg_1 \quad (12.15)$$

from (1) with $t_0 = 1$, $t_1 = 0$ and the induction hypothesis.

To prove part (3) notice that

$$D(g^{-1}) = g(Dg)^{-1} g^{-1} = [g^{-1}, Dg](Dg)^{-1}. \quad (12.16)$$

By the induction hypothesis the map g^{-1} is $G_{\bullet/d}$-polynomial, the map Dg is $G_{\bullet+1}$-polynomial, and the map $(Dg)^{-1}$ is $G_{\bullet+1}$-polynomial. Thus also $D(g^{-1})$ is $G_{\bullet+1}$-polynomial by (1) and the induction hypothesis. □

Discarding some technical information that was necessary for the inductive proof we can write the above theorem succinctly as follows.

Corollary 12.17 ([Lei02, Proposition 3.7]). *Let $G_\bullet$ be a prefiltration of length d. Then the set $P(\mathbb{Z}^r, G_\bullet)$ of $G_\bullet$-polynomials on $\mathbb{Z}^r$ is a group under pointwise operations and admits a canonical prefiltration of length d given by*

$$P(\mathbb{Z}^r, G_\bullet) \geq P(\mathbb{Z}^r, G_{\bullet+1}) \geq \cdots \geq P(\mathbb{Z}^r, G_{\bullet+d+1}).$$

Remark. In [Lei02] a polynomial has a "vector degree" that is given by a sequence $\bar{d} = (d_i)_{i\in\mathbb{N}} \subset \mathbb{N}$ that is superadditive in the sense that $d_{i+j} \geq d_i + d_j$ for all $i, j \in \mathbb{N}$; by convention $d_{-1} = -\infty$. This is included in our treatment: a map has vector degree $\bar{d}$ with respect to a prefiltration $G_\bullet$ if and only if it is $G^{\bar{d}}_\bullet$-polynomial, where the prefiltration $G^{\bar{d}}_\bullet$ is given by

$$G^{\bar{d}}_i = G_j \quad \text{whenever} \quad d_{j-1} < i \leq d_j. \tag{12.18}$$

Remark. Variants of the above definition of polynomials include prefiltrations indexed by partially ordered semigroups more general than the natural numbers $\mathbb{N} = \{0, 1, \dots\}$, see [GTZ12, Appendix B].

12.3 Integer Lagrange interpolation

A polynomial of degree d on $\mathbb{Z}$ is determined by its values at $0, \dots, d$. Similiarly, a polynomial of degree d on $\mathbb{Z}^r$ is determined by its values on the set

$$\Delta_{r,d} := \{k \in \mathbb{Z}^r : k_i \geq 0, \sum_{i=1}^{r} k_i \leq d\}.$$

This is proved in two steps: firstly, any polynomial of degree $\leq d$ that vanishes on $\Delta_{r,d}$ vanishes everywhere. Secondly, the dimension of the space of polynomials of deree $\leq d$ has dimension $|\Delta_{r,d}|$, so every function on $\Delta_{r,d}$ can be interpolated by a polynomial of degree $\leq d$. Also, a polynomial maps integer points to integers iff its restriction to $\Delta_{r,d}$ does.

Similar results hold for polynomials in nilpotent groups.

Lemma 12.19. *Let $G_\bullet$ be a filtration of length $\leq d$, $g \in P(\mathbb{Z}^r, G_\bullet)$, and suppose that g vanishes on $\Delta_{m,d}$. Then g vanishes identically.*

Proof. By induction on d. In the case $d = 0$ the map g is constant, and since it vanishes at 0 it vanishes identically.

Suppose that the claim holds for d and consider it with d replaced by $d+1$. Then for every basis vector $e_i \in \mathbb{Z}^r$ the derivative $D_{e_i} g$ vanishes on $\Delta_{m,d+1} \cap \Delta_{m,d+1} - e_i \supset \Delta_{m,d}$. By the inductive hypothesis the derivative vanishes identically, and the claim follows. □

Lemma 12.20. *Let $G_\bullet$ be a filtration of length $\leq d$ and $g \in P(\mathbb{Z}^r, G_\bullet)$. Then we can write*

$$g = \prod_{a\in\Delta_{r,d}} g_a^{\binom{n}{a}}, \quad \text{where} \quad \binom{n}{a} = \prod_{j=1}^{r} \binom{n_j}{a_j}, \quad g_a \in G_{|a|},$$

the product taken e.g. in increasing lexicographic order with top level ordering by $|a| = \sum_j a_j$.

Proof. This follows from the following inductive claim.

Claim 12.21. Let $g \in P(\mathbb{Z}^r, G_\bullet)$ vanish on $\Delta_{r,l-1}$. Then $g(n) = \prod_{a\in\Delta_{r,l}\setminus\Delta_{r,l-1}} g(a)^{\binom{n}{a}} \tilde{g}(n)$, where $\tilde{g} \in P(\mathbb{Z}^r, G_\bullet)$ vanishes on $\Delta_{r,l}$.

To show the claim note that gG_{l+1} is $G_{\bullet/l}$-polynomial, hence it vanishes identically by Lemma 12.19. Therefore $g(a) \in G_l$, so that the maps $n \to g(a)^{\binom{n}{a}}$ are $G_\bullet$-polynomial. Moreover, if $\sum_j a_j = l$, then

$$\binom{n}{a} = \begin{cases} 1, & n = a \\ 0, & n \in \Delta_{r,l}, n \neq a. \end{cases}$$

It follows that the remainder term $\tilde{g}$ is $G_\bullet$-polynomial and vanishes on $\Delta_{r,l}$. □

Despite the fact that every polynomial can be written in such explicit form, it is usually more convenient to use the abstract definition, particularly when passing between different prefiltrations.

12.4 Commensurable lattices

We summarize the properties of nilmanifolds that will be necessary for our discussion. For our purpose one can think of these properties as being provided by the structure theorem for Host–Kra factors.

Definition 12.22. A *nilmanifold* consists of the following pieces of information:

1. a nilpotent Lie group G,
2. a prefiltration $G_\bullet$ on G consisting of closed (Lie) subgroups, and
3. a finitely generated discrete subgroup $\Gamma \leq G$.

We assume that the homogeneous space G_i/Γ is compact for every group G_i in the prefiltration $G_\bullet$. We call a group Γ as above a *lattice* and write $\Gamma_i = \Gamma \cap G_i$. A closed subgroup $\tilde{G} \leq G$ such that $\tilde{G}/\Gamma$ is compact is called Γ-rational.

Recall that two subgroups $A, B \leq G$ are called *commensurable* if $A \cap B$ has finite index in both A and B.

Lemma 12.23. *Let G/Γ be a nilmanifold and $\tilde{\Gamma} \leq G$ be a group that is commensurable with Γ. Then the following assertions hold.*

1. *$\tilde{\Gamma}$ is also a discrete cocompact subgroup.*
2. *Every Γ-rational subgroup $G' \leq G$ is also $\tilde{\Gamma}$-rational.*

Proof. To see (1) note that if $\tilde{\Gamma} \leq \Gamma$, then the natural map $G/\tilde{\Gamma} \to G/\Gamma$ is a covering map with finitely many sheets, and it follows that $G/\tilde{\Gamma}$ is compact. If $\Gamma \leq \tilde{\Gamma}$, then $G/\tilde{\Gamma}$ is a quotient space of G/Γ, so it is clearly compact. From this it follows that $\tilde{\Gamma}$ is cocompact in general. Also, it is clear that $\tilde{\Gamma}$ is discrete if and only if Γ is discrete.

The assertion (2) follows since the groups $\Gamma \cap G'$ and $\tilde{\Gamma} \cap G'$ are commensurable whenever Γ and $\tilde{\Gamma}$ are commensurable. □

An important class of examples of commensuarble lattices arises when one needs to replace a nilmanifold by a connected one.

Lemma 12.24. *Let G/Γ be a nilmanifold. Then there exists a lattice $\Gamma \leq \tilde{\Gamma} \leq G$ such that Γ has finite index in $\tilde{\Gamma}$ and $G_i/\tilde{\Gamma}_i$ is connected for every i.*

Here and later we denote the connected component of the identity in a group G by G^o.

Proof. We use induction on the length of the filtration. If $G_\bullet$ is trivial, then there is nothing to show, so suppose that the conclusion holds for filtrations of length $d-1$ and consider a Γ-rational filtration $G_\bullet$ of length d.

By the rationality assumption we can write $G_d = G_d^o \oplus A$ in such a way that $\Gamma \cap A \leq A$ is a finite index subgroup. Since A is central in G, this implies that Γ has finite index in ΓA. Replacing Γ by ΓA if necessary, we may assume that $\Gamma G_d = \Gamma G_d^o$.

By the inductive assumption $\Gamma G_d/G_d$ is a finite index subgroup of a lattice $\tilde{\Gamma}_{/d}$ such that $(G_i/G_d)/\tilde{\Gamma}_{/d}$ is connected for every i. Let $\{\tilde{\gamma}_j\} \subset G/G_d$ be a finite set that together with $\Gamma G_d/G_d$ generates $\tilde{\Gamma}_{/d}$. We can write $\tilde{\gamma}_j = g_j G_d$, and we have $g_j^r \in \Gamma G_d$ for some r and all j. Now recall that $\Gamma G_d = \Gamma G_d^o$ and that in the connected commutative Lie group G_d^o arbitrary roots exist. Hence, multiplying g_j by an element of G_d^o if necessary, we may assume that $g_j^r \in \Gamma$.

By Lemma 12.9 Γ has finite index in the group generated by Γ and the elements g_j. It remains to show that $G_i/\tilde{\Gamma}_i$ is connected for every i. We have an exact sequence of topological spaces

$$G_d/\tilde{\Gamma} \to G_i/\tilde{\Gamma} \to G_i/\tilde{\Gamma} G_d$$

and the outer two are connected. □

12.5 Reduction of polynomials to connected Lie groups

The next step is vaguely parallel to separating rational and irrational coefficients of a polynomial over the reals. Given a prefiltration $G_\bullet$, we define a prefiltration $G_\bullet^o$ by $(G^o)_i = (G_i)^o$.

Lemma 12.25. *Let G/Γ be a nilmanifold such that G_i/Γ_i is connected for each i. Then every $G_\bullet$-polynomial sequence $g(n)$ can be written in the form*

$$g(n) = g^o(n)\gamma(n),$$

where g^o is a $G_\bullet^o$-polynomial sequence, and γ is a $\Gamma_\bullet$-polynomial sequence.

Proof. It suffices to show the following:

Claim 12.26. Let $g \in P(\mathbb{Z}^r, G_\bullet)$ be a polynomial that vanishes on $\Delta_{r,l-1}$. Then we can factorize

$$g = g^o \tilde{g} \gamma,$$

where g^o is $G_\bullet^o$-polynomial, γ is $\Gamma_\bullet$-polynomial, and $\tilde{g}$ is $G_\bullet$-polynomial and vanishes on $\Delta_{r,l}$.

The group G_l/G_{l+1} is a commutative Lie group, and its quotient modulo Γ is a compact connected space. Hence $G_l/G_{l+1} = G_l^o/G_{l+1} \times A$, where A is a discrete subgroup contained in Γ/G_{l+1}.

The map gG_l is $G_{\bullet/l}$-polynomial, so by Lemma 12.19 it vanishes identically. Hence g takes values in G_l. By the hypothesis that G_l/Γ is connected we have $G_l = G_l^o \Gamma_l$. For every $a \in \Delta_{r,l} \setminus \Delta_{r,l-1}$ write $g(a) = g_a \gamma_a$ with $g_a \in G_l^o$, $\gamma_a \in \Gamma_l$. Define

$$g^o(n) := \prod_{a \in \Delta_{r,l} \setminus \Delta_{r,l-1}} g_a^{\binom{n}{a}}, \quad \gamma(n) := \prod_{a \in \Delta_{r,l} \setminus \Delta_{r,l-1}} \gamma_a^{\binom{n}{a}},$$

the product being taken in any fixed order, say lexicographic. The polynomials $n \mapsto \binom{n}{a}$ have degree l and $g_a, \gamma_a \in G_l$, so these maps are in fact polynomial with respect to the required filtrations. Also, $\tilde{g} = (g^o)^{-1} g(\gamma)^{-1}$ vanishes on $\Delta_{r,l}$ by construction. □

On general nilmanifolds we can use Lemma 12.25 together with Lemma 12.24 and obtain a splitting in which the second factor takes values in a group $\tilde{\Gamma} \geq \Gamma$ in which Γ has finite index. The next lemma shows that the latter factor is periodic modulo Γ; this can be compared to the fact that rational polynomials are periodic modulo $\mathbb{Z}$.

Lemma 12.27. *Let G be a nilpotent group with a prefiltration $G_\bullet$ and let $\Gamma \leq G$ be a finite index subgroup. Then for every $G_\bullet$-polynomial sequence $g(n)$ the sequence $g(n)\Gamma$ is periodic (that is, constant on cosets of a finite index subgroup of $\mathbb{Z}^r$).*

Proof. Replacing Γ by a finite index subgroup that is normal in G and working modulo Γ, we may assume that G is finite and Γ is trivial.

We factorize g as in Lemma 12.20 and observe that each term in the factorization is periodic. □

13 Equidistribution criterion for polynomials on nilmanifolds

13.1 Cube group

We outline a special case of the cube construction of Green, Tao, and Ziegler [GTZ12, Definition B.2] using notation of Green and Tao [GT12, Proposition 7.2]. We will only have to perform it on filtrations, but even in this case the result is in general only a prefiltration.

Definition 13.1 (Cube filtration)**.** Given a prefiltration $G_\bullet$ we define the prefiltration $G_\bullet^\square$ by

$$G_i^\square := \left\langle G_i^\triangle, G_{i+1} \times G_{i+1} \right\rangle$$

where $G^\triangle = \{(g_0, g_1) \in G^2 : g_0 = g_1\}$ is the diagonal group corresponding to G. By an abuse of notation we refer to the filtration obtained from $G_\bullet^\square$ by replacing $G_0^\square$ with $G_1^\square$ as the "filtration $G_\bullet^\square$".

To see that this indeed defines a prefiltration note first that $G_i^\square$ is normal in $G_0^\square$. Using this and Hall identities it suffices to verify the commutator property on generators, which is straightforward.

Lemma 13.2 (Rationality of the cube filtration)**.** *Let G/Γ be a nilmanifold. Then $G^\square/\Gamma^2$ is a nilmanifold (where $G^\square = G_1^\square$).*

Proof. We have to verify that $G_i^\square/\Gamma^2$ is compact for every i. We induct on the length d of the filtration $G_\bullet$.

The group $G_d^\square$ is just the diagonal group, so its quotient modulo Γ^2 is compact by the hypothesis. Let $i < d$. On the quotient space $G_i^\square/\Gamma^2$ we have a continuous action of the compact abelian group $(G_d/\Gamma)^2$. The quotient of $G_i^\square/\Gamma^2$ modulo this action is compact by the inductive hypothesis, and the claim follows. □

Lemma 13.3. *Let $g \in P(\mathbb{Z}^r, G_\bullet)$. Then for every $k \in \mathbb{Z}^r$ the map*

$$g_k^{\tilde{\square}}(n) := (g(n+k), g(n))$$

is $G_\bullet^\square$-polynomial.

Proof. We use induction on the length l of the prefiltration $G_\bullet$. Indeed, for $l = -\infty$ there is nothing to show. If $l \geq 0$, then $g_k^{\tilde{\square}}$ takes values in $G_0^\square$ since $g(n)^{-1}g(n+k) = D_k g(n) \in G_1$ by definition of a polynomial. Moreover $D_{k'}(g_k^{\tilde{\square}}) = (D_{k'}g)_k^{\tilde{\square}}(n)$, so that $D_{k'}(g_k^{\tilde{\square}})$ is $G_{\bullet+1}^\square$-polynomial by the induction hypothesis. □

13.2 Vertical characters

Let G/Γ be a nilmanifold of nilpotency class l. Then G/Γ is a smooth principal bundle with the compact commutative Lie structure group G_l/Γ_l. The fibers of this bundle are called "vertical" tori.

Definition 13.4 (Vertical character)**.** Let G/Γ be a nilmanifold of nilpotency class l. A measurable function F on G/Γ is called a *vertical character* if there exists a character $\chi \in \widehat{G_l/\Gamma_l}$ such that for every $g_l \in G_l$ and a.e. $y \in G/\Gamma$ we have $F(g_l y) = \chi(g_l\Gamma_l)F(y)$.

Definition 13.5 (Vertical Fourier series)**.** Let G/Γ be a nilmanifold of nilpotency class l. For every $F \in L^2(G/\Gamma)$ and $\chi \in \widehat{G_l/\Gamma_l}$ let

$$F_\chi(y) := \int_{G_l/\Gamma_l} F(g_l y)\overline{\chi}(g_l)\mathrm{d}g_l. \tag{13.6}$$

With this definition F_χ is defined almost everywhere and is a vertical character as witnessed by the character χ. The usual Fourier inversion formula implies that $F = \sum_{\chi \in \widehat{G_l/\Gamma_l}} F_\chi$ in $L^2(G/\Gamma)$.

Remark. The correct analog of the Plancherel identity for vertical Fourier series reads

$$\sum_\chi \|F_\chi\|_{U^l(G/\Gamma)}^{2^l} = \|F\|_{U^l(G/\Gamma)}^{2^l},$$

where U^l stands for appropriate Gowers–Host–Kra seminorms, see [ET12, Lemma 10.2] for the case $l = 3$.

13.3 Fractional part map

Lemma 13.7 (Fundamental domain)**.** *Let $\Gamma \leq G$ be a cocompact lattice. Then there exists a relatively compact set $K \subset G$ and a map $G \to K$, $g \mapsto \{g\}$ such that $g\Gamma = \{g\}\Gamma$ and $\{\{g\}\} = \{g\}$ for each $g \in G$.*

This follows readily from local homeomorphy of G and G/Γ, from local compactness of G and from compactness of G/Γ. For example, for $G = \mathbb{R}$ and $\Gamma = \mathbb{Z}$ the fundamental domain K can be taken to be the interval $[0, 1)$ with the usual fractional part map $\{\cdot\}$. In case of a general connected Lie group the fundamental domain can be taken to be $[0, 1)^d$ in Mal'cev coordinates [GT12, Lemma A.14], but we do not need this information.

For each nilmanifold that we consider we fix some map $\{\cdot\}$ as above and write $g = \{g\}\lfloor g\rfloor$ with $\lfloor g\rfloor \in \Gamma$.

13.4 Linear equidistribution criterion

We will use the following notation for multiparameter averages. A box in $\mathbb{Z}^r$ is denoted by the letter I. We write $\mathrm{Av}_{n\in I} = |I|^{-1} \sum_{n\in I}$ and write $\lim_I$ for the limit as the minimal side length of the box goes to ∞ (similarly for lim inf and lim sup).

Lemma 13.8. *Let $\delta > 0$ and $g : \mathbb{Z}^r \to \mathbb{U}^s$ (a torus, $\mathbb{U} = \mathbb{R}/\mathbb{Z}$) be a linear sequence. Suppose that there exists a Lipschitz function $F : \mathbb{U}^s \to \mathbb{R}$ such that*

$$\limsup_I \Big| \mathrm{Av}_{n\in I} F(g(n)) - \int F \Big| > \delta \|F\|_{\mathrm{Lip}}.$$

Then there exists a $0 \neq k \in \mathbb{Z}^s$, $k = O(\delta^{-1})$, such that $k \cdot g(n) \equiv \mathrm{const}$.

Proof. Replacing F by $(F - \int F)/\|F - \int F\|_{\mathrm{Lip}}$ we may assume $\int F = 0$, $\|F\|_{\mathrm{Lip}} = 1$. We may also assume $\delta \leq 1$. Let

$$F' := |\delta/10m|^{-2} F * 1_{\frac{\delta}{10m}[-1,1]^m} * 1_{\frac{\delta}{10m}[-1,1]^m}.$$

Then $\|F' - F\|_\infty < \|F\|_{\mathrm{Lip}}\delta/10$, hence

$$\limsup_I \Big| \mathrm{Av}_{n\in I} F'(g(n)) \Big| > \delta/2.$$

On the other hand,

$$|\widehat{F'}(\xi)| = |\hat{F}(\xi)| \cdot ||\widehat{\delta/10m|^{-1} 1_{\frac{\delta}{10m}[-1,1]^m}}(\xi)|^2 \lesssim (1 + |\xi|/\delta)^{-2},$$

so truncating F' in frequency at C/δ we obtain

$$\limsup_I \Big| \mathrm{Av}_{n\in I} F''(g(n)) \Big| > \delta/4$$

with a function F'' whose Fourier transform is supported in a cube with side length C/δ, bounded by 1, and vanishes at 0. By the pigeonhole principle it follows that

$$\limsup_I \Big| \mathrm{Av}_{n\in I}\, e(k\cdot g(n)) \Big| > C\delta^{s+1}$$

with some k as in the conclusion of the lemma. It remains to observe that $\lim_I \mathrm{Av}_{n\in I}\, e(k\cdot g(n)) = 0$ unless $k\cdot g(n) \equiv \mathrm{const}$ (this is a linear exponential sum that can be computed explicitly). ☐

13.5 Polynomial equidistribution criterion

The following multiparameter version of the van der Corput inequality is proved in exactly the same way as the one-parameter version.

Proposition 13.9. *Let V be a Hilbert space and let $(v_n)_{n\in\mathbb{Z}^r}$ be a bounded sequence in V. Then*

$$\limsup_I \| \mathrm{Av}_{n\in I}\, v_n \|^2 \le \liminf_{I'} \mathrm{Av}_{k\in I'} \limsup_I | \mathrm{Av}_{n\in I} \langle v_{n+k}, v_n\rangle |.$$

Leibman's equidistribution criterion [Lei05] tells that the only obstruction to equidistribution of $G_\bullet$-polynomial sequences on a connected nilmanifold G/Γ are *characters*, that is, continuous homomorphisms $\eta : G \to \mathbb{U}$ that vanish on Γ:

Theorem 13.10. *Let G/Γ be a nilmanifold associated to a connected group G and $G_\bullet$ a Γ-rational filtration on G. Let $F \in C(G/\Gamma)$, $\Lambda \le \mathbb{Z}^r$ a finite index subgroup, and $\delta > 0$. Then there exists a finite set of non-trivial characters on G such that for every $g \in P(\mathbb{Z}^r, G_\bullet)$ with*

$$\limsup_{I\subset\Lambda+\rho} \Big| \mathrm{Av}_{n\in I}\, F(g(n)\Gamma) - \int_{G/\Gamma} F \Big| > \delta$$

for some $\rho \in \mathbb{Z}^r$ there exists a character η on this list such that $\eta\circ g = \mathrm{const}$.

We will give a qualitative version of the proof that is due to Green and Tao [GT12; GT14]. The proof proceeds by induction on the length of the filtration and on $\dim G_2$. In each step one performs the cube construction and factors out the diagonal central subgroup. Uniformity over all polynomials is essential for inductive purposes.

The connectedness hypothesis is needed in order to ensure that any non-zero multiple of a non-trivial character is again a non-trivial character; this observation will be used without further reference.

Proof. Replacing F by $F - \int F$ we may assume $\int F = 0$. First we reduce to the case that $G_\bullet$ consists of connected groups, $\Lambda = \mathbb{Z}^r$, and $g(0) = \mathrm{id}$. To this end we split $g = g^o\gamma$, where g^o is $G^o_\bullet$-polynomial and γ is $\tilde\Gamma_\bullet$-polynomial for some finite index surgroup $\tilde\Gamma \ge \Gamma$ that does not depend on g. In particular, $\gamma\Gamma$ is periodic with period $\Lambda' \le \Lambda \le \mathbb{Z}^d$ that does not depend on g. By the pigeonhole principle there is a coset $\rho' + \Lambda'$ such that

$$\limsup_{I\subset\Lambda'+\rho'} | \mathrm{Av}_{n\in I}\, F(g(n)\Gamma)| > \delta.$$

This can be wriiten as

$$\limsup_{I\subset\Lambda'} | \mathrm{Av}_{n\in I}\, F(\{g(\rho)\}\{g(\rho)\}^{-1} g(n+\rho)\lfloor g(\rho)\rfloor^{-1}\Gamma)| > \delta.$$

Since $\{g(\rho)\}$ lies in a fixed compact set, the set of functions $x \mapsto F(\{g(\rho)\}x)$ is compact, so it can be covered by finitely many balls of radius $\delta/2$, the covering being independent of g. Hence we may assume $g(\rho) = \mathrm{id}$. Applying the connected case to the group $\Lambda' \cong \mathbb{Z}^r$ we obtain a character such that $\eta(\{g(\rho)\}^{-1} g(\Lambda'+\rho)\lfloor g(\rho)\rfloor^{-1}) = \mathrm{const}$, and therefore $\eta(g(\Lambda'+\rho)) = \text{const}$. It follows that $\eta \circ g \equiv \text{const} \mod \frac{1}{R}$ for some bounded R, and the claim follows with η replaced by $R\eta$. This completes the reduction.

It remains to prove the conclusion under the additional assumptions that $G_\bullet$ consists of connected groups, $\Lambda = \mathbb{Z}^r$, and $g(0) = \mathrm{id}$. Replacing g by the sequence

$$g(n) \prod_{i=1}^{r} \lfloor g(e_i) \rfloor^{-n_i}$$

we may also assume that $g(e_i) = \{g(e_i)\}$. Write

$$g_{\mathrm{lin}}(n) = \prod_{i=1}^{r} g(e_i)^{n_i}, \quad g_{\mathrm{nlin}}(n) = g_{\mathrm{lin}}(n)^{-1} g(n),$$

so that g_{nlin} takes values in G_2. By uniform approximation we may assume that F is smooth.

The case $l = 1$ is contained in Lemma 13.8.

Suppose now that $l \geq 2$. Analogously to the commutative case, smoothness implies that the vertical Fourier series $F = \sum_\chi F_\chi$ (Definition 13.5) converges absolutely, so, decreasing δ if necessary, we can assume that F has a vertical frequency χ. If this frequency vanishes, then we can factor out G_l and use induction on the length of filtration.

Assume now that the vertical frequency χ is non-trivial. By the van der Corput difference lemma the set of h such that

$$\limsup_I |\mathrm{Av}_{n\in I} F(g(h+n)\Gamma)\overline{F(g(n)\Gamma)}| > \delta^2$$

has positive lower Banach density (bounded below by a constant depending only on δ and F). We write

$$F(g(h+n)\Gamma)\overline{F(g(n)\Gamma)} = \underbrace{F \otimes \bar{F}((\{g_{\mathrm{lin}}(h)\}, \mathrm{id})}_{=:F^\square_{g,h}} \underbrace{(\{g_{\mathrm{lin}}(h)\}^{-1} g(h+n) \lfloor g_{\mathrm{lin}}(h) \rfloor^{-1}, g(n))}_{=:g^\square_h(n)} \Gamma^2)$$

Since the fractional part function $\{\cdot\}$ has relatively compact range, the set of functions $F^\square_{g,h}$ is relatively compact. Choosing a g-independent $\delta^2/2$-dense subset and pigeonholing we obtain a function $F^\square$ that does not depend on h such that

$$\limsup_I |\mathrm{Av}_{n\in I} F^\square(g^\square_h(n))| > \delta^2/2$$

for a set of h of positive lower Banach density. Note that $F^\square$ has a non-trivial vertical frequency with respect to G_l^2 and is G_l^Δ-invariant. Hence, factoring out G_l^Δ, we see that $g^\square_h$ is polynomial with respect to the filtration $G^\square_\bullet/G_l^\Delta$ that has length $l-1$ and $F^\square$ has zero integral on $G^\square_1/G^\Delta_l\Gamma^2$.

By the induction hypothesis we obtain a finite list of characters $\eta : G^\square/G^\Delta_l \to \mathbb{U}$ such that for each h in our positive lower Banach density set there exists a character on this list such that $\eta \circ g^\square_h$ vanishes. By the pigeonhole principle we may assume that the character η does not depend on h. Write

$$\begin{aligned}
\eta(g^\square_h(n)) &= \eta(\{g_{\mathrm{lin}}(h)\}^{-1} g(h+n) \lfloor g_{\mathrm{lin}}(h) \rfloor^{-1}, g(n)) \\
&= \eta(g(n), g(n)) + \eta(\{g_{\mathrm{lin}}(h)\}^{-1} g(h+n) \lfloor g_{\mathrm{lin}}(h) \rfloor^{-1} g(n)^{-1}, \mathrm{id}) \\
&= \eta_1(g(n)) + \eta_2(\{g_{\mathrm{lin}}(h)\}^{-1} g(h+n) \lfloor g_{\mathrm{lin}}(h) \rfloor^{-1} g(n)^{-1}),
\end{aligned}$$

where $\eta_1 : G_1 \to \mathbb{U}, g \mapsto \eta(g,g)$ and $\eta_2 : G_2 \to \mathbb{U}, g \mapsto \eta(g, \mathrm{id})$ are characters that vanish on the normal subgroups $[G,G]$ and $[G,G_2]$, respectively, and on Γ. If the character η_2 is trivial, then η_1 is non-trivial and we obtain the conclusion with the character η_1. Hence we may assume that η_2 is non-trivial.

Note

$$\begin{aligned}
\eta_2(\{g_{\text{lin}}(h)\}^{-1}g(h+n)\lfloor g_{\text{lin}}(h)\rfloor^{-1}g(n)^{-1}) \\
&= \eta_2(\{g_{\text{lin}}(h)\}^{-1}g(h+n)g(h)^{-1}\{g_{\text{lin}}(h)\}g(n)^{-1}) \\
&= \eta_2([\{g_{\text{lin}}(h)\}, g_{\text{lin}}(h)g(h+n)^{-1}]g(h+n)g_{\text{lin}}(h)^{-1}g(n)^{-1}) \\
&= \eta_2([\{g(h)\}, g_{\text{lin}}(h)g(h+n)^{-1}]) + \eta_2(g(h+n)g_{\text{lin}}(h)^{-1}g(n)^{-1}) \\
&= -\sum_{j=1}^{r} n_j\eta_2([\{g(h)\}, g(e_j)]) + \eta_2(g_{\text{lin}}(h+n)g_{\text{lin}}(h)^{-1}g_{\text{lin}}(n)^{-1}) \\
&\qquad + \eta_2(g_{\text{nlin}}(h+n)) - \eta_2(g_{\text{nlin}}(n)).
\end{aligned}$$

(Here and later we repeatedly use that the commutator induces an antisymmetric bihomomorphism $G/G_2 \times G/G_2 \to G_2/[G, G_2]$). In the second term we note

$$\begin{aligned}
g_{\text{lin}}(n)g_{\text{lin}}(h) &\equiv g_{\text{lin}}(n+h)\prod_{i<j}[g(e_j)^{h_j}, g(e_i)^{n_i}] \mod [G, G_2] \\
&\equiv g_{\text{lin}}(n+h)\prod_{i<j}[g(e_j), g(e_i)]^{h_j n_i} \mod [G, G_2],
\end{aligned}$$

so

$$\eta_2(g_{\text{lin}}(h+n)g_{\text{lin}}(h)^{-1}g_{\text{lin}}(n)^{-1}) = -\eta_2(\prod_{i<j}[g(e_j), g(e_i)]^{h_j n_i}) = \sum_{i<j} n_i h_j \eta_2([g(e_i), g(e_j)]).$$

Overall we obtain that for a bounded below density set of $h \in \mathbb{Z}^r$ and every $n \in \mathbb{Z}^r$ we have

$$P(n) + Q(n+h) - Q(n) + \sum_{i=1}^{r} \sigma_i(h)n_i = 0_{\mathbb{U}} \tag{13.11}$$

with $P(n) = \eta_1(g(n))$, $Q(n) = \eta_2(g_{\text{nlin}}(n))$, and

$$\sigma_i(h) = -\eta_2([\{g(h)\}, g(e_i)]) + \sum_{i<j} h_j\eta_2([g(e_i), g(e_j)]).$$

Since this holds for a positive density set of h, this holds in particular for h, h' with $h' = h + \delta_i e_i$ and bounded δ_i for all $i = 1, \dots, r$. Hence some bounded discrete derivatives of Q have degree 1 modulo $\mathbb{Z}$. By integer Lagrange interpolation these discrete derivatives have integer coefficients of orders > 1, so the coefficients of Q of order > 2 are rational with bounded denominator. A fortiori, the coefficients of P of order > 1 are rational with bounded denominator. Moreover, by construction P has no constant term and Q has no constant and no linear terms. Multiplying η by a bounded non-zero integer we may assume $\deg Q \le 2$, $\deg P \le 1$. It follows that

$$\sum_{i=1}^{r} \frac{1}{2} q_{ii}h_i^2 + \sum_{i=1}^{r}(p_i + \sigma_i(h) + \sum_{j=1}^{r} q_{ij}h_j)n_i = 0_{\mathbb{U}}, \tag{13.12}$$

where p_i are linear coefficients of P and q_{ij} are quadratic coefficients of Q (multiplied by 2 in the case $i = j$). Since this holds for all n, we have

$$p_i + \sigma_i(h) + \sum_{j=1}^{r} q_{ij}h_j = 0_{\mathbb{U}}. \tag{13.13}$$

At this point we start expanding the definition of σ. The group G/G_2 is a connected commutative Lie group, so there is an isomorphism $\psi : G/G_2 \to \mathbb{R}^s/\mathbb{Z}^{s'} \times \{0\}$,

$s' \leq s = \dim G - \dim G_2$. We may assume that $\psi(\Gamma/G_2) = \mathbb{Z}^s/\mathbb{Z}^{s'}$ and that the fractional part map on G coincides modulo G_2 with the usual coordinatewise fractional part map. We lift ψ to a map with codomain $\mathbb{R}^s$. In coordinates $\eta_2([\cdot,\cdot])$ is given by an antisymmetric bilinear form with integer coefficients, $\eta_2([x,y]) = \psi(x)A\psi(y)$ (that is well-defined modulo $\mathbb{Z}^{s'}$ in both arguments).

$$\begin{tikzcd}
 & G \times G \arrow[r, "{[\cdot,\cdot]}"] \arrow[dl, "\psi\times\psi"'] \arrow[d] & G_2 \arrow[r, "\eta_2"] \arrow[d] & \mathbb{U} \\
\mathbb{R}^s \times \mathbb{R}^s \arrow[r] & G/G_2 \times G/G_2 \arrow[r, dashed] & G_2/[G,G_2] \arrow[ur, dashed] &
\end{tikzcd}$$

Hence

$$\begin{aligned}
\sigma_i(h) = -\psi(\{g(h)\})A\psi(g(e_i)) + \sum_{i<j} h_j \psi(g(e_i))A\psi(g(e_j)) \\
= -\{\sum_j h_j g_j\}Ag_i + \sum_{i<j} h_j g_i A g_j,
\end{aligned}$$

where $g_i = \psi(g(e_i)) \in \mathbb{R}^s$. Inserting this into the previous display we obtain

$$\mathbb{Z} \ni p_i - \{\sum_{j=1}^r h_j g_j\}Ag_i + \sum_{j=1}^r (q_{ij} + \delta_{i<j} g_i A g_j)h_j =: p_i - \{\sum_{j=1}^r h_j g_j\} \cdot \xi_i + \sum_{j=1}^r \tilde{q}_{ij} h_j,$$

where $\xi_i = Ag_i \in \mathbb{R}^s$ is bounded and $\tilde{q}_{ij} \in \mathbb{R}$. Let $i \in \{1,\dots,r\}$ be arbitrary. For a positive lower density set H of $h \in \mathbb{Z}^r$ we have

$$p_i - \{\sum_{j=1}^r h_j g_j\} \cdot \xi + \sum_{j=1}^r \tilde{q}_{ij} h_j \in \mathbb{Z}. \tag{13.14}$$

Hence the sequence $h \mapsto (\{\sum_{j=1}^r h_j g_j\}, \{\sum_{j=1}^r \tilde{q}_{ij} h_j\}) \in [0,1]^{s+1}$ takes values in a union U of parallel planes with distance $\|(\xi,1)\|^{-1}/R$ for $h \in H$, and this distance is bounded from below. In particular, this sequence is not equidistributed, and applying Lemma 13.8 with $F(\cdot) = (1 - C\operatorname{dist}(\cdot, U))_+$ (distance taken on the torus) we find that this sequence is contained in a proper subtorus described by the equation $\{x : k \cdot x = \text{const} \mod \mathbb{Z}\}$ with a bounded non-zero $k = (k_s, k') \in \mathbb{Z}^s \times \mathbb{Z}$. We have therefore

$$\mathbb{Z} \ni k \cdot (g_j, \tilde{q}_{ij}) = k_s \cdot g_j + \tilde{q}_{ij} k'$$

for all $j = 1,\dots,r$.

Suppose first that $k' = 0$, so that $k_s \cdot g_j \in \mathbb{Z}$ for all j. Then $g \mapsto k_s \cdot \psi(g)$ is a non-trivial character on G that vanishes on Γ and $g(e_j)$, $j = 1,\dots,r$, hence satisfies the conclusion of the theorem.

Suppose now that $k' \neq 0$. Multiplying (13.14) by k' we obtain

$$\begin{aligned}
k' p_i &\equiv \{\sum_{j=1}^r h_j g_j\} \cdot k'\xi - \sum_{j=1}^r \tilde{q}_{ij} k' h_j \\
&\equiv \{\sum_{j=1}^r h_j g_j\} \cdot k'\xi + \sum_{j=1}^r h_j g_j \cdot k_s \\
&\equiv \{\sum_{j=1}^r h_j g_j\} \cdot (k'\xi + k_s) \mod \mathbb{Z}.
\end{aligned}$$

If $k'\xi + k_s \neq (0,\dots,0)$, then repeating the above argument we obtain

$$\tilde{k} \cdot g_j \in \mathbb{Z}, \quad j = 1,\dots,r$$

with some bounded non-zero $\tilde{k} \in \mathbb{Z}^s$, and we can conclude in the same way as in the case $k' = 0$.

If we have not arrived at the conclusion after running the above argument for each i, then we have $\xi_i \in \frac{1}{R'}\mathbb{Z}^s$ for each $i = 1, \dots, r$. Multiplying η by R' we may assume $R' = 1$ (we also obtain $p_i \in \mathbb{Z}$, so that $P(n) = \eta_1(g(n)) = 0$, but this information cannot be used because we cannot exclude the possibility of η_1 being trivial).

Consider the characters $\tau_\iota : G \to \mathbb{R}, g \mapsto e_\iota A\psi(g)$ for $\iota = 1, \dots, s$. These characters vanish on Γ, and $\tau_\iota(g(e_j)) = e_\iota A\psi(g(e_j)) = e_\iota \cdot \xi_j \equiv 0 \mod \mathbb{Z}$. If one of these characters is non-trivial, then it satisfies the conclusion of the theorem.

It remains to consider the case when each τ_ι is trivial. Then $A = 0$, so that η_2 vanishes on the commutator subgroup $[G, G]$. It follows that the sequence

$$G'_1 := G_2, \quad G'_i := G_i \cap \ker \eta_2, i \geq 2 \tag{13.15}$$

is a filtration with $\dim G'_2 < \dim G_2$.

Moreover, $\eta_2([G, G]) = 0$ implies $\sigma_i(h) = 0$ in (13.11). The differentiation argument now shows that the coefficients of Q of order > 1 are rational (with bounded denominator), so, after multiplying η by a bounded number, we may assume $Q(n) = \eta_2(g_{\mathrm{nlin}}(n)) = 0$. In other words, g_{nlin} takes values in G'_2. It follows that g is polynomial with respect to the filtration $G'_\bullet$, and we obtain the conclusion by induction on $\dim G_2$. □

14 Jointly intersective polynomials

We call a set of the form $\Lambda = \prod_{i=1}^{r}(r_i + a_i\mathbb{Z}) \subset \mathbb{Z}^r$, $a_i \neq 0$, a *lattice*. Every coset of a finite index subgroup of $\mathbb{Z}^r$ is a lattice modulo a change of coordinates.

14.1 P-sequences in nilpotent groups

Let G be a group and P a ring of functions $\mathbb{Z} \to \mathbb{Z}$. A *$P$-sequence in G* is a sequence of the form $n \mapsto \prod_{i=1}^{l} g_i^{p_i(n)}$, where $g_i \in G$ and $p_i \in P$. We call a ring P of functions $\mathbb{Z} \to \mathbb{Z}$ *gcd-normalized* if for every $p \in P$ and $c \geq 0$ such that $c|p(n)$ for all n also the function $n \mapsto p(n)/c$ is in P. Note that if P is gcd-normalized, $j \geq 1$, and $p \in p$, then also the function $n \mapsto \binom{p(n)}{j}$ is in P.

Lemma 14.1. *Let G be a nilpotent group, P a gcd-normalized ring of functions, and g a P-sequence in G with values in a subgroup $H \leq G$. Then g is a P-sequence in H.*

The conclusion means that $g(n)$ can be written as $\prod_{i=1}^{l} h_i^{p_i(n)}$ with $h_i \in H$ and $p_i \in P$.

Proof. We induct on the nilpotency degree of G. If the nilpotency degree equals 0, then g vanishes identically and the conclusion holds trivially.

Suppose that G has nilpotency degeree d and the lemma is known for groups with nilpotency degree $< d$. By the hypothesis $g(n) = \prod_{i=1}^{l} g_i^{p_i(n)}$. Replacing G by the subgroup generated by the g_i's we may assume that G is finitely generated. We will factorize $g = hg'$ with h being a P-sequence in H and g' a P-sequence in G with values in $G_2 := [G, G]$. It will then follow that g' takes values in $G_2 \cap H$, so it is a P-sequence in H by the inductive hypothesis.

The factorization proceeds in two steps. Let $\tilde{H}$ be the subgroup of G generated by H and G_2. We first factorize $g = \tilde{h}g'$ with $\tilde{h}$ being a P-sequence in $\tilde{H}$. To this end note that G/G_2 is a finitely generated abelian group, and by the structure theorem for submodules of $\mathbb{Z}^d$ we can find a set of generators $r_1, \dots, r_d$ for G/G_2 and integers $c_1, \dots, c_l$ such that

$$n_1 r_1 + \dots + n_d r_d \in \tilde{H}/G_2 \leq G/G_2 \iff c_1|n_1, \dots, c_l|n_l, n_{l+1} = \dots = n_d = 0.$$

We choose representatives for the congruence classes $r_1, \dots, r_d$ in G. Then every element $\gamma \in G$ can be written as $\prod_{i=1}^{d} r_i^{k_i} \cdot \rho$ with $\rho \in G_2$. Note that the sequences $n \mapsto \gamma^n$ and $n \mapsto \prod_{i=1}^{d} r_i^{k_i n}$ are both polynomial and coincide modulo G_2. Hence we have

$$\gamma^n = \prod_{i=1}^{d} r_i^{k_i n} \cdot \rho(n),$$

where the sequence $\rho(n)$ vanishes at 0 and is polynomial with respect to the filtration $G'_\bullet$ given by

$$G'_0 = G'_1 = G'_2 = G_2, \quad G'_i = G_i, i > 2.$$

By Lagrange interpolation we obtain

$$\rho(n) = \prod_{j \geq 1} \rho_j^{\binom{n}{j}}$$

with some $\rho_j \in G_2$. It follows that

$$g(n) = \prod_{i=1}^{l} \Big(\prod_{j=1}^{d} r_j^{k_{i,j} p_i(n)} \prod_{j \geq 1} \rho_{i,j}^{\binom{p_i(n)}{j}} \Big)$$

with $\rho_{i,j} \in G_2$. Note that all functions in the exponents are elements of P. Now we collect the r_j terms; this will produce commutator terms which are P-sequences in G_2 as we will show next. Indeed, let $\gamma, \delta \in G$. Then the sequence

$$(n,m) \mapsto [\gamma^n, \delta^m]$$

is polynomial with respect to the filtration $G'_\bullet$ and vanishes if $n = 0$ or $m = 0$. By Lagrange interpolation it can be written in the form

$$\prod_{j,j' \geq 1} \rho_{j,j'}^{\binom{n}{j}\binom{m}{j'}}.$$

Substituting functions in P for n and m we see that the commutator is indeed a P-sequence in G_2.

This allows us to write

$$g(n) = \tilde{h}(n)g'(n), \quad \tilde{h}(n) = \prod_{j=1}^{d} r_j^{\sum_i k_{i,j}p_i(n)},$$

with g' being a P-sequence in G_2. By the hypothesis that g takes values in H and by the choice of r_j's we have

$$c_j | \sum_i k_{i,j}p_i(n), j = 1, \ldots, l, \quad \sum_i k_{i,j}p_i(n) = 0, j > l$$

for all $n \in \mathbb{Z}$. It follows that

$$\tilde{h} = \prod_{j=1}^{l} \tilde{h}_j^{\tilde{p}_j(n)},$$

where $\tilde{h}_j = r_j^{c_j} \in \tilde{H}$ and $\tilde{p}_j = \sum_i k_{i,j}p_i(n)/c_j \in P$. This completes the first step in the factorization.

Now we factorize $\tilde{h} = hg'$ with h being a P-sequence in H and g' a P-sequence in G_2. This is similar to the first factorization step, but this time we split $\tilde{H} \ni g_i = h_i \rho_i$ with $h_i \in H$ and $\rho_i \in G_2$. □

14.2 Jointly intersective polynomials

A polynomial $p : \mathbb{Z}^r \to \mathbb{Z}$ is called *intersective* if it has a zero modulo every integer.

Example. The polynomial $(n^2 - 13)(n^2 - 17)(n^2 - 13 \cdot 17)$ is intersective. Note first that it suffices to find zeros modulo powers of prime numbers. We have $17 \equiv 1 \mod 8$, from which it follows that 17 is a quadratic residue modulo every power of 2 (Gauss, DA, 103). Moreover, 13 is a quadratic residue modulo 17 and 17 is a quadratic residue modulo 13. From multiplicativity of Legendre symbol we obtain that at least one of 13, 17, $13 \cdot 17$ is a quadratic residue modulo p, where p is a prime $\neq 2, 13, 17$. We conclude using the fact that if p is an odd prime, $(p, q) = 1$, and q is a quadratic residue modulo p, then q is a quadratic residue modulo every power of p (Gauss, DA, 101).

In this example $13, 17$ can be replaced by any primes that are $\equiv 1 \mod 4$ and which are quadratic residues modulo each other (by quadratic reciprocity it suffices to check this only one way).

Remark. A criterion for intersectivity that applies to general polynomials is given in [BB96].

Polynomials $p_1, \ldots, p_k : \mathbb{Z}^r \to \mathbb{Z}$ are called *jointly intersective* if they have a common zero modulo every integer.

Lemma 14.2. *Let $p_1, \dots, p_k : \mathbb{Z}^r \to \mathbb{Z}$ be jointly intersective polynomials and $m \geq 1$ an integer. Then there exists a lattice $\Lambda \subset \mathbb{Z}^r$ such that the restrictions of the p_i's to Λ are jointly intersective and vanish modulo m.*

Proof. The polynomials p_i are periodic modulo m, that is, there exists a finite index subgroup $\Lambda_0 \leq \mathbb{Z}^r$ such that each p_i is constant modulo m on each coseet of Λ_0. Let $r \geq 1$ be an integer, then by the pigeonhole principle there exists a coset Λ_r of Λ_0 such that the p_i's have a common zero modulo r on Λ_r. By the pigeonhole principle there exists a coset Λ such that $\Lambda = \Lambda_{r!}$ for arbitrarily large r. This is the required lattice. □

Lemma 14.3. *Let $p_1, \dots, p_k$ be jointly intersective functions and $\alpha_1, \dots, \alpha_k \in \mathbb{R}$. If $\sum_{i=1}^k \alpha_i p_i(n) \equiv c \mod \mathbb{Z}$ for some constant $c \in \mathbb{R}$ and all n, then $c \in \mathbb{Z}$.*

Proof. Choose rationally independent numbers $\frac{1}{R}, \beta_1, \dots, \beta_l$ such that

$$\alpha_i = n_i/R + \sum_{j=1}^{l} n_{i,j}\beta_j$$

for some integers $n_i, n_{i,j}$. By rational independence we then have

$$\sum_{i=1}^{k} n_{i,j} p_i(n) = \text{const},$$

and since the polynomial on the left-hand side is intersective, it must vanish identically. Therefore

$$\sum_{i=1}^{k} \alpha_i p_i(n) = \frac{1}{R} \sum_{i=1}^{k} n_i p_i(n).$$

Since the sum on the right-hand side is an intersective polynomial, it has a zero modulo R, so its constant value modulo $\mathbb{Z}$ must be 0. □

If $G_\bullet$ is a prefiltration and $\bar{d} = (d_i)_{i\in\mathbb{N}} \subset \mathbb{N}$ is a superadditive sequence (i.e. $d_{i+j} \geq d_i + d_j$ for all $i, j \in \mathbb{N}$; by convention $d_{-1} = -\infty$) then $G^{\bar{d}}_\bullet$, defined by

$$G^{\bar{d}}_i = G_j \quad \text{whenever} \quad d_{j-1} < i \leq d_j, \tag{14.4}$$

is again a prefiltration. In particular, if $p : \mathbb{Z}^r \to \mathbb{Z}$ is a polynomial and $g \in G$, then the map $n \mapsto g^{p(n)}$ is polynomial with respect to $G^{\bar{d}}_\bullet$, where $d_i = i \deg p$.

Proposition 14.5. *Let $p_1, \dots, p_k : \mathbb{Z}^r \to \mathbb{Z}$ be jointly intersective polynomials and P the gcd-normalized ring generated by them. Let also G/Γ be a nilmanifold and g a P-sequence in G. Then* $\mathrm{id}\Gamma \in \overline{g(\mathbb{Z}^r)\Gamma} \subset G/\Gamma$.

Proof. By the above remark a P-sequence is necessarily polynomial with respect to a suitable filtration. Since G/Γ is compact, $G^o\Gamma$ is a finite index subgroup of G. It follows from Lemma 14.2 that, passing to a lattice in $\mathbb{Z}^r$, we may assume that g is a P-sequence in $G^o\Gamma$. The algorithm in Lemma 14.1 gives a factorization into a P-sequence in G^o and a P-sequence in Γ. Hence we may assume that g is a P-sequence in G^o.

If g is equidistributed in G^o/Γ, then we are done. Otherwise, by the equidistribution criterion there exists a non-trivial character η on G^o such that $\eta \circ g$ is constant. By Lemma 14.3 this constant must be 0. Hence g takes values in the subgroup $\ker \eta \leq G^o$ that has strictly smaller dimension. By Lemma 14.1 the sequence g is a P-sequence in $\ker \eta$. We conclude by induction on $\dim G$. □

See [BLL08] for the deduction of the Szemerédi theorem for jointly intersective polynomials from Proposition 14.5.

15 Orbit closure theorem

Definition 15.1. Let $\Lambda \subset \mathbb{Z}^r$ be a lattice. A sequence $(x_n)_{n\in\Lambda}$ in a regular measure space (X,μ) is called *well-distributed on* X if for every $f \in C(X)$ we have

$$\lim_{I\subset\Lambda} \mathrm{Av}_{n\in I} f(x_n) = \int f \mathrm{d}\mu$$

and *totally well-distributed on* X if the above display holds for every sublattice $\Lambda' \subset \Lambda$.

An arbitrary polynomial can be factorized into a "totally equidistributed" and a "rational" part as follows.

Lemma 15.2 (Factorization). *Let G/Γ be a nilmanifold. For every $g \in P_0(\mathbb{Z}^r, G_\bullet)$ there exists a closed connected rational subgroup $H \leq G$ such that g can be written in the form $g = h\gamma$, where $\gamma \in P_0(\mathbb{Z}^r, \sqrt{\Gamma}_\bullet)$, $h \in P_0(\mathbb{Z}^r, H_\bullet)$, and for every finite index subgroup $\tilde\Gamma \leq \Gamma$ the sequence $g\tilde\Gamma$ is totally well-distributed on $H/\tilde\Gamma$.*

Proof. We induct on the dimension of G. If $\dim G = 0$, then $G \leq \sqrt{\Gamma}$, and we can set $h \equiv \mathrm{id}$, $\gamma = g$. Suppose now that the conclusion is known for rational subgroups of dimension $< \dim H$.

Consider the splitting $g = g^o\gamma$ into a connected and a rational part; by construction we have $g^o(0) = \gamma(0) = \mathrm{id}$. Suppose that g^o is not totally equidistributed on $G^o/\tilde\Gamma$ for some finite index subgroup $\tilde\Gamma \leq \Gamma$. Then by the equidistribution criterion we have $\eta \circ g^o \equiv 0$ for some non-trivial character η on the group G^o that vanishes on $\tilde\Gamma \cap G^o$. Multiplying η by an integer we may assume that it vanishes on $\Gamma \cap G^o$. Hence g^o takes values in the proper rational subgroup $\ker \eta \leq G^o$, and we can conclude by the induction hypothesis. □

Corollary 15.3 (Point orbit closure). *Let G/Γ be a nilmanifold and $g \in P(\mathbb{Z}^r, G_\bullet)$. Let $H \leq G$ and $g(0)^{-1}g = h\gamma$ be the factorization from Lemma 15.2 and $\Lambda \leq \mathbb{Z}^r$ be a finite index subgroup modulo which $\gamma\Gamma$ is periodic. The for every coset Λ' of Λ the sequence $(g(n)\Gamma)_{n\in\Lambda'}$ is totally well-distributed on the subnilmanifold $g(0)H\gamma(m)\Gamma$, where $m \in \Lambda'$ is arbitrary.*

In order to obtain the precise statement of [Lei05, Theorem B] one could consider $g(0)Hg(0)^{-1}$ instead of H. Note that this subgroup is in general not Γ-rational.

Corollary 15.4 (Subnilmanifold orbit closure [Lei05, Corollary 1.9]). *Let G/Γ be a nilmanifold and $\tilde X = g_0\tilde G\gamma_0\Gamma$ a connected subnilmanifold, where $\tilde G \leq G$ is a connected rational subgroup, $\gamma_0 \in \sqrt{\Gamma}$, and $g_0 \in G$. Let also $g \in P(\mathbb{Z}^r, G_\bullet)$. Then there exists a finite index subgroup $\Lambda \leq \mathbb{Z}^r$ such that for every coset Λ' of Λ the orbit closure $Y := \overline{\cup_{n\in\Lambda'} g(n)\tilde X}$ is also a connected subnilmanifold and the sequence of subnilmanifolds $(g(n)\tilde X)_{n\in\Lambda'}$ is totally well-distributed in Y in the sense that*

$$\lim_{I\subset\Lambda''\subset\Lambda'} \mathrm{Av}_{n\in I}\, g(n)_*\mu_{\tilde X} = \mu_Y$$

in the weak topology on the space of probability measures for every sublattice $\Lambda'' \subset \Lambda'$.*

For simplicity assume $g_0 = \gamma_0 = \mathrm{id}$ (this is the case for the diagonal submanifold of a power of a nilmanifold).

Proof. Let $(\tilde g(m))_{m\in\mathbb{Z}^s}$ be a polynomial sequence in $\tilde G$ that is well-distributed on $\tilde X$ (by the equidistribution criterion it suffices to ensure that it is dense modulo $\Gamma[\tilde G, \tilde G]$).

Consider the polynomial sequence $(n,m) \mapsto g(n)\tilde g(m)$. By Corollary 15.3 there exista finite index subgroup $\Lambda \leq \mathbb{Z}^{r+s}$ such that the sequence $(g(n)g(m)\Gamma)_{(n,m)\in\Lambda'}$ is totally well-distributed on its orbit closure, which is a connected subnilmanifold, for every coset Λ' of Λ. Passing to a subgroup of Λ we may assume $\Lambda = \tilde{\Lambda}_n \times \Lambda_m$, where $\Lambda_n \leq \mathbb{Z}^n$, $\Lambda_m \leq \mathbb{Z}^m$. Then for every subgroup $\Lambda_n \leq \tilde{\Lambda}_n$ we have

$$\begin{aligned}
\mu_Y &= \lim_{I \subset \Lambda_n \times \Lambda_m + (n_0, m_0)} \mathrm{Av}_{(n,m)\in I} (g(n)\tilde{g}(m))_* \delta_\Gamma \\
&= \lim_{I_n \times I_m \subset \Lambda_n + n_0 \times \Lambda_m + m_0} \mathrm{Av}_{n\in I_n} \mathrm{Av}_{m \in I_m} g(n)_* \tilde{g}(m)_* \delta_\Gamma \\
&= \lim_{I_n \subset \Lambda_n + n_0} \lim_{I_m \subset \Lambda_m + m_0} \mathrm{Av}_{n\in I_n} g(n)_* \mathrm{Av}_{m \in I_m} \tilde{g}(m)_* \delta_\Gamma \\
&= \lim_{I_n \subset \Lambda_n + n_0} \mathrm{Av}_{n\in I_n} g(n)_* \lim_{I_m \subset \Lambda_m + m_0} \mathrm{Av}_{m \in I_m} \tilde{g}(m)_* \delta_\Gamma \\
&= \lim_{I_n \subset \Lambda_n + n_0} \mathrm{Av}_{n\in I_n} g(n)_* \mu_{\tilde{X}}
\end{aligned}$$

as required. $\square$

Let P be a set of jointly intersective polynomials $\mathbb{Z}^r \to \mathbb{Z}$ and g a P-sequence in G. Then for every $g_0 \in G$ also $g_0 g g_0^{-1}$ is a P-sequence. It follows from the orbit closure proposition for P-sequences that $\overline{g(\mathbb{Z}^r) g_0 \Gamma} \ni g_0 \Gamma$.

Let Λ be as in the above corollary. There exists some coset Λ' of Λ on which P is still jointly intersective. Therefore also $\overline{\cup_{n\in\Lambda'} g(n)\tilde{X}} \supseteq \tilde{X}$. We apply this to the diagonal in a power of a nilmanifold, see [BLL08] for details.

References

[Aus10] T. Austin. "On the norm convergence of non-conventional ergodic averages". In: *Ergodic Theory Dynam. Systems* 30.2 (2010), pp. 321–338. arXiv: 0805.0320 [math.DS]. MR: 2599882(2011h:37006) (cit. on p. 26).

[BB96] D. Berend and Y. Bilu. "Polynomials with roots modulo every integer". In: *Proc. Amer. Math. Soc.* 124.6 (1996), pp. 1663–1671. MR: 1307495(96h:11107) (cit. on p. 63).

[BL02] V. Bergelson and A. Leibman. "A nilpotent Roth theorem". In: *Invent. Math.* 147.2 (2002), pp. 429–470. MR: 1881925(2003a:37002) (cit. on p. 49).

[BLL08] V. Bergelson, A. Leibman, and E. Lesigne. "Intersective polynomials and the polynomial Szemerédi theorem". In: *Adv. Math.* 219.1 (2008), pp. 369–388. arXiv: 0710.4862 [math.DS]. MR: 2435427(2009e:37004) (cit. on pp. 64, 66).

[ET12] T. Eisner and T. Tao. "Large values of the Gowers-Host-Kra seminorms". In: *J. Anal. Math.* 117 (2012), pp. 133–186. arXiv: 1012.3509 [math.CO]. MR: 2944094 (cit. on p. 56).

[EW11] M. Einsiedler and T. Ward. *Ergodic theory with a view towards number theory.* Vol. 259. Graduate Texts in Mathematics. Springer-Verlag London, Ltd., London, 2011, pp. xviii+481. MR: 2723325(2012d:37016) (cit. on p. 11).

[Fur77] H. Furstenberg. "Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions". In: *J. Analyse Math.* 31 (1977), pp. 204–256. MR: 0498471(58#16583) (cit. on p. 2).

[Fur81] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory.* M. B. Porter Lectures. Princeton, N.J.: Princeton University Press, 1981, pp. xi+203. MR: 603625(82j:28010) (cit. on p. 11).

[FW96] H. Furstenberg and B. Weiss. "A mean ergodic theorem for $\frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$". In: *Convergence in ergodic theory and probability (Columbus, OH, 1993).* Vol. 5. Ohio State Univ. Math. Res. Inst. Publ. Berlin: de Gruyter, 1996, pp. 193–227. MR: 1412607(98e:28019) (cit. on pp. 37, 40, 46).

[Gla03] E. Glasner. *Ergodic theory via joinings.* Vol. 101. Mathematical Surveys and Monographs. Providence, RI: American Mathematical Society, 2003, pp. xii+384. MR: 1958753(2004c:37011).

[GT12] B. Green and T. Tao. "The quantitative behaviour of polynomial orbits on nilmanifolds". In: *Ann. of Math. (2)* 175.2 (2012), pp. 465–540. arXiv: 0709.3562 [math.NT]. MR: 2877065 (cit. on pp. 55–57).

[GT14] B. Green and T. Tao. "On the quantitative distribution of polynomial nilsequences—erratum". In: *Ann. of Math. (2)* 179.3 (2014), pp. 1175–1183. arXiv: 1311.6170 [math.NT]. MR: 3171762 (cit. on p. 57).

[GTZ12] B. Green, T. Tao, and T. Ziegler. "An inverse theorem for the Gowers $U^{s+1}[N]$-norm". In: *Ann. of Math. (2)* 176.2 (2012), pp. 1231–1372. arXiv: 1009.3998 [math.CO]. MR: 2950773 (cit. on pp. 51, 55).

[Hal33] P. Hall. "A contribution to the theory of groups of prime-power order." In: *Proc. Lond. Math. Soc., II. Ser.* 36 (1933), pp. 29–95 (cit. on p. 48).

[HK05] B. Host and B. Kra. "Nonconventional ergodic averages and nilmanifolds". In: *Ann. of Math. (2)* 161.1 (2005), pp. 397–488. MR: 2150389(2007b:37004) (cit. on pp. 21, 40, 41, 45).

[Kec95] A. S. Kechris. *Classical descriptive set theory.* Vol. 156. Graduate Texts in Mathematics. New York: Springer-Verlag, 1995, pp. xviii+402. MR: 1321597(96e:03057) (cit. on p. 37).

[Lei02] A. Leibman. "Polynomial mappings of groups". In: *Israel J. Math.* 129 (2002). with erratum, pp. 29–60. MR: 1910931(2003g:20060) (cit. on p. 51).

[Lei05] A. Leibman. "Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translations on a nilmanifold". In: *Ergodic Theory Dynam. Systems* 25.1 (2005), pp. 215–225. MR: 2122920(2006j:37005) (cit. on pp. 57, 65).

[MKS66] W. Magnus, A. Karrass, and D. Solitar. *Combinatorial group theory: Presentations of groups in terms of generators and relations.* Interscience Publishers [John Wiley & Sons, Inc.], New York-London-Sydney, 1966, pp. xii+444. MR: 0207802(34#7617) (cit. on p. 48).

[Rot53] K. F. Roth. "On certain sets of integers". In: *J. London Math. Soc.* 28 (1953), pp. 104–109. MR: 0051853(14,536g) (cit. on p. 19).

[Sár78] A. Sárközy. "On difference sets of sequences of integers. I". In: *Acta Math. Acad. Sci. Hungar.* 31.1–2 (1978), pp. 125–149. MR: 0466059(57#5942) (cit. on p. 2).

[Sze75] E. Szemerédi. "On sets of integers containing no k elements in arithmetic progression". In: *Acta Arith.* 27 (1975). Collection of articles in memory of Jurij Vladimirovič Linnik, pp. 199–245. MR: 0369312(51#5547) (cit. on p. 2).

[Tak02] M. Takesaki. *Theory of operator algebras. I.* Vol. 124. Encyclopaedia of Mathematical Sciences. Reprint of the first (1979) edition, Operator Algebras and Non-commutative Geometry, 5. Springer-Verlag, Berlin, 2002, pp. xx+415. MR: 1873025(2002m:46083) (cit. on p. 5).

[Tao09] T. Tao. *Poincaré's legacies, pages from year two of a mathematical blog. Part I.* Providence, RI: American Mathematical Society, 2009, pp. x+293. MR: 2523047(2010h:00003) (cit. on p. 11).

[Zie07] T. Ziegler. "Universal characteristic factors and Furstenberg averages". In: *J. Amer. Math. Soc.* 20.1 (2007), 53–97 (electronic). arXiv: math/0403212 [math.DS]. MR: 2257397(2007j:37004) (cit. on pp. 40, 46).