ISEM 28: Ergodic Structure Theory and Applications

Asgar Jamneshan, Henrik Kreidler

April 5, 2025

Contents

Lecture 1

Ergodic theory examines the behavior of measure-preserving maps. In this first lecture, we introduce these maps, and already prove an early basic result of the field: Poincaré's recurrence theorem. In the remainder of the chapter, we then explore several alternative approaches to measure-preserving transformations via the measure algebra and the Koopman operator.

1.1 Measure-Preserving Transformations and Poincaré's Recurrence Theorem

To start, let us briefly recall some basic measure theoretic concepts and introduce notation. A **measurable space** (X, Σ_X) consists of a set X together with a σ-algebra Σ_X of subsets of X (which are called **measurable**). A **probability space** (X, Σ_X, μ_X) is given by a measurable space (X, Σ_X) and a probability measure $\mu_X \colon \Sigma_X \to [0,1]$. In the following, we write just X for a measurable space (X, Σ_X) or a probability space (X, Σ_X, μ_X) if there is no risk of confusion. We assume that the reader is familiar with these concepts and knows basic measure and integration theory. Introductory texts on the subject are, e.g., [Rud87], [Bar95], [Ran02], [Tao11], and [BBP16].

The following are the "structure preserving maps" between measurable and probability spaces, respectively.

Definition 1.1.1. (i) A map $\tau \colon X \to Y$ between measurable spaces X and Y is **measurable** if $\tau^{-1}(A) \subseteq X$ is measurable for every measurable subset $A \subseteq Y$.

(ii) A measurable map $\tau \colon X \to Y$ between probability spaces X and Y is **measure-preserving** if $\mu_X(\tau^{-1}(A)) = \mu_Y(A)$ for every measurable subset $A \subseteq Y$.

It is clear from the definition that the composition of measurable or measure-preserving maps is again measurable or measure-preserving, respectively.

Remark 1.1.2. For any measurable map $\tau\colon X \to Y$ between measurable spaces X and Y and any probability measure μ_X on X, one can define the **pushforward measure** $\tau_*\mu_X$ on Y via $\tau_*\mu_X(A) := \mu_X(\tau^{-1}(A))$ for every measurable subset $A \subseteq Y$. Thus, a measurable map $\tau\colon X \to Y$ between probability spaces is measure-preserving precisely when $\tau_*\mu_X = \mu_Y$.

Here are a few basic examples of measure-preserving maps. Additional and more interesting ones can be found in the Exercises below and in the next lectures.

Examples 1.1.3. (i) For any probability space X the identity map $\mathrm{id}_X\colon X \to X$, $x \mapsto x$ is measure-preserving.

(ii) Equip a finite set $X = \{0, \dots, k-1\}$ for some $k \in \mathbb{N}$ with the power set $\mathcal{P}(X)$ as its σ-algebra. Recall that if $p = (p_0, \dots, p_{k-1}) \in [0,1]^k$ is a **probability vector**, i.e., $\sum_{i=0}^{k-1} p_i = 1$, then we obtain a probability measure $\sum_{i=0}^{k-1} p_i\delta_i\colon \mathcal{P}(X) \to [0,1]$, $A \mapsto \sum_{i\in A} p_i$. One can easily check that every bijection $\tau\colon \{0, \dots, k-1\} \to \{0, \dots, k-1\}$ is then measure-preserving with respect to the measure defined by the vector $(p_0, \dots, p_{k-1}) = (\frac{1}{k}, \dots, \frac{1}{k})$.

(iii) Consider $X = [0, 1)$ with the Borel σ-algebra and the Lebesgue measure, and for fixed $\alpha \in [0, 1)$ define the map $\tau_\alpha\colon [0, 1) \to [0, 1)$ by $\tau_\alpha(x) := x + \alpha \mod 1$ for $x \in [0, 1)$. If $A \subseteq [0, 1)$ is any Borel measurable set, then the preimage

$$\begin{aligned}\tau_\alpha^{-1}(A) &= (\tau_\alpha^{-1}(A) \cap [0, 1-\alpha)) \cup (\tau_\alpha^{-1}(A) \cap [1-\alpha, 1)) \\ &= ((A-\alpha) \cap [0, 1-\alpha)) \cup ((A-\alpha+1) \cap [1-\alpha, 1))\end{aligned}$$

is measurable since translates of Borel measurable subsets are Borel measurable. Moreover, we obtain by translation invariance of the Lebesgue measure, that

$$\begin{aligned}\mu_X(\tau_\alpha^{-1}(A)) &= \mu_X((A-\alpha) \cap [0, 1-\alpha)) + \mu_X((A-\alpha+1) \cap [1-\alpha, 1)) \\ &= \mu_X(A \cap [\alpha, 1)) + \mu_X(A \cap [0, \alpha)) = \mu_X(A).\end{aligned}$$

Thus, τ_α is a measure-preserving map.

In many instances it is difficult or tedious to directly check with the definition that a map is indeed measure-preserving. We therefore collect some equivalent characterizations. Recall here that for a measurable space X a subset $\mathcal{E} \subseteq \Sigma_X$ is a **generator of** Σ_X if Σ_X is the smallest σ-algebra over X containing $\mathcal{E}$, and $\cap$**-stable** if $A \cap B \in \mathcal{E}$ whenever $A, B \in \mathcal{E}$.

Proposition 1.1.4. (i) *Let X and Y be measurable spaces and $\mathcal{E} \subseteq \Sigma_Y$ a generator. For a map $\tau\colon X \to Y$ the following assertions are equivalent.*

(a) *τ is measurable.*

(b) *$\tau^{-1}(A) \subseteq X$ is measurable for every $A \in \mathcal{E}$.*

(c) *$f \circ \tau$ is measurable for every measurable function $f \colon Y \to \mathbb{C}$.*

(ii) *For a measurable map $\tau \colon X \to Y$ between probability spaces X and Y and a $\cap$-stable generator $\mathcal{E} \subseteq \Sigma_Y$ the following assertions are equivalent.*

(a) *τ is measure-preserving.*

(b) *$\mu_X(\tau^{-1}(A)) = \mu_Y(A)$ for every $A \in \mathcal{E}$.*

(c) *$\int_X f \circ \tau = \int_Y f$ for every measurable function $f \colon Y \to [0, \infty)$.*

(d) *$f \circ \tau \colon X \to \mathbb{C}$ is integrable with $\int_X f \circ \tau = \int_Y f$ for every integrable function $f \colon Y \to \mathbb{C}$.*

We leave the proof (using standard measure theoretic arguments) to the interested reader. For the equivalence "(a) $\Leftrightarrow$ (b)" of part (ii), the Carathéodory uniqueness theorem from measure theory is needed (see, e.g., [Bil95, Theorem 10.3] for a reference).

As for many measure theoretic concepts, it is irrelevant in most situations what a measure-preserving map does on nullsets. It is therefore convenient to identify measure-preserving maps which agree almost everywhere. Given probability spaces X and Y, denote the set of all measure-preserving maps $\tau \colon X \to Y$ by $\mathscr{M}(X, Y)$. Write $\tau_1 \sim \tau_2$ for $\tau_1, \tau_2 \in \mathscr{M}(X, Y)$ if the equality $\tau_1(x) = \tau_2(x)$ holds for almost every $x \in X$. One can check that $\sim$ is an equivalence relation on $\mathscr{M}(X, Y)$, and we write $\mathrm{M}(X, Y) \coloneqq \mathscr{M}(X, Y)/\sim$ for the set of equivalence classes.

Similarly as for L^p-spaces, we still write τ for an element of $\mathrm{M}(X, Y)$ and pick a representative in $\mathscr{M}(X, Y)$ (i.e., an actual measure-preserving map $X \to Y$) whenever necessary. In particular, using this convention, for probability spaces X, Y and Z the composition of measure-preserving maps gives us a (well-defined!) map

$$\circ \colon \mathrm{M}(Y, Z) \times \mathrm{M}(X, Y) \to \mathrm{M}(X, Z), \quad (\sigma, \tau) \mapsto \sigma \circ \tau.$$

The following concept of invertible measure-preserving transformations is therefore natural.

Definition 1.1.5. Let X and Y be probability spaces. Then $\tau \in \mathrm{M}(X, Y)$ is **invertible** if there is a (necessarily unique) $\tau' \in \mathrm{M}(Y, X)$ such that

$$\tau' \circ \tau = \mathrm{id}_X \text{ in } \mathrm{M}(X, X) \text{ and } \tau \circ \tau' = \mathrm{id}_Y \text{ in } \mathrm{M}(Y, Y).$$

In this case, τ is an **isomorphism of probability spaces**. If such an isomorphism exists, then the probability spaces X and Y are **isomorphic**.

All the measure-preserving maps in Examples 1.1.3 define isomorphisms of probability spaces. An example which is not an isomorphism is discussed in Exercise 1.2 below.

We now discuss one of the early insights of ergodic theory. Given a measure-preserving map $\tau\colon X \to X$ we want to study recurrence properties: Given any measurable subset $A \subseteq X$ with $\mu_X(A) > 0$, do we find an element $x \in A$ and some "time" $n \in \mathbb{N}$ such that $\tau^n(x)$ is again in A?[1] Put differently, are there elements in the intersection $A \cap \tau^{-n}(A)$ for some $n \in \mathbb{N}$? The following result shows that we can indeed find "many" points in A returning to A.

Theorem 1.1.6 (Poincaré's recurrence theorem). *Let $\tau\colon X \to X$ be a measure-preserving map on a probability space X. If $A \subseteq X$ is measurable with $\mu_X(A) > 0$, then there is an $n \in \mathbb{N}$ with $\mu_X(A \cap \tau^{-n}(A)) > 0$.*

Proof. Assume that $\mu_X(A \cap \tau^{-n}(A)) = 0$ for every $n \in \mathbb{N}$. Since τ is measure-preserving, we then have $\mu_X(\tau^{-m}(A) \cap \tau^{-(n+m)}(A)) = 0$ for every $m \in \mathbb{N}_0$ and $n \in \mathbb{N}$. This yields $\mu_X(\tau^{-m}(A) \cap \tau^{-k}(A)) = 0$ for all $m, k \in \mathbb{N}$ with $m \neq k$. Thus, the sets $\tau^{-n}(A)$ for $n \in \mathbb{N}$ are pairwise disjoint "up to nullsets", and consequently we have

$$\mu_X\left(\bigcup_{n\in\mathbb{N}} \tau^{-n}(A)\right) = \sum_{n=1}^{\infty} \mu_X(\tau^{-n}(A)) = \sum_{n=1}^{\infty} \mu_X(A) = \infty,$$

a contradiction. □

Studying different notions of recurrence for measure-preserving transformations will be one focus of this course. One of the deeper results of ergodic theory is Furstenberg's multiple recurrence theorem telling that we can return to the set A after finitely many multiples $n, 2n, 3n, \ldots, kn$ of some "time" $n \in \mathbb{N}$ for each $k \in \mathbb{N}$:

Theorem 1.1.7 (Furstenberg's multiple recurrence theorem). *Let $\tau\colon X \to X$ be a measure-preserving map on a probability space X. If $A \subseteq X$ is measurable with $\mu_X(A) > 0$ and $k \in \mathbb{N}$, then there is $n \in \mathbb{N}$ with $\mu_X(A \cap \tau^{-n}(A) \cap \cdots \cap \tau^{-kn}(A)) > 0$.*

This result is of particular interest due to its consequences in additive combinatorics. It implies a celebrated theorem on the existence of arithmetic progressions in "asymptotically large" sets of natural numbers. Here we write $|A|$ for the number of elements of a finite set A.

Theorem 1.1.8 (Szemerédi). *Let $A \subseteq \mathbb{N}$ with $\limsup_{N\to\infty} \frac{|A\cap\{1,\ldots,N\}|}{N} > 0$. For every $k \in \mathbb{N}$ there is a starting number $a \in \mathbb{N}$ and a distance $d \in \mathbb{N}$ such that $a, a+d, \ldots, a+kd \in A$.*

This fruitful connection between combinatorial number theory and ergodic theory, known as *Furstenberg's correspondence principle*, will be discussed in Lecture 4. However, showing Theorem 1.1.7 (and related results) requires a substantial amount

[1] Here and in the following, $\mathbb{N}$ is the set of all integers $n \geq 1$, while $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

of work. We return to it in a later lecture once we have developed the necessary structure theory for measure-preserving systems.

1.2 Measure Algebras and Their Homomorphisms

While measure-preserving transformations are defined as concrete maps $\tau\colon X \to X$ on probability spaces, all we actually need to formulate and prove Poincaré recurrence is the induced transformation

$$\Sigma_X \to \Sigma_X, \quad A \mapsto \tau^{-1}(A)$$

of measurable sets. As we will see, ergodic structure theory also does not rely on the underlying map directly, but rather on the induced transformations of measurable subsets and functions. We can therefore "forget" about the concrete map on X altogether, and instead consider transformations on the level of measurable sets. As we can also ignore what happens on nullsets, this leads to studying transformations of so-called "measure algebras".

To make this precise, recall that for sets A and B the set $A \,\Delta\, B \coloneqq (A \setminus B) \cup (B \setminus A)$ is their **symmetric difference**. For a probability space X we then say that measurable sets $A, B \subseteq X$ **agree up to null-sets** if $\mu_X(A \,\Delta\, B) = 0$, and write $A \sim B$ in this case. One can easily check that $\sim$ defines an equivalence relation on the σ-algebra Σ_X.

Definition 1.2.1. For a probability space X the set

$$\Sigma(X) \coloneqq \Sigma_X/{\sim} = \{[A] \mid A \in \Sigma_X\}$$

is called the **measure algebra** of X.

Remark 1.2.2. One of the advantages of $\Sigma(X)$ is that we can make sense of "uncountable unions", see Exercises 1.4 and 1.6 for this and further properties.

Once again it is convenient to still denote the equivalence classes in $\Sigma(X)$ by the letters $A, B, C, \ldots$ and pick an identically denoted representative in Σ_X whenever necessary. With this convention, we can form $A \cap B, A \cup B, A \setminus B \in \Sigma(X)$ for $A, B \in \Sigma(X)$, and define a map

$$\Sigma(X) \to [0, 1], \quad A \mapsto \mu_X(A).$$

Of course, one has to check that these constructions do not depend on nullsets, and hence are independent of the chosen representative. By a slight abuse of notation, we also write $\emptyset$ and X for the elements of $\Sigma(X)$ defined by the empty set and the entire space, respectively. Now if $\tau\colon X \to Y$ is a measure-preserving map between probability spaces, then, since pre-images of nullsets are again nullsets, τ gives rise to a map between the respective measure algebras:

Definition 1.2.3. Let $\tau\colon X \to Y$ be a measure-preserving map between probability spaces. Then we call

$$\tau^*\colon \Sigma(Y) \to \Sigma(X), \quad A \mapsto \tau^{-1}(A)$$

the **pullback (modulo null sets)** of τ.

Remark 1.2.4. Note that pullbacks are compatible with compositions: If $\tau\colon X \to Y$ and $\sigma\colon Y \to Z$ are measure-preserving maps between probability spaces, then $(\sigma \circ \tau)^* = \tau^* \circ \sigma^*$.

The pullback of a measure-preserving map evidently preserves the algebraic operations of union and intersection, and the measure.

Proposition and Definition 1.2.5. *For a measure-preserving map $\tau\colon X \to Y$ between probability spaces the map $T = \tau^*\colon \Sigma(Y) \to \Sigma(X)$ is a* ***measure algebra homomorphism****, i.e., it satisfies*

(i) $T(A \cup B) = T(A) \cup T(B)$ *for all* $A, B \in \Sigma(Y)$,

(ii) $T(A \cap B) = T(A) \cap T(B)$ *for all* $A, B \in \Sigma(Y)$,

(iii) $\mu_X(T(A)) = \mu_Y(A)$ *for all* $A \in \Sigma(Y)$.[2]

In the following, given probability spaces X and Y, write $\mathrm{M}(\Sigma(Y), \Sigma(X))$ for the set of measure algebra homomorphisms $T\colon \Sigma(Y) \to \Sigma(X)$. We list some important properties. It is Exercise 1.5 to prove these.

Lemma 1.2.6. *Let X and Y be probability spaces and $T\colon \Sigma(Y) \to \Sigma(X)$ a measure algebra homomorphism. Then the following statements hold.*

(i) $T(\emptyset) = \emptyset$ *and* $T(Y) = X$.

(ii) $T(A \setminus B) = T(A) \setminus T(B)$ *for all* $A, B \in \Sigma(Y)$.

(iii) $T(A \,\Delta\, B) = T(A) \,\Delta\, T(B)$ *for all* $A, B \in \Sigma(Y)$.

(iv) $T\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \bigcup_{n\in\mathbb{N}} T(A_n)$ *for every sequence* $(A_n)_{n\in\mathbb{N}}$ *in* $\Sigma(Y)$.

(v) T *is injective.*

(vi) *If T is surjective, then $T^{-1}\colon \Sigma(X) \to \Sigma(Y)$ is also a measure algebra homomorphism.*

[2] As observed by the local group in Ljubljana, part (ii) can actually be inferred from (i) and (iii) as follows. Note first that (iii) implies $T(Y) = X$. By (i) we therefore obtain $X = T(A) \cup T(Y \setminus A)$, which yields $\mu_X((X \setminus T(A)) \setminus T(Y \setminus A)) = 0$. Since

$$\mu_X(X \setminus T(A)) = 1 - \mu_X(T(A)) = 1 - \mu_Y(A) = \mu_Y(Y \setminus A) = \mu_X(T(Y \setminus A)),$$

this already implies $X \setminus T(A) = T(Y \setminus A)$. Combined with (i) and De Morgan's laws this yields (ii).

Are there any examples of measure algebra homomorphisms $T\colon \Sigma(Y) \to \Sigma(X)$ other than pullbacks of measure-preserving transformations $\tau\colon X \to Y$? In general, this can be the case (see [JT23b, Section 5]). Moreover, two distinct elements $\sigma, \tau \in \mathrm{M}(X, Y)$ can induce the same pullback between the corresponding measure algebras, see Exercise 1.7 below. However, for certain "nice" probability spaces such behavior does not occur. Recall that a metric space (X, d_X) is **separable** if there is a countable subset of X which is dense in X, and **complete** if every Cauchy sequence in X converges. Recall also that the **Borel σ-algebra** $\mathcal{B}(X)$ of any topological space is the smallest σ-algebra over X containing all open sets. By equipping a separable and complete metric space (X, d_X) with $\mathcal{B}(X)$ we obtain a measurable space called a **standard Borel space**. Most examples of probability spaces we will encounter in this course are given by a probability measure on such a standard Borel space. The following concept allows even some more flexibility.

Definition 1.2.7. A probability space X is a **Lebesgue space** if there is a separable and complete metric space (Y, d_Y) and a probability measure $\mu_Y\colon \mathcal{B}(Y) \to [0,1]$ such that X is isomorphic to $(Y, \mathcal{B}(Y), \mu_Y)$ (in the sense of Definition 1.1.5).

For Lebesgue spaces the following result establishes a one-to-one correspondence between (equivalence classes of) measure-preserving maps and measure algebra homomorphisms between the corresponding measure algebras.

Theorem 1.2.8. *For every probability space X and every Lebesgue space Y the map $\mathrm{M}(X, Y) \to \mathrm{M}(\Sigma(Y), \Sigma(X))$, $\tau \mapsto \tau^*$ is a bijection.*

We do not show the result in the main part of the course (as it is not crucial for what follows), but a proof is included as a supplement at the end of this lecture.

1.3 Koopman's Approach

We now take yet another perspective on measure-preserving transformations using functional analysis.[3] Observe that for any measure-preserving map $\tau\colon X \to Y$ between probability spaces and any square-integrable function $f\colon Y \to \mathbb{C}$ we obtain from Proposition 1.1.4 that the composition $f \circ \tau$ is again square-integrable with the same L^2-norm. Since for functions $f_1, f_2\colon Y \to \mathbb{C}$ agreeing almost everywhere, the compositions $f_1 \circ \tau$ and $f_2 \circ \tau$ also agree almost everywhere, we obtain a well-defined map on the level of L^2-spaces:

Definition 1.3.1. Let $\tau\colon X \to Y$ be a measure-preserving map between probability

[3] While we assume the reader to be familiar with basic measure and integration theory, we briefly recall all necessary functional analytic notions and results for the course in the appendix.

spaces X and Y. The map

$$U_\tau\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X), \quad f \mapsto f \circ \tau$$

is called the **Koopman operator** induced by τ.

Koopman operators are bounded linear operators, but have important additional properties.

Proposition and Definition 1.3.2. *Let $\tau\colon X \to Y$ be a measure-preserving map between probability spaces. Then $U = U_\tau\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ is a* ***Markov embedding****, i.e.,*

(i) *U is a linear isometry,*

(ii) *$|Uf| = U|f|$ for all $f \in \mathrm{L}^2(Y)$, and*

(iii) *$U\mathbb{1} = \mathbb{1}$.*

Here, $\mathbb{1}$ denotes the (equivalence class) of the constant one-function, and, as usual, the absolute value of functions is defined pointwise. The verification of these properties is straightforward from the definitions, and is left to the reader.

Markov embeddings (even when not given as Koopman operators) have a number of useful properties, which we list below. Again all operations on functions (e.g., real and imaginary parts, suprema and infima, etc.) here are defined pointwise.

Lemma 1.3.3. *For probability spaces X and Y and a Markov embedding $U\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ the following assertions hold.*

(i) *$U(\mathrm{L}^2(Y, [0,\infty))) \subseteq \mathrm{L}^2(X, [0,\infty))$.*

(ii) *$U(\mathrm{L}^2(Y, \mathbb{R})) \subseteq \mathrm{L}^2(X, \mathbb{R})$.*

(iii) *$U \sup(f,g) = \sup(Uf, Ug)$ and $U \inf(f,g) = \inf(Uf, Ug)$ for $f, g \in \mathrm{L}^2(Y, \mathbb{R})$.*

(iv) *$Uf_+ = (Uf)_+$ and $Uf_- = (Uf)_-$ for all $f \in \mathrm{L}^2(Y, \mathbb{R})$.*

(v) *$U(\mathrm{Re}\, f) = \mathrm{Re}\, Uf$ and $U(\mathrm{Im}\, f) = \mathrm{Im}\, Uf$ for all $f \in \mathrm{L}^2(Y)$.*

(vi) *$U\overline{f} = \overline{Uf}$ for all $f \in \mathrm{L}^2(Y)$.*

(vii) *$\int_X Uf = \int_Y f$ for every $f \in \mathrm{L}^2(Y)$.*

(viii) *$U(\mathrm{L}^\infty(Y)) \subseteq \mathrm{L}^\infty(X)$ and $\|Uf\|_\infty = \|f\|_\infty$ for all $f \in \mathrm{L}^\infty(Y)$.*

(ix) *U is injective.*

(x) *If U is surjective, then $U^{-1}\colon \mathrm{L}^2(X) \to \mathrm{L}^2(Y)$ is also a Markov embedding.*

Proof. For part (i) note that if $f \in \mathrm{L}^2(Y, [0,\infty))$, then $f = |f|$, hence $Uf = U|f| = |Uf| \in \mathrm{L}^2(X, [0,\infty))$. Part (ii) is a direct consequence of (i) since every $f \in \mathrm{L}^2(Y, \mathbb{R})$ can be written as $f = f_+ - f_-$, hence as a difference of elements in $\mathrm{L}^2(Y, [0,\infty))$.

For (iii) notice that $\sup(f,g) = \frac{f+g+|f-g|}{2}$ for $f, g \in \mathrm{L}^2(Y,\mathbb{R})$. Thus,

$$U\sup(f,g) = U\left(\frac{f+g+|f-g|}{2}\right) = \frac{Uf+Ug+|Uf-Ug|}{2} = \sup(Uf,Ug).$$

But then also $U\inf(f,g) = U(-\sup(-f,-g)) = -\sup(-Uf,-Ug) = \inf(Uf,Ug)$. Part (iv) is a special case of (iii). Part (v) follows from linearity of U and part (ii), and (vi) then follows from (v).

We now show (vii). Since U is an isometry on a Hilbert space, it respects inner products. Thus

$$\int_X Uf = (Uf|\mathbb{1}) = (Uf|U\mathbb{1}) = (f|\mathbb{1}) = \int_Y f \quad \text{for } f \in \mathrm{L}^2(Y).$$

To check (viii), let $f \in \mathrm{L}^\infty(Y)$ and take $c \geq 0$. Then $|f| \leq c\mathbb{1}$ almost everywhere precisely when $(|f| - c\mathbb{1})_+ = 0$ in $\mathrm{L}^2(X)$, i.e., $\|(|f| - c\mathbb{1})_+\|_2 = 0$. But

$$\|(|f| - c\mathbb{1})_+\|_2 = \|U(|f| - c\mathbb{1})_+\|_2 = \|(|Uf| - c\mathbb{1})_+\|_2$$

by the definition of Markov embeddings and part (iv). This shows that the inequality $|f| \leq c\mathbb{1}$ holds almost everywhere precisely when $|Uf| \leq c\mathbb{1}$ almost everywhere. This shows (viii). Finally, part (ix) follows from the fact that U is isometric, while (x) is straightforward to see. □

The next result shows that, just as every measure-preserving map, each measure algebra homomorphism also induces a "Koopman operator".

Proposition 1.3.4. *Let X and Y be probability spaces. If $T\colon \Sigma(Y) \to \Sigma(X)$ is a measure algebra homomorphism, then there is a unique Markov embedding $U_T\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ with $U_T\mathbb{1}_A = \mathbb{1}_{T(A)}$ for all $A \in \Sigma(Y)$.*

Remark 1.3.5. Note that the uniqueness property immediately implies a compatibility with compositions: For probability spaces X, Y, and Z, and measure algebra homomorphisms $S\colon \Sigma(Z) \to \Sigma(Y)$ and $T\colon \Sigma(Y) \to \Sigma(X)$ we have $U_{T\circ S} = U_T \circ U_S$.

We need the following basic observation from measure theory (see, e.g., [BBP16, Proposition 9.24] for a proof).

Lemma 1.3.6. *Let X be a probability space. Then the linear hull $\mathrm{lin}\{\mathbb{1}_A \mid A \in \Sigma(X)\}$ is dense in $\mathrm{L}^p(X)$ for every $p \in [1,\infty]$.*

Proof of Proposition 1.3.4. First note that uniqueness is clear: Since the (equivalence classes of) characteristic functions span a dense subspace of $\mathrm{L}^2(Y)$, any bounded linear operator $U\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ is uniquely determined by its values $U\mathbb{1}_A$ for $A \in \Sigma(Y)$. For existence write E_Y and E_X for the spaces of (equivalence classes) of simple functions in $\mathrm{L}^2(Y)$ and $\mathrm{L}^2(X)$, respectively. If $A_1, \dots, A_k \in \Sigma(Y)$ and $a_1, \dots, a_k \in \mathbb{C}$, we obtain

$$\begin{aligned}\Big\|\sum_{i=1}^k a_i \mathbb{1}_{T(A_i)}\Big\|_2^2 &= \sum_{i=1}^k \sum_{j=1}^k a_i \overline{a_j} \int_X \mathbb{1}_{T(A_i)} \cdot \mathbb{1}_{T(A_j)} = \sum_{i=1}^k \sum_{j=1}^k a_i \overline{a_j} \mu_X(T(A_i) \cap T(A_j)) \\ &= \sum_{i=1}^k \sum_{j=1}^k a_i \overline{a_j} \mu_Y(A_i \cap A_j) = \Big\|\sum_{i=1}^k a_i \mathbb{1}_{A_i}\Big\|_2^2.\end{aligned}$$

A moment's thought reveals that we now obtain a (well-defined!) linear map $U_T \colon E_Y \to E_X$, $\sum_{i=1}^k a_i \mathbb{1}_{A_i} \to \sum_{i=1}^k a_i \mathbb{1}_{T(A_i)}$ which is isometric with respect to the L^2-norms. Since E_Y is dense in $\mathrm{L}^2(Y)$ by Lemma 1.3.6, we obtain from basic functional analysis (see Proposition A.1.1) that U_T uniquely extends to a linear isometry $U_T \colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$. To check that U_T is a Markov embedding, first take a simple function $f = \sum_{i=1}^k a_i \mathbb{1}_{A_i} \in E_Y$. We can assume that $\mu_Y(A_i \cap A_j) = 0$ and hence also $\mu_X(T(A_i) \cap T(A_j)) = 0$ for $i, j \in \{1, \dots, k\}$ with $i \neq j$. Then

$$U_T|f| = U_T \sum_{i=1}^k |a_i| \mathbb{1}_{A_i} = \sum_{i=1}^k |a_i| \mathbb{1}_{T(A_i)} = \Big|\sum_{i=1}^k a_i \mathbb{1}_{T(A_i)}\Big| = |U_T f|.$$

For general $f \in \mathrm{L}^2(Y)$ we find a sequence $(f_n)_{n\in\mathbb{N}}$ in E_Y with $\lim_{n\to\infty} f_n = f$ in $\mathrm{L}^2(Y)$. Then also $\lim_{n\to\infty} U_T f_n = U_T f$ since U_T is continuous. Since $||f_n| - |f|| \leq |f_n - f|$ for every $n \in \mathbb{N}$, also $\lim_{n\to\infty} |f_n| = |f|$ in $\mathrm{L}^2(Y)$, and, by the same argument, $\lim_{n\to\infty} |U_T f_n| = |U_T f|$. We conclude that $U_T|f| = \lim_{n\to\infty} U_T|f_n| = \lim_{n\to\infty} |U_T f_n| = |U_T f|$. Finally, $U_T \mathbb{1} = U_T \mathbb{1}_Y = \mathbb{1}_{T(Y)} = \mathbb{1}_X = \mathbb{1}$ by Lemma 1.2.6 (i). Thus, U_T is a Markov embedding. □

It turns out that measure algebra homomorphisms and Markov embeddings are equivalent concepts. Given probability spaces X and Y, write $\mathrm{M}(\mathrm{L}^2(Y), \mathrm{L}^2(X))$ for the set of Markov embeddings $U \colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$. Then the following correspondence holds.

Theorem 1.3.7. *For all probability spaces X and Y the map $\mathrm{M}(\Sigma(Y), \Sigma(X)) \to \mathrm{M}(\mathrm{L}^2(Y), \mathrm{L}^2(X))$, $T \mapsto U_T$ is a bijection.*

Proof. To see that the map is injective, observe that if $T_1, T_2 \colon \Sigma(Y) \to \Sigma(X)$ are two measure algebra homomorphisms with $U_{T_1} = U_{T_2}$, then

$$\mathbb{1}_{T_1(A)} = U_{T_1} \mathbb{1}_A = U_{T_2} \mathbb{1}_A = \mathbb{1}_{T_2(A)}$$

and hence $T_1(A) = T_2(A)$ for every $A \in \Sigma(Y)$.

We now show that the map is surjective. Let $U \colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ be a Markov embedding. If $A \in \Sigma(Y)$, then, by Lemma 1.3.3 (iii), the element $f := U\mathbb{1}_A \in \mathrm{L}^2(X)$

satisfies

$$\inf(f, \mathbb{1} - f) = \inf(U\mathbb{1}_A, U\mathbb{1} - U\mathbb{1}_A) = U\inf(\mathbb{1}_A, \mathbb{1} - \mathbb{1}_A) = U0 = 0.$$

Thus, $\min(f(x), 1 - f(x)) = 0$ for almost every $x \in X$. This implies $f = \mathbb{1}_B$ for a unique $B \in \Sigma(X)$ and we set $T(A) := B$. For $A_1, A_2 \in \Sigma(Y)$ we then have

$$\begin{aligned}\mathbb{1}_{T(A_1 \cap A_2)} &= U(\mathbb{1}_{A_1 \cap A_2}) = U\inf(\mathbb{1}_{A_1}, \mathbb{1}_{A_2}) = \inf(U\mathbb{1}_{A_1}, U\mathbb{1}_{A_2}) = \inf(\mathbb{1}_{T(A_1)}, \mathbb{1}_{T(A_2)}) \\ &= \mathbb{1}_{T(A_1) \cap T(A_2)}\end{aligned}$$

and hence $T(A_1 \cap A_2) = T(A_1) \cap T(A_2)$. Similarly, $T(A_1 \cup A_2) = T(A_1) \cup T(A_2)$. Finally, for $A \in \Sigma(Y)$ we have

$$\mu_X(T(A)) = \|\mathbb{1}_{T(A)}\|^2 = \|U\mathbb{1}_A\|^2 = \|\mathbb{1}_A\| = \mu_Y(A).$$

Therefore T is a measure algebra homomorphism. Since U and U_T agree on every $\mathbb{1}_A$ for $A \in \Sigma(X)$, we obtain $U = U_T$ again from the fact that U and U_T are bounded linear operators and Lemma 1.3.6. □

Corollary 1.3.8. *Let X and Y be probability spaces. Then every Markov embedding $U \colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ restricts to an algebra homomorphism $U|_{\mathrm{L}^\infty(Y)} \colon \mathrm{L}^\infty(Y) \to \mathrm{L}^\infty(X)$, i.e., $U(f \cdot g) = Uf \cdot Ug$ for all $f, g \in \mathrm{L}^\infty(Y)$.*

Proof. By Theorem 1.3.7 we may assume that $U = U_T$ for a measure algebra homomorphism $T \colon \Sigma(Y) \to \Sigma(X)$. By Lemma 1.3.3 (viii), U_T restricts to a linear isometry $U_T|_{\mathrm{L}^\infty(Y)} \colon \mathrm{L}^\infty(Y) \to \mathrm{L}^\infty(X)$. If $f = \mathbb{1}_A$ and $g = \mathbb{1}_B$ for $A, B \in \Sigma(Y)$, then

$$U_T(f \cdot g) = U_T\mathbb{1}_{A \cap B} = \mathbb{1}_{T(A \cap B)} = \mathbb{1}_{T(A) \cap T(B)} = \mathbb{1}_{T(A)} \cdot \mathbb{1}_{T(B)} = U_T f \cdot U_T g.$$

By bilinearity of multiplication, $U_T(f \cdot g) = U_T f \cdot U_T g$ for (equivalence classes of) simple functions $f, g \in \mathrm{L}^\infty(Y)$. Finally, we can approximate general $f, g \in \mathrm{L}^\infty(Y)$ with sequences $(f_n)_{n \in \mathbb{N}}$ and $(g_n)_{n \in \mathbb{N}}$ of simple functions to obtain $U_T(f \cdot g) = \lim_{n\to\infty} U_T(f_n \cdot g_n) = \lim_{n\to\infty} U_T(f_n) \cdot U_T(g_n) = U_T f \cdot U_T g$ with the limit now taken with respect to the L^∞-norm (since multiplication of functions is continuous with respect to this norm). □

Corollary 1.3.9. *Let X be a probability space and Y a Lebesgue space. Then the map $\mathrm{M}(X, Y) \to \mathrm{M}(\mathrm{L}^2(Y), \mathrm{L}^2(X))$, $\tau \mapsto U_\tau$ is a bijection.*

Proof. One can readily check that the map is the composition of the bijections of Theorems 1.2.8 and 1.3.7. □

To conclude the lecture, let us emphasize once more that we have (at least for Lebesgue spaces) three different, but equivalent approaches to measure-preserving transformations between probability spaces X and Y:

(i) measure-preserving maps $\tau\colon X \to Y$,

(ii) measure algebra homomorphisms $T\colon \Sigma(Y) \to \Sigma(X)$, and

(iii) Markov embeddings $U\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$.

To pass from (i) to (ii) we take the pullback. The transition from (ii) to (iii) is achieved by extending the induced map of (equivalence classes of) characteristic functions to a linear isometry between the L^2-spaces.

1.4 Supplement: Realization of Homomorphisms

We now provide the promised proof of Theorem 1.2.8. First recall its content.

Theorem. *For every probability space X and every Lebesgue space Y the map $\mathrm{M}(X,Y) \to \mathrm{M}(\Sigma(Y), \Sigma(X))$, $\tau \mapsto \tau^*$ is a bijection.*

Proof. For a measure algebra homomorphism $T\colon \Sigma(Y) \to \Sigma(X)$ we have to find a unique $\tau \in \mathrm{M}(X,Y)$ with $T = \tau^*$. We first improve the situation: By assumption we find a separable and complete metric space (Z, d_Z), a Borel probability measure $\mu_Z\colon \mathcal{B}(Z) \to [0,1]$ and an isomorphism $\vartheta \in \mathrm{M}(Y,Z)$. Let $\vartheta^{-1} \in \mathrm{M}(Z,Y)$ be its inverse. Then $S \coloneqq T \circ \vartheta^*\colon \Sigma(Z) \to \Sigma(X)$ is again a measure algebra homomorphism. If we find a unique $\sigma \in \mathrm{M}(X,Z)$ with $\sigma^* = S$, then one can check that $\vartheta^{-1} \circ \sigma \in \mathrm{M}(Y,X)$ is the unique $\tau \in \mathrm{M}(X,Y)$ with $\tau^* = T$.

Now pick a countable dense subset $\{z_k \mid k \in \mathbb{N}\}$ of Z. Observe then that, for fixed $n \in \mathbb{N}$, the open balls $\mathrm{B}(z_k, 1/n) \coloneqq \{z \in Z \mid \mathrm{d}_Z(z_k, z) < 1/n\}$ for $k \in \mathbb{N}$ cover the entire space Z.

Uniqueness: Assume that $\sigma_1, \sigma_2\colon X \to Z$ are measure-preserving maps with $(\sigma_1)^* = (\sigma_2)^* = S$. We have to show that the set $M \coloneqq \{x \in X \mid \sigma_1(x) \neq \sigma_2(x)\}$ is a nullset. Since for any $A \in \Sigma_Y$ the symmetric difference of $(\sigma_1)^{-1}(A)$ and $(\sigma_2)^{-1}(A)$ is a nullset, it suffices to show the identity

$$M = \bigcup_{n\in\mathbb{N}} \bigcup_{k\in\mathbb{N}} (\sigma_1)^{-1}(\mathrm{B}(z_k, 1/n)) \cap (X \setminus (\sigma_2)^{-1}(\mathrm{B}(z_k, 1/n)))$$

as this implies that M is a countable union of nullsets, hence a nullset itself.

The inclusion "$\supseteq$" is clear. Conversely, if $x \in X$ with $\sigma_1(x) \neq \sigma_2(x)$, then there is some $n \in \mathbb{N}$ with $\mathrm{d}_Z(\sigma_1(x), \sigma_2(x)) \geq 2/n$. Choose $k \in \mathbb{N}$ with $\sigma_1(x) \in \mathrm{B}(z_k, 1/n)$. By the triangle inequality, $\sigma_2(x)$ cannot also be an element of $\mathrm{B}(z_k, 1/n)$. This shows $x \in (\sigma_1)^{-1}(\mathrm{B}(z_k, 1/n)) \cap (X \setminus (\sigma_2)^{-1}(\mathrm{B}(z_k, 1/n)))$.

Existence: The proof of existence of $\sigma\colon X \to Z$ with $\sigma^* = S$ requires several steps.

Step 1: Construct suitable measurable partitions of X and Z. Fix $n \in \mathbb{N}$. We "disjointify" the cover of open balls from above by setting $A_1^n \coloneqq \mathrm{B}(z_1, 1/n)$ and recursively $A_k^n \coloneqq \mathrm{B}(z_k, 1/n) \setminus \bigcup_{j=1}^{k-1} A_j^n$ for $k \geq 2$. The sets A_k^n for $k \in \mathbb{N}$ then form a measurable partition of Z.

For $k \in \mathbb{N}$ we choose a measurable subset $B_k^n \subseteq X$ with $S(A_k^n) = B_k^n$ in $\Sigma(X)$. Since S is a measure algebra homomorphism, we obtain

$$\mu_X(B_k^n \cap B_l^n) = \mu_X(S(A_k^n \cap A_l^n)) = \mu_Z(A_k^n \cap A_l^n) = 0$$

for $k \neq l$. Moreover, by Lemma 1.2.6 (iv) we have

$$\mu_X\left(\bigcup_{k\in\mathbb{N}} B_k^n\right) = \mu_Z\left(\bigcup_{k\in\mathbb{N}} A_k^n\right) = \mu_Z(Z) = 1.$$

A moment's thought reveals that by replacing B_1^n with $(B_1^n)' := B_1^n \cup X \setminus \bigcup_{l\in\mathbb{N}} B_l^n$ and recursively B_k^n with $(B_k^n)' := B_k^n \setminus \bigcup_{l=1}^{k-1}(B_l^n)'$ for $k \geq 2$, we may assume that the sets B_k^n for $k \in \mathbb{N}$ form a measurable partition of X.

Step 2: Construct approximating sequences of measurable maps. We claim that there are measurable maps $\varrho_n \colon Z \to Z$ and $\sigma_n \colon X \to Z$ for $n \in \mathbb{N}$ with the following two properties.

(i) $S(\varrho_n^{-1}(A)) = (\sigma_n)^{-1}(A)$ in $\Sigma(X)$ for every measurable $A \subseteq Z$ and $n \in \mathbb{N}$.

(ii) $\lim_{n\to\infty} \varrho_n(z) = z$ for every $z \in Z$.

To construct these, notice that for every $z \in Z$ there is precisely one $k \in \mathbb{N}$ with $z \in A_k^n$. We then set $\varrho_n(z) := z_k$. Similarly, for every $x \in X$ we find exactly one $k \in \mathbb{N}$ with $x \in B_k^n$ and we define $\sigma_n(x) := z_k$. As the preimages of any subset of Z with respect to ϱ_n and σ_n are countable unions of the measurable sets A_k^n and B_k^n, respectively, both maps are measurable. To check property (i) observe that for any measurable subset $A \subseteq Z$ and $n \in \mathbb{N}$ the identities

$$\varrho_n^{-1}(A) = \bigcup_{\substack{k\in\mathbb{N}\\ z_k\in A}} A_k^n \quad \text{and} \quad \sigma_n^{-1}(A) = \bigcup_{\substack{k\in\mathbb{N}\\ z_k\in A}} B_k^n$$

hold. Thus,

$$S(\varrho_n^{-1}(A)) = S\left(\bigcup_{\substack{k\in\mathbb{N}\\ z_k\in A}} A_k^n\right) = \bigcup_{\substack{k\in\mathbb{N}\\ z_k\in A}} B_k^n = (\sigma_n)^{-1}(A)$$

in $\Sigma(X)$ by Lemma 1.2.6 (iv). For property (ii) note that for $z \in Z$ and $n \in \mathbb{N}$ we have $z \in \mathrm{B}(\varrho_n(z), 1/n)$ by definition of the map ϱ_n and the sets A_k^n for $k \in \mathbb{N}$. But this implies $\lim_{n\to\infty} \varrho_n(z) = z$.

Step 3: Construct the measure-preserving map σ. We show that for almost every $x \in X$ the sequence $(\sigma_n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence in Z. First fix $n, m \in \mathbb{N}$. Then the sets $B_k^n \cap B_l^m$ for $k, l \in \mathbb{N}$ form a measurable partition of X. If $k, l \in \mathbb{N}$ with $\mu_X(B_k^n \cap B_l^m) > 0$, then $\mu_Y(A_k^n \cap A_l^m) > 0$, and hence there is some $z \in A_k^n \cap A_l^m$. For every $x \in B_k^n \cap B_l^m$ we therefore obtain

$$d_Z(\sigma_n(x), \sigma_m(x)) = d_Z(z_k, z_l) \leq d_Z(z_k, z) + d_Z(z, z_l) < \frac{1}{n} + \frac{1}{m}.$$

by definition of the sets A_k^n and A_l^m. Thus, $d_Z(\sigma_n(x), \sigma_m(x)) < \frac{1}{n} + \frac{1}{m}$ for almost every $x \in X$. Since the set $\mathbb{N} \times \mathbb{N}$ is countable, we even find a nullset $N \subseteq X$ such that $d(\sigma_n(x), \sigma_m(x)) \leq \frac{1}{n} + \frac{1}{m}$ holds for all $x \in X \setminus N$ and *all* $n, m \in \mathbb{N}$. Thus, $(\sigma_n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence for all $x \in X \setminus N$.

Redefining the sequence σ_n on a nullset (setting it to some fixed, arbitrary value), we still have property (i) above but can achieve that $(\sigma_n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence for *every* $x \in X$. Since the metric space (Z, d_Z) is complete, the limit $\sigma(x) := \lim_{n\to\infty} \sigma_n(x)$ exists for all $x \in X$. Moreover, $\sigma\colon X \to Z$ is measurable as the pointwise limit of a sequence of measurable maps (see, e.g., [Kal97, Lemma 1.10]). For each measurable $A \subseteq Z$ we obtain by property (i), Lemma 1.2.6 (iii), and Exercise 1.4 that

$$\begin{aligned}\mu_X(S(A)\,\Delta\,\sigma^{-1}(A)) &\leq \mu_X(S(A)\,\Delta\,S(\varrho_n(A))) + \mu_X(S(\varrho_n(A))\,\Delta\sigma^{-1}(A))\\ &= \mu_Z(A\,\Delta\,\varrho_n^{-1}(A)) + \mu_X(\sigma_n^{-1}(A)\,\Delta\,\sigma^{-1}(A)).\end{aligned}$$

Since $\lim_{n\to\infty} \varrho_n(z) = z$ for every $z \in Z$ by property (ii) above, and $\lim_{n\to\infty} \sigma_n(x) = \sigma(x)$ for every $x \in X$ by definition of σ, we obtain from Lebesgue's theorem,

$$\begin{aligned}&\lim_{n\to\infty} \mu_Z(A\,\Delta\,\varrho_n^{-1}(A)) = \lim_{n\to\infty} \|\mathbb{1}_A - \mathbb{1}_A \circ \varrho_n\|_1 = 0, \text{ and}\\ &\lim_{n\to\infty} \mu_X(\sigma_n^{-1}(A)\,\Delta\,\sigma^{-1}(A))) = \lim_{n\to\infty} \|\mathbb{1}_A \circ \sigma_n - \mathbb{1}_A \circ \sigma\|_1 = 0.\end{aligned}$$

Thus, $S(A) = \sigma^{-1}(A)$ in $\Sigma(X)$. In particular, $\mu_X(\sigma^{-1}(A)) = \mu_X(S(A)) = \mu_Z(A)$. Therefore, $\sigma\colon X \to Z$ is measure-preserving with $\sigma^* = S$. $\square$

1.5 Comments and Further Reading

What is now known as Poincaré's recurrence theorem has its roots in the work [Poi90] of the French mathematician and physicist Henri Poincaré on "the three body problem" and predates modern ergodic theory (and abstract measure theory). For more information, also on the physical context, see, e.g., [EW11, Section 2.2], [EFHN15, Section 6.2], and [VO16, Chapter 1].

The abstract approach to measure-preserving transformations, formalized using measure algebras, has been implicitly present in the literature; see, e.g., [Fur14, Chapter 5] and [Gla03, Chapter 2]. This "point-free" framework of ergodic theory has been systematically developed in a series of papers by the first author and Terence Tao, see [JT23a, JT23b, JT22].

Taking pullbacks of maps on the level of sets or functions is a common procedure in many mathematical areas. Bernhard Koopman, a coauthor of John von Neumann, introduced his operator in [Koo31] and paved the way for applying functional analysis to study measure-preserving maps. This operator theoretic approach to ergodic theory is thoroughly discussed in the book manuscript [DNP87], and the monograph [EFHN15], which is based on the 12th edition of the ISem from 2008/09 [EFHN09]. In particular, we refer to [EFHN15, Chapter 13] for more information on Markov embeddings.

An early version of Theorem 1.2.8 on the connection between measure-preserving maps and measure algebra homomorphisms was already proven by von Neumann in [vN32a], see also [EFHN15, Sections 6.1 and 7.3]. Our proof of Theorem 1.2.8 follows the one of [JT23b] (see Proposition 3.2 there), where the realization of measure algebra homomorphisms as pullbacks is thoroughly discussed.

1.6 Exercises

Exercise 1.1. Let $\tau\colon X \to Y$ be a measure-preserving map between probability spaces. Assume that $\sigma\colon Y \to X$ is measurable with $(\sigma \circ \tau)(x) = x$ for almost every $x \in X$. Show that σ is measure-preserving.

Exercise 1.2. (i) Equip $X = [0,1)$ with the Borel σ-algebra and the Lebesgue measure, and consider the **doubling map** $\tau\colon X \to X$, $x \mapsto 2x \mod 1$. Show that τ is a measurable and measure-preserving map which does *not* define an invertible element of $\mathrm{M}(X,X)$.

(ii) Equip $Y = [0,1] \times [0,1]$ with the Borel σ-algebra and the (two-dimensional) Lebesgue measure, and consider the **baker's transformation**

$$\sigma\colon Y \to Y, \quad (x,y) \mapsto \begin{cases} (2x, y/2) & \text{if } 0 \le x < 1/2, \\ (2x-1, y+1/2) & \text{if } 1/2 \le x \le 1. \end{cases}$$

Show that σ is a measurable and measure-preserving map which defines an invertible element of $\mathrm{M}(Y,Y)$.
Hint: Use Proposition 1.1.4.

Exercise 1.3. Let $\tau\colon X \to X$ be a measure-preserving map on a probability space X and $A \subseteq X$ a measurable subset with $\mu_X(A) > 0$. Show that there are infinitely many $n \in \mathbb{N}$ with $\mu_X(A \cap \tau^{-n}(A)) > 0$.

Exercise 1.4. Let X be a probability space. We set

$$d(A,B) := \mu_X(A \,\Delta\, B) = \int_X |\mathbb{1}_A - \mathbb{1}_B| = \|\mathbb{1}_A - \mathbb{1}_B\|_1$$

for $A, B \in \Sigma(X)$. Show the following assertions.

(i) d is a metric on $\Sigma(X)$.

(ii) The metric space $(\Sigma(X), d)$ is complete.

(iii) $d(X \setminus A, X \setminus B) = d(A,B)$ for all $A, B \in \Sigma(X)$.

(iv) $d(A \cap B, C \cap D) \le d(A,C) + d(B,D)$ for all $A,B,C,D \in \Sigma(X)$.

(v) $d(A \setminus B, C \setminus D) \le d(A,C) + d(B,D)$ for all $A,B,C,D \in \Sigma(X)$.

(vi) $d(A \cup B, C \cup D) \le d(A,C) + d(B,D)$ for all $A,B,C,D \in \Sigma(X)$.

(vii) $|\mu_X(A) - \mu_X(B)| \le d(A,B)$ for all $A, B \in \Sigma(X)$.

(viii) $|\mu_X(A \cap B) - \mu_X(C \cap D)| \le d(A,C) + d(B,D)$ for all $A,B,C,D \in \Sigma(X)$.

(ix) If $(A_n)_{n\in\mathbb{N}}$ is a sequence in $\Sigma(X)$ with $\mu_X(A_n \setminus A_{n+1}) = 0$ for every $n \in \mathbb{N}$, then $\lim_{n\to\infty} A_n = \bigcup_{n\in\mathbb{N}} A_n$ with respect to d.

Exercise 1.5. Prove Lemma 1.2.6.

Exercise 1.6. Let X be a probability space. For $A, B \in \Sigma(X)$ write $A \leq B$ if $\mu_X(A \setminus B) = 0$. Show the following assertions.

(i) $\leq$ is a partial order on $\Sigma(X)$.

(ii) For a probability space Y, a measure algebra homomorphism $T\colon \Sigma(Y) \to \Sigma(X)$, and $A, B \in \Sigma(Y)$, we have $A \leq B$ precisely when $T(A) \leq T(B)$.

Recall for any partially ordered set Λ and a subset $M \subseteq \Lambda$, the smallest element $s \in \Lambda$ with $m \leq s$ for every $m \in M$ (if it exists) is called the **supremum** $\sup M$ of M. We now investigate suprema in $\Sigma(X)$ with respect to the partial order above. Show the following assertions.

(iii) For a countable subset $M \subseteq \Sigma(X)$ the supremum $\sup M$ exists and is given by $\sup M = \bigcup_{A \in M} A$.

(iv) For any subset $M \subseteq \Sigma(X)$ the supremum $\sup M$ exists and there is a countable subset $N \subseteq M$ with $\sup M = \sup N = \bigcup_{A \in N} A$.
Hint: Find a sequence $(N_k)_{k \in N}$ of countable subsets $N_k \subseteq M$ with $N_k \subseteq N_{k+1}$ for every $k \in \mathbb{N}$ as well as

$$\lim_{k \to \infty} \mu_X\Bigg(\bigcup_{A \in N_k} A\Bigg) = \sup\Bigg\{\mu_X\Bigg(\bigcup_{A \in C} A\Bigg) \mid C \subseteq M \text{ countable}\Bigg\},$$

and then consider $N := \bigcup_{k \in \mathbb{N}} N_k$.

(v) For a probability space Y and a measure algebra homomorphism $T\colon \Sigma(Y) \to \Sigma(X)$ we have $T(\sup M) = \sup T(M)$ for every subset $M \subseteq \Sigma(Y)$.

Exercise 1.7. Give an example of probability spaces X and Y and distinct $\sigma, \tau \in \mathrm{M}(X, Y)$ with $\sigma^* = \tau^*$.
Hint: Equip $\{0, 1\}$ with the trivial σ-algebra $\{\emptyset, \{0, 1\}\}$.

Lecture 2

In this second lecture we introduce the key objects of the course: Measure-preserving systems. We then discuss Bernoulli shifts as an important class of examples. Finally, we introduce the property of "ergodicity" and study different approaches to subsystems of a measure-preserving system.

2.1 Measure-Preserving Systems

In ergodic theory, we are interested in the asymptotic behavior of measure-preserving transformations. In this course, we will also assume that these are invertible, and introduce the following notation for a probability space X.

(i) $\mathrm{Aut}(X) := \{\tau \in \mathrm{M}(X, X) \mid \tau \text{ invertible}\}$,

(ii) $\mathrm{Aut}(\Sigma(X)) := \{T \in \mathrm{M}(\Sigma(X), \Sigma(X)) \mid T \text{ bijective}\}$, and

(iii) $\mathrm{Aut}(\mathrm{L}^2(X)) := \{U \in \mathrm{M}(\mathrm{L}^2(X), \mathrm{L}^2(X)) \mid U \text{ bijective}\}$.

Elements of these sets are called **automorphisms** of X, **measure algebra automorphisms** of $\Sigma(X)$, and **Markov automorphisms** of $\mathrm{L}^2(X)$, respectively.

Note that all these sets equipped with the natural compositions are groups. Theorem 1.3.7 yields a group isomorphism between $\mathrm{Aut}(\Sigma(X))$ and $\mathrm{Aut}(\mathrm{L}^2(X))$ (cf. Remark 1.3.5). If X is a Lebesgue space, then by Theorem 1.2.8 (and Remark 1.2.4) all three groups are isomorphic.

We could now study the behavior of the iterates of an element in any one of these automorphism groups. However, we start from a more general definition of measure-preserving systems.

Let Γ be an arbitrary (discrete) abelian[1] group (e.g., $\Gamma = \mathbb{Z}^n$, $n \geq 1$, $\Gamma = \mathbb{F}_p^\omega = \bigoplus_{n=1}^\infty \mathbb{F}_p$, where $\mathbb{F}_p$ is the finite field of prime order p) and fix this for now.

[1]Some topics of these lectures can be treated in the framework of arbitrary groups (or even semigroups). However, we restrict to the case of abelian groups throughout the entire course.

Definition 2.1.1. An **(abelian) measure-preserving system** (X, T) consists of a probability space X and a group homomorphism

$$T\colon \Gamma \to \mathrm{Aut}(\Sigma(X)), \quad \gamma \mapsto T_\gamma,$$

i.e., $T_{\gamma_1+\gamma_2} = T_{\gamma_1}T_{\gamma_2}$ for $\gamma_1, \gamma_2 \in \Gamma$.

For a measure-preserving system (X, T) the induced group homomorphism

$$U_T\colon \Gamma \to \mathrm{Aut}(\mathrm{L}^2(X)), \quad \gamma \mapsto U_{T_\gamma}$$

is the **Koopman representation** of (X, T).

The following are trivial, but important examples of measure-preserving systems.

Example 2.1.2. For any probability space X we have a system (X, Id) given by the group homomorphism $\mathrm{Id}\colon \Gamma \to \mathrm{Aut}(\Sigma(X))$, $\gamma \mapsto \mathrm{Id}_{\Sigma(X)}$. A very special case is the **trivial system** $(\{0\}, \mathrm{Id})$ where $\{0\}$ is equipped with the point measure δ_0 (cf. Example 1.1.3 (ii)).

Most examples are given by concrete measure-preserving maps. We therefore introduce a second type of measure-preserving systems.

Definition 2.1.3. A **concrete measure-preserving system** (X, τ) consists of a probability space X and a group homomorphism

$$\tau\colon \Gamma \to \mathrm{Aut}(X), \quad \gamma \mapsto \tau_\gamma.$$

Remark 2.1.4. One has to be a bit careful here: If we choose a concrete measure-preserving map representing τ_γ for $\gamma \in \Gamma$, the identity $\tau_{\gamma_1+\gamma_2} = \tau_{\gamma_1}\tau_{\gamma_2}$ only holds outside of a nullset depending on γ_1 and γ_2. In particular, notions like the orbit of a point do not make sense in this framework.[2]

We immediately obtain the following fact (cf. Remark 1.3.5).

Proposition 2.1.5. *Every concrete measure-preserving system* (X, τ) *gives rise to a measure-preserving system* (X, τ^*) *as in Definition 2.1.1 by setting*

$$(\tau^*)_\gamma \coloneqq (\tau_\gamma)^*\colon \Sigma(X) \to \Sigma(X), \quad A \mapsto \tau_\gamma^{-1}(A)$$

for $\gamma \in \Gamma$.

If X is a Lebesgue space, then – as a consequence of Theorem 1.2.8 – every measure-preserving system (X, T) is of the form (X, τ^*) for a uniquely determined concrete measure-preserving system (X, τ), see Exercise 2.6 below.

[2] In the case of a countable group Γ (which suffices for most applications), one can circumvent this problem and transition to a proper action of Γ on X by measure-preserving maps. However, we avoid these measure-theoretic intricacies altogether by using Definition 2.1.1 and working on the level of measure algebras.

Remark 2.1.6. Notice that if $\Gamma = \mathbb{Z}$, then a measure-preserving system (X, T) is completely determined by $T_1\colon \Sigma(X) \to \Sigma(X)$ since $T_m = T_1^m$ for every $m \in \mathbb{Z}$. Conversely, if $S\colon \Sigma(X) \to \Sigma(X)$ is a measure algebra automorphism, then $T\colon \mathbb{Z} \to \mathrm{Aut}(\Sigma(X))$, $n \mapsto S^n$ yields a measure-preserving system with respect to the group $\mathbb{Z}$. This establishes a one-to-one correspondence between measure-preserving systems over $\mathbb{Z}$ and automorphisms $S\colon \Sigma(X) \to \Sigma(X)$. We will occasionally abuse notation and, for an automorphism $S\colon \Sigma(X) \to \Sigma(X)$, refer to (X, S) as a measure-preserving system over $\mathbb{Z}$. An analogous correspondence holds for concrete measure-preserving systems (X, τ) over $\mathbb{Z}$ and elements $\sigma \in \mathrm{Aut}(X)$, and we use a similar convention.

In view of the previous remark, we already obtain some examples of concrete measure-preserving systems from the previous lecture. To introduce another interesting and important class of examples, we need the concept of product measures for possibly infinitely many spaces (see [HS65, Section 22]).

Proposition and Definition 2.1.7. *Let $I \neq \emptyset$ be an index set and X_i a probability space for each $i \in I$. For a finite set $\{i_1, \dots, i_m\} \subseteq I$ and measurable sets $A_j \in \Sigma_{X_{i_j}}$ for $j \in \{1, \dots, m\}$ we call*

$$\mathrm{Z}(i_1, \dots, i_m; A_1, \dots, A_m) \coloneqq \left\{ (x_i)_{i \in I} \in \prod_{i \in I} X_i \mid x_{i_j} \in A_j \text{ for all } j \in \{1, \dots, m\} \right\}$$

*a **(measurable) cylinder set**. The smallest σ-algebra over $\prod_{i\in I} X_i$ containing all such cylinder sets is the **product-σ-algebra** $\bigotimes_{i\in I} \Sigma_{X_i}$. There is a unique probability measure $\bigotimes_{i\in I} \mu_{\mathrm{X}_i}\colon \bigotimes_{i\in I} \Sigma_{X_i} \to [0,1]$, called the **product measure**, with*

$$\bigotimes_{i\in I} \mu_{\mathrm{X}_i}(\mathrm{Z}(i_1, \dots, i_m; A_1, \dots, A_m)) = \mu_{X_{i_1}}(A_1) \cdots \mu_{X_{i_m}}(A_m)$$

for each cylinder set $\mathrm{Z}(i_1, \dots, i_m; A_1, \dots, A_m)$.

If $I \neq \emptyset$ is an index set and X is a probability space, we also write X^I for the product probability space $\prod_{i\in I} X$ (i.e., in case $X_i = X$ for all $i \in I$). With this notation, we introduce a new example for a measure-preserving system.

Example 2.1.8. Let X be any probability space, e.g., $X = \{0, \dots, k-1\}$ with the probability measure defined by the probability vector $p = (\frac{1}{k}, \dots, \frac{1}{k})$, see Example 1.1.3 (ii). Then $\tau\colon \Gamma \to \mathrm{Aut}(X^\Gamma)$, $\gamma \mapsto \tau_\gamma$ with $\tau_\gamma(x_\delta)_{\delta\in\Gamma} = (x_{\delta+\gamma})_{\delta\in\Gamma}$ for $(x_\delta)_{\delta\in\Gamma} \in X^\Gamma$ and $\gamma \in \Gamma$ defines a concrete measure-preserving system (X^Γ, τ): For $\gamma \in \Gamma$ and a cylinder set $\mathrm{Z}(\delta_1, \dots, \delta_m; A_1, \dots, A_m)$ we obtain

$$\begin{aligned}\tau_\gamma^{-1}(\mathrm{Z}(\delta_1, \dots, \delta_m; A_1, \dots, A_m)) &= \{(x_\delta)_{\delta\in\Gamma} \mid x_{\delta_j+\gamma} \in A_j \text{ for every } j \in \{1, \dots, m\}\} \\ &= \mathrm{Z}(\delta_1 + \gamma, \dots, \delta_m + \gamma; A_1, \dots, A_m),\end{aligned}$$

and therefore

$$\begin{aligned}\mu_{X^\Gamma}(\tau_\gamma^{-1}(\mathrm{Z}(\delta_1,\dots,\delta_m;A_1,\dots,A_m))) &= \mu_{X^\Gamma}(\mathrm{Z}(\delta_1+\gamma,\dots,\delta_m+\gamma;A_1,\dots,A_m))\\ &= \mu_X(A_1)\cdots\mu_X(A_m)\\ &= \mu_{X^\Gamma}(\mathrm{Z}(\delta_1,\dots,\delta_m;A_1,\dots,A_m)).\end{aligned}$$

Since cylinder sets form a $\cap$-stable generator of the σ-algebra Σ_{X^Γ}, we obtain that the map $\tau_\gamma\colon X^\Gamma \to X^\Gamma$ is measurable and measure-preserving for every $\gamma \in \Gamma$ by Proposition 1.1.4. Moreover, it is clear that $\tau_{\gamma_1+\gamma_2} = \tau_{\gamma_1} \circ \tau_{\gamma_2}$ for all $\gamma_1, \gamma_2 \in \Gamma$. We call the system (X^Γ, τ) (and also the induced measure-preserving system (X^Γ, τ^*) on the level of measure algebras) a **Bernoulli shift**.

We now introduce a crucial irreducibility property for measure-preserving systems.

Definition 2.1.9. For a measure-preserving system (X, T) we say that $A \in \Sigma(X)$ is **invariant** if $T_\gamma(A) = A$ for every $\gamma \in \Gamma$. We set

$$\Sigma(X)_{\mathrm{inv}} := \{A \in \Sigma(X) \mid A \text{ invariant}\} \subseteq \Sigma(X).$$

The system (X, T) is called **ergodic** if every invariant $A \in \Sigma(X)$ already satisfies $\mu_X(A) \in \{0, 1\}$, i.e., $\Sigma(X)_{\mathrm{inv}} = \{\emptyset, X\}$.

Examples 2.1.10. (i) For a system with trivial dynamics (X, Id) (see Example 2.1.2) every element $A \in \Sigma(X)$ is invariant. Hence, such a system (X, Id) is ergodic precisely when $\Sigma(X) = \{\emptyset, X\}$. Thus, (X, Id) is "almost never" ergodic.

(ii) Consider the set $X = \{0, 1\}$ with the probability measure defined by the probability vector $(\frac{1}{2}, \frac{1}{2})$ and the measure preserving system (X, τ^*) over $\Gamma = \mathbb{Z}$ defined by the measure-preserving map $\tau\colon X \to X$, $m \mapsto 1 - m$, see Example 1.1.3 (ii) and Remark 2.1.6. Then (X, τ^*) is ergodic.

What about Bernoulli shifts? For the trivial group $\Gamma = \{0\}$, we have $(X^\Gamma, \tau^*) = (X^\Gamma, \mathrm{Id})$, and this is only ergodic if $\Sigma(X) = \{\emptyset, X\}$. The same is true for any finite abelian group Γ (see Exercise 2.4). However, for infinite groups we obtain the following.

Proposition 2.1.11. *If the group Γ is infinite, then the Bernoulli shift (X^Γ, τ^*) is ergodic. Moreover, for all measurable subsets $A, B \subseteq X^\Gamma$ and every $\varepsilon > 0$ there is a finite subset $F \subseteq \Gamma$ such that*

$$|\mu_{X^\Gamma}(\tau_\gamma^{-1}(A) \cap B) - \mu_{X^\Gamma}(A) \cdot \mu_{X^\Gamma}(B)| \leq \varepsilon \quad \textit{for every } \gamma \in \Gamma \setminus F.$$

For the proof of Proposition 2.1.11 we need two statements from measure theory. The first one is a general approximation result for measurable sets, see, e.g., [Bil95, Theorem 11.4].

Lemma 2.1.12. *Let X be a probability space and $\mathcal{E} \subseteq \Sigma_X$ a generator of the σ-algebra which is an algebra over X, i.e., $\emptyset, X \in \mathcal{E}$ and $A \cup B, A \cap B, A \setminus B \in \mathcal{E}$ whenever $A, B \in \mathcal{E}$. Then for every measurable subset $A \subseteq X$ and $\varepsilon > 0$ there is $B \in \mathcal{E}$ with $\mu_X(A \,\Delta\, B) \leq \varepsilon$.*

The second auxiliary result is a compatibility statement for product measures. If we decompose the index set of a product space into two parts, then the product measure can also be written as a product of two product measures:

Lemma 2.1.13. *Assume that $I \neq \emptyset$ is an index set and X_i is a probability space for every $i \in I$. If for $\emptyset \neq J \subsetneq I$ we naturally identify*

$$\prod_{i \in I} X_i = \prod_{i \in J} X_i \times \prod_{i \in I \setminus J} X_i$$

as sets, then

$$\bigotimes_{i \in I} \Sigma_{X_i} = \bigotimes_{i \in J} \Sigma_{X_i} \otimes \bigotimes_{i \in I \setminus J} \Sigma_{X_i} \quad \text{and} \quad \bigotimes_{i \in I} \mu_{X_i} = \bigotimes_{i \in J} \mu_{X_i} \otimes \bigotimes_{i \in I \setminus J} \mu_{X_i}.$$

A proof can be found in [HS65, Lemmas 22.4 and 22.12].

Proof of Proposition 2.1.11. We show the second claim first and start with particularly simple measurable sets $A, B \in \Sigma_{X^\Gamma}$. For every finite subset $E \subseteq \Gamma$ we may identify $X^\Gamma = X^E \times X^{\Gamma \setminus E}$, hence for each $C \in \Sigma_{X^E}$ the set $C \times X^{\Gamma \setminus E}$ defines a measurable subset $\mathrm{Z}(E, C)$ of X^Γ by Lemma 2.1.13. We make the following basic observations about these measurable sets which "depend only on a finite number of coordinates":

(i) For disjoint finite subsets $E, E' \subseteq \Gamma$, $C \in \Sigma_{X^E}$, and $C' \in \Sigma_{X^{E'}}$ we have $\mathrm{Z}(E, C) \cap \mathrm{Z}(E', C') = \mathrm{Z}(E \cup E', C \times C')$.

(ii) For every finite subset $E \subseteq \Gamma$ and $C \in \Sigma_{X^E}$ we have $\mu_{X^\Gamma}(\mathrm{Z}(E, C)) = \mu_{X^E}(C)$.

(iii) If $E \subseteq \Gamma$ is a finite subset and $\gamma \in \Gamma$, then (since E and $E + \gamma$ have the same cardinality) we can identify X^E with $X^{E+\gamma}$. With this identification, we obtain $\tau_\gamma^{-1}(\mathrm{Z}(E, C)) = \mathrm{Z}(E + \gamma, C)$ for every $C \in X^E$.

Now set $\mathcal{E} := \{\mathrm{Z}(E, C) \mid E \subseteq \Gamma \text{ finite}, C \in \Sigma_{X^E}\}$. It is easy to check that $\mathcal{E}$ is an algebra over X^Γ, and, since it contains all measurable cylinder sets, it generates the σ-algebra Σ_{X^Γ}. We now show the second statement for sets $A, B \in \mathcal{E}$, even in a stronger version. The idea is to shift A far enough that it "lives on coordinates disjoint from those of B".

To make this precise, take finite subsets $E, E' \subseteq \Gamma$ as well as measurable sets $C \in \Sigma_{X^E}$ and $C' \in \Sigma_{X^{E'}}$. Then $F := E' - E = \{\gamma' - \gamma \mid \gamma' \in E', \gamma \in E\}$ is a finite subset of Γ. For $\gamma \in \Gamma \setminus F$, the sets $E+\gamma$ and E' are disjoint. Using properties (i) – (iii) above and Lemma 2.1.13, we therefore obtain

$$\begin{aligned}\mu_{X^\Gamma}(\tau_\gamma^{-1}(\mathrm{Z}(E,C)) \cap \mathrm{Z}(E',C')) &= \mu_{X^\Gamma}(\mathrm{Z}(E+\gamma,C) \cap \mathrm{Z}(E',C')) \\ &= \mu_{X^\Gamma}(\mathrm{Z}((E+\gamma) \cup E', C \times C') = \mu_{X^{(E+\gamma)\cup E'}}(C \times C') \\ &= \mu_{X^{E+\gamma}}(C) \cdot \mu_{X^{E'}}(C') = \mu_{X^E}(C) \cdot \mu_{X^{E'}}(C') \\ &= \mu_{X^\Gamma}(\mathrm{Z}(E,C)) \cdot \mu_{X^\Gamma}(\mathrm{Z}(E',C')).\end{aligned}$$

For general measurable subsets $A, B \subseteq X$ we establish the claimed inequality by using an approximation argument. For $\varepsilon > 0$ we find by Lemma 2.1.12 sets $A', B' \in \mathcal{E}$ with $\mu_{X^\Gamma}(A \,\Delta A') \leq \frac{\varepsilon}{4}$ and $\mu_{X^\Gamma}(B \,\Delta B') \leq \frac{\varepsilon}{4}$. By the above, we find a finite subset $F \subseteq \Gamma$ with $\mu_{X^\Gamma}(\tau_\gamma^{-1}(A') \cap B') = \mu_{X^\Gamma}(A') \cdot \mu_{X^\Gamma}(B')$ for all $\gamma \in \Gamma \setminus F$. For each $\gamma \in \Gamma$ we obtain by Exercise 1.4 and Lemma 1.2.6 (iii),

$$\begin{aligned}|\mu_{X^\Gamma}(\tau_\gamma^{-1}(A) \cap B) - \mu_{X^\Gamma}(\tau_\gamma^{-1}(A') \cap B')| &\leq \mu_{X^\Gamma}(\tau_\gamma^{-1}(A) \,\Delta\, \tau_\gamma^{-1}(A')) + \mu_{X^\Gamma}(B \,\Delta\, B') \\ &= \mu_{X^\Gamma}(A \,\Delta\, A') + \mu_{X^\Gamma}(B \,\Delta\, B') \leq \frac{\varepsilon}{2},\end{aligned}$$

and, by the triangle inequality and Exercise 1.4,

$$\begin{aligned}|\mu_{X^\Gamma}(A)\mu_{X^\Gamma}(B) - \mu_{X^\Gamma}(A')\mu_{X^\Gamma}(B')| &\leq |\mu_{X^\Gamma}(A) - \mu_{X^\Gamma}(A')| + |\mu_{X^\Gamma}(B) - \mu_{X^\Gamma}(B')| \\ &\leq \mu_{X^\Gamma}(A \,\Delta\, A') + \mu_{X^\Gamma}(B \,\Delta\, B') \leq \frac{\varepsilon}{2}.\end{aligned}$$

Applying the triangle inequality once more, we obtain the desired estimate.

To prove ergodicity, let $A \in \Sigma(X)$ be invariant. For $\varepsilon > 0$ we find a finite subset $F \subseteq \Gamma$ with

$$|\mu_{X^\Gamma}((\tau_\gamma)^*(A) \cap A) - \mu_{X^\Gamma}(A) \cdot \mu_{X^\Gamma}(A)| \leq \varepsilon$$

for every $\gamma \in \Gamma \setminus F$. Since Γ is infinite, such a γ actually exists. From A being invariant we conclude $(\tau_\gamma)^*(A) = A$, and therefore $\mu_{X^\Gamma}((\tau_\gamma)^*(A) \cap A) = \mu_{X^\Gamma}(A)$. This implies, $|\mu_{X^\Gamma}(A) - \mu_{X^\Gamma}(A)^2| \leq \varepsilon$. Since $\varepsilon > 0$ was arbitrarily chosen, we obtain $\mu_{X^\Gamma}(A)^2 = \mu_{X^\Gamma}(A)$, and therefore $\mu_{X^\Gamma}(A) \in \{0,1\}$. □

2.2 Subsystems and Extensions

To understand the relation between different measure-preserving systems we also need "structure-preserving maps" between them. This leads to the following definition.

Definition 2.2.1. An **extension** (or **homomorphism**) $J\colon (Y,S) \to (X,T)$ of measure-preserving systems is a measure algebra homomorphism $J\colon \Sigma(Y) \to \Sigma(X)$ such that the diagram

$$\begin{array}{ccc} \Sigma(X) & \xrightarrow{T_\gamma} & \Sigma(X) \\ {\scriptstyle J}\uparrow & & \uparrow{\scriptstyle J} \\ \Sigma(Y) & \xrightarrow{S_\gamma} & \Sigma(Y) \end{array}$$

commutes for every $\gamma \in \Gamma$. In this case, we call (Y,S) a **subsystem** of (X,T). Moreover, J is an **isomorphism** of measure-preserving systems if, in addition, J is bijective.

On the level of concrete measure-preserving systems, the "arrows reverse", and we speak of factors instead of subsystems:

Definition 2.2.2. A **factor map** $q\colon (X,\tau) \to (Y,\sigma)$ of concrete measure-preserving systems is an element $q \in \mathrm{M}(X,Y)$ such that the identity $\sigma_\gamma \circ q = q \circ \tau_\gamma$ holds in $\mathrm{M}(X,Y)$ for all $\gamma \in \Gamma$. In this case, we call (Y,σ) a **factor** of (X,τ). Moreover, q is an **isomorphism** of concrete measure-preserving systems if, in addition, q is an isomorphism of probability spaces.

A simple example is the following.

Example 2.2.3. For $\Gamma = \mathbb{Z}$ and $\alpha \in [0,1)$ consider the system (X,τ) defined by the measure-preserving map $[0,1) \to [0,1)$, $x \mapsto x + \alpha \mod 1$ from Example 1.1.3 (iii) (cf. Remark 2.1.6). If we pick another $\beta \in [0,1)$, the addition map $q\colon [0,1) \to [0,1)$, $x \mapsto x + \beta \mod 1$ defines an isomorphism $q\colon (X,\tau) \to (X,\tau)$ of concrete measure-preserving systems (since addition modulo 1 is commutative).

The following analogue of Proposition 2.1.5 is evident.

Proposition 2.2.4. *If $q\colon (X,\tau) \to (Y,\sigma)$ is a factor map of concrete measure-preserving systems, then $q^*\colon \Sigma(Y) \to \Sigma(X)$, $A \mapsto q^{-1}(A)$ defines an extension $q^*\colon (Y,\sigma^*) \to (X,\tau^*)$ in the sense of Definition 2.2.1. If q is an isomorphism, then so is q^*.*

We now discuss an alternative perspective on subsystems and extensions.

Proposition and Definition 2.2.5. *If $J\colon (Y,S) \to (X,T)$ is an extension of measure-preserving systems, then $\Lambda := J(\Sigma(Y)) \subseteq \Sigma(X)$ is an **invariant sub-σ-algebra**, i.e.,*

(i) $\bigcup_{n\in\mathbb{N}} A_n \in \Lambda$ *whenever* $A_n \in \Lambda$ *for* $n \in \mathbb{N}$,

(ii) $X \setminus A \in \Lambda$ *for every* $A \in \Lambda$,

(iii) $\emptyset, X \in \Lambda$, *and*

(iv) $T_\gamma(A) \in \Lambda$ *for all* $A \in \Lambda$ *and* $\gamma \in \Gamma$.

Proof. Parts (i)–(iii) are a direct consequence of Lemma 1.2.6 (i), (ii), and (iv). For part (iv) consider $A = J(B) \in J(\Sigma(Y))$ for some $B \in \Sigma(Y)$. For $\gamma \in \Gamma$ we then obtain

$$T_\gamma(A) = T_\gamma(J(B)) = (T_\gamma \circ J)(B) = (J \circ S_\gamma)(B) = J(S_\gamma(B)) \in J(\Sigma(Y)).$$

□

Remark 2.2.6. Note that the associated invariant sub-σ-algebra completely determines a subsystem up to an isomorphism: If $J_1\colon (Y_1, S_1) \to (X, T)$ and $J_2\colon (Y_2, S_2) \to (X, T)$ are extensions with $J_1(\Sigma(Y_1)) = J_2(\Sigma(Y_2))$, then $J_2^{-1} \circ J_1\colon \Sigma(Y_1) \to \Sigma(Y_2)$ defines an isomorphism between (Y_1, S_1) and (Y_2, S_2).

We now follow the converse direction: For a given measure-preserving system (X, T) and an invariant sub-σ-algebra $\Lambda \subseteq \Sigma(X)$ we build a subsystem of (X, T) in the following way:

(i) The set $\Sigma_\Lambda := \{A \subseteq X \text{ measurable} \mid [A] \in \Lambda\}$ is a σ-algebra over X and the restriction $(\mu_X)|_{\Sigma_\Lambda}\colon \Sigma_\Lambda \to [0, 1]$ is a probability measure. Thus $X_\Lambda := (X, \Sigma_\Lambda, \mu|_{\Sigma_\Lambda})$ is a probability space.

(ii) The map $J_\Lambda\colon \Sigma(X_\Lambda) \to \Sigma(X)$, $A \mapsto A$ is a measure algebra homomorphism with range $J_\Lambda(\Sigma(X_\Lambda)) = \Lambda$.

(iii) For every $\gamma \in \Gamma$ we obtain a measure algebra automorphism

$$(T_\Lambda)_\gamma := J_\Lambda^{-1} \circ T_\gamma|_\Lambda \circ J_\Lambda\colon \Sigma(X_\Lambda) \to \Sigma(X_\Lambda).$$

This gives us a group representation $T_\Lambda\colon \Gamma \to \mathrm{Aut}(\Sigma(X_\Lambda))$, $\gamma \mapsto (T_\Lambda)_\gamma$, thus (X_Λ, T_Λ) is a measure-preserving system.

It is then clear by construction that the following holds.

Proposition 2.2.7. *Let* (X, T) *be a measure-preserving system and* $\Lambda \subseteq \Sigma(X)$ *an invariant sub-*σ*-algebra. Then* J_Λ *defines an extension* $J_\Lambda\colon (X_\Lambda, T_\Lambda) \to (X, T)$ *with* $J_\Lambda(\Sigma(X_\Lambda)) = \Lambda$.

Thus, up to isomorphism, subsystems of a given measure-preserving system (X, T) are in one-to-one correspondence with invariant sub-σ-algebras of the measure algebra $\Sigma(X)$. We use this observation in the following special case.

Example 2.2.8. Let (X, T) be any measure-preserving system (X, T). By Lemma 1.2.6, the set $\Lambda := \Sigma(X)_{\mathrm{inv}}$ from Definition 2.1.9 is an invariant sub-σ-algebra of $\Sigma(X)$, hence defines an extension $J_\Lambda\colon (X_\Lambda, T_\Lambda) \to (X, T)$ of measure-preserving systems. Note that by construction $(T_\Lambda)_\gamma = \mathrm{Id}_{\Sigma(X_\Lambda)}$ for every $\gamma \in \Gamma$, so the dynamics on X_Λ in this case are trivial. We write $J_{\mathrm{inv}} := J_\Lambda$ and $(X_{\mathrm{inv}}, \mathrm{Id}) := (X_\Lambda, T_\Lambda)$.

We also describe subsystems on a functional analytic level: If $J\colon (Y,S) \to (X,T)$ is an extension of measure-preserving systems, then the induced Koopman operator $U_J\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ from Proposition 1.3.4 satisfies $U_J U_{S_\gamma} = U_{T_\gamma} U_J$ for each $\gamma \in \Gamma$, cf. Remark 1.3.5. Combining this observation with Proposition 1.3.3, we immediately obtain the properties of the range of U_J.

Proposition and Definition 2.2.9. *Let $J\colon (Y,S) \to (X,T)$ be an extension of measure-preserving systems. Then $E = U_J(\mathrm{L}^2(Y))$ is an* ***invariant Markov sublattice*** *of $\mathrm{L}^2(X)$, i.e.,*

(i) *E is a closed linear subspace of $\mathrm{L}^2(X)$,*

(ii) *$\mathbb{1} \in E$,*

(iii) *$|f|$, $\mathrm{Re}(f)$, $\mathrm{Im}(f) \in E$ for every $f \in E$, and*

(iv) *$U_{T_\gamma} f \in E$ for every $f \in E$ and $\gamma \in \Gamma$.*

Remark 2.2.10. As a consequence of Theorem 1.3.7, a subsystem is again uniquely determined up to an isomorphism by the corresponding Markov sublattice. The details are discussed in Exercise 2.7 below.

We again go the converse direction and construct a subsystem from an invariant Markov sublattice. We use the following lemma.

Lemma 2.2.11. *Assume that (X,T) is a measure-preserving system and further that $E \subseteq \mathrm{L}^2(X)$ is an invariant Markov sublattice. Then*

$$\Lambda_E \coloneqq \{A \in \Sigma(X) \mid \mathbb{1}_A \in E\} \subseteq \Sigma(X)$$

is an invariant sub-σ-algebra.

Proof. We check the properties listed in Definition 2.2.5. To obtain property (i), note first that for $A, B \in \Lambda_E$ we also have $A \cup B \in \Lambda_E$ since

$$\mathbb{1}_{A\cup B} = \sup(\mathbb{1}_A, \mathbb{1}_B) = \frac{\mathbb{1}_A + \mathbb{1}_B + |\mathbb{1}_A - \mathbb{1}_B|}{2} \in E.$$

We thus obtain $A_1 \cup \dots \cup A_n \in \Lambda_E$ for all $A_1, \dots, A_n \in \Lambda_E$ and $n \in \mathbb{N}$ by induction. Now if $(A_n)_{n\in\mathbb{N}}$ is a sequence in Λ and $A \coloneqq \bigcup_{n\in\mathbb{N}} A_n \in \Sigma(X)$, then $\mathbb{1}_A = \lim_{n\to\infty} \mathbb{1}_{A_1\cup\dots\cup A_n}$ pointwise and hence in $\mathrm{L}^2(X)$ by Lebesgue's theorem. This shows $A \in \Lambda_E$, and hence establishes property (i) of Definition 2.2.5. For part (ii), note that if $A \in \Lambda_E$, then $\mathbb{1}_{X\setminus A} = \mathbb{1} - \mathbb{1}_A \in E$, hence $X \setminus A \in \Lambda_E$. For (iii) observe that $\mathbb{1}_\emptyset = 0 \in E$ (since E is a linear subspace) and $\mathbb{1}_X = \mathbb{1} \in E$, hence $\emptyset, X \in \Lambda_E$. Finally, (iv) of Definition 2.2.5 is obvious by part (iv) of Definition 2.2.9. □

We again consider the construction in a special case. For this we need the concept of the **fixed space** $\mathrm{fix}(V) \coloneqq \{f \in E \mid Vf = f\}$ of a linear map $V\colon E \to E$ on a complex vector space E.

Example 2.2.12. Let (X, T) be a measure-preserving system. Then, by the properties of Markov embeddings (see Definition 1.3.2 and Lemma 1.3.3), the **fixed space**

$$\mathrm{fix}(U_T) := \bigcap_{\gamma \in \Gamma} \mathrm{fix}(U_{T_\gamma}) = \{f \in \mathrm{L}^2(X) \mid U_{T_\gamma} f = f \text{ for every } \gamma \in \Gamma\}$$

is an invariant Markov sublattice of $\mathrm{L}^2(X)$. Since $U_{T_\gamma} \mathbb{1}_A = \mathbb{1}_{T_\gamma(A)}$ for $A \in \Sigma(X)$ and $\gamma \in \Gamma$, the corresponding sub-σ-algebra $\Lambda_{\mathrm{fix}(U_T)}$ is precisely $\Sigma(X)_{\mathrm{inv}}$ from Examples 2.1.9 and 2.2.8.

For a measure-preserving system (X, T) we now build subsystems from an invariant Markov sublattice $E \subseteq \mathrm{L}^2(X)$: First construct the invariant sub-σ-algebra $\Lambda_E := \{A \in \Sigma(X) \mid \mathbb{1}_A \in E\}$ of Lemma 2.2.11, and from this the extension $J_{\Lambda_E} \colon (X_{\Lambda_E}, T_{\Lambda_E}) \to (X, T)$, see Proposition 2.2.7. We abbreviate $(X_E, T_E) := (X_{\Lambda_E}, T_{\Lambda_E})$ and $J_E := J_{\Lambda_E}$, and show the following result.

Proposition 2.2.13. *Let (X, T) be a measure-preserving system and $E \subseteq \mathrm{L}^2(X)$ an invariant Markov sublattice. Then $J_E \colon (X_E, T_E) \to (X, T)$ is an extension with $U_{J_E}(\mathrm{L}^2(X_E)) = E$.*

Proof. The inclusion "$\subseteq$" follows from the facts that E is a closed linear subspace, U_{J_E} is a bounded linear map, and the subspace spanned by (equivalence classes of) characteristic functions is dense in $\mathrm{L}^2(X_E)$ by Lemma 1.3.6.

For the converse inclusion, we make a few observations.

(i) For $f, g \in E \cap \mathrm{L}^2(X, \mathbb{R})$ we have

$$\inf(f, g) = \frac{f + g - |f - g|}{2} \in E \quad \text{and} \quad \sup(f, g) = \frac{f + g + |f - g|}{2} \in E.$$

(ii) Recall that X_E is given by the set X, the σ-algebra $\Sigma_{\Lambda_E} = \{A \subseteq X \text{ measurable} \mid [A] \in \Lambda_E\}$ and the restriction $\mu_X|_{\Sigma_{\Lambda_E}} \colon \Sigma_{\Lambda_E} \to [0, 1]$. Moreover, the Koopman operator U_{J_E} induced by the measure algebra homomorphism $J_E \colon \Sigma(X_\Lambda) \to \Sigma(X)$, $A \mapsto A$ (see Proposition 1.3.4) is explicitly given as $U_{J_E} \colon \mathrm{L}^2(X_E) \to \mathrm{L}^2(X)$, $f \mapsto f$.

(iii) If $A \in \Sigma_{\Lambda_E}$, then $\int_X \mathbb{1}_A = \mu_X(A) = \int_{X_E} \mathbb{1}_A$. By the "standard measure-theoretic procedure" we also obtain that for a function $f \colon X \to [0, \infty)$ which is measurable with respect to Σ_{Λ_E} the integrals $\int_X f$ and $\int_{X_E} f$ agree. In particular, this implies that if $f \colon X \to [0, \infty)$ is Σ_{Λ_E}-measurable and square-integrable on the probability space $X = (X, \Sigma_X, \mu_X)$, then f is also square-integrable on $X_E = (X, \Sigma_{\Lambda_E}, \mu_X|_{\Sigma_{\Lambda_E}})$ (with the same L^2-norm).

Now pick $f \in E$ and show that $f \in U_{J_E}(\mathrm{L}^2(X_E))$. By decomposing into real and imaginary parts we may assume that $f \in \mathrm{L}^2(X, \mathbb{R})$. Take a identically named representative of f. By observations (ii) and (iii) above, it suffices to check that f is measurable with respect to Σ_{Λ_E}. So for $c \in \mathbb{R}$ we show that $[f > c]$ is an element of Λ_E, i.e., $\mathbb{1}_{[f>c]} \in E$. Replacing f by $f - c$ we may assume that $c = 0$. Now notice that for every $r \in \mathbb{R}$ we have

$$\lim_{n\to\infty} \min(n \cdot \max(r, 0), 1) = \begin{cases} 1 & \text{if } r > 0, \\ 0 & \text{if } r \leq 0. \end{cases}$$

This implies $\mathbb{1}_{\{f>0\}} = \lim_{n\to\infty} \inf(n \sup(f, 0), \mathbb{1})$ in $\mathrm{L}^2(X)$. By observation (i) above and the fact that E is closed in $\mathrm{L}^2(X)$, we therefore obtain $\mathbb{1}_{\{f>0\}} \in E$ as claimed. □

As a consequence, we obtain the following important functional analytic characterization of ergodic systems.

Corollary 2.2.14. *Let (X, T) be a measure-preserving system. Then the linear hull* $\mathrm{lin}\{\mathbb{1}_A \mid A \in \Sigma(X)_{\mathrm{inv}}\}$ *is dense in* $\mathrm{fix}(U_T)$. *In particular, the following assertions are equivalent.*

(a) *(X, T) is ergodic.*

(b) $\mathrm{fix}(U_T) = \{c\mathbb{1} \mid c \in \mathbb{C}\}$.

Proof. By Example 2.2.12 and Proposition 2.2.13 we obtain that the Markov embedding $U_{J_{\mathrm{inv}}} \colon \mathrm{L}^2(X_{\mathrm{inv}}) \to \mathrm{L}^2(X)$ has the range $U_{J_{\mathrm{inv}}}(\mathrm{L}^2(X_{\mathrm{inv}})) = \mathrm{fix}(U_T)$. Since $\mathrm{lin}\{\mathbb{1}_B \mid B \in \Sigma(X_{\mathrm{inv}})\}$ is dense in $\mathrm{L}^2(X_{\mathrm{inv}})$ by Lemma 1.3.6, and $U_{J_{\mathrm{inv}}}$ is a bounded linear operator, the subspace

$$\mathrm{lin}\{U_{J_{\mathrm{inv}}}\mathbb{1}_B \mid B \in \Sigma(X_{\mathrm{inv}})\} = \mathrm{lin}\{\mathbb{1}_{J_{\mathrm{inv}}(B)} \mid B \in \Sigma(X_{\mathrm{inv}})\} = \mathrm{lin}\{\mathbb{1}_A \mid A \in \Sigma(X)_{\mathrm{inv}}\}$$

is dense in $\mathrm{fix}(U_T)$. □

The last result of this lecture allows a way to build invariant Markov sublattices (and therefore subsystems of a measure-preserving system) from certain “invariant subalgebras”.

Proposition 2.2.15. *Let (X, T) be a measure-preserving system. Assume that $F \subseteq \mathrm{L}^\infty(X)$ is a linear subspace with*

(i) *$f \cdot g \in F$ for all $f, g \in F$,*

(ii) *$\overline{f} \in F$ for all $f \in F$,*

(iii) *$\mathbb{1} \in F$, and*

(iv) *$U_{T_\gamma} f \in F$ for all $f \in F$ and $\gamma \in \Gamma$.*

Then the closure $\overline{F}$ in $\mathrm{L}^2(X)$ is an invariant Markov sublattice.

For the proof we use the following identity. Recall that the binomial coefficients $\binom{\alpha}{k}$ for $\alpha \in \mathbb{C}$ are

$$\binom{\alpha}{k} := \frac{\alpha \cdot (\alpha - 1) \cdots (\alpha - k + 1)}{k!} \quad \text{for } k \in \mathbb{N}, \quad \text{and} \quad \binom{\alpha}{0} := 1.$$

Lemma 2.2.16. *Let X be a probability space and $f \in \mathrm{L}^\infty(X)$ with $\|f\|_\infty \leq 1$. Then*

$$|f| = \lim_{n\to\infty} \sum_{k=0}^{n} \binom{\frac{1}{2}}{k} (-1)^k (\mathbb{1} - |f|^2)^k$$

in $\mathrm{L}^\infty(X)$.

Proof. For $\alpha \in (0, \infty)$ series $\sum_{k=0}^{\infty} \binom{\alpha}{k} z^k$ converges absolutely and uniformly to $(1+z)^\alpha$ for $z \in \mathbb{C}$ with $|z| \leq 1$ (see, e.g., [AE05, Theorem V.3.10]). Observing that $|z| = (1 - (1 - |z|^2))^{\frac{1}{2}}$ for $z \in \mathbb{C}$, we therefore find for $\varepsilon > 0$ some $n_0 \in \mathbb{N}$ with

$$\left| |z| - \sum_{k=0}^{n} \binom{\frac{1}{2}}{k} (-1)^k (1 - |z|^2)^k \right| \leq \varepsilon$$

all $z \in \mathbb{C}$ with $|z| \leq 1$ and for every $n \geq n_0$. This implies

$$\left| |f| - \sum_{k=0}^{n} \binom{\frac{1}{2}}{k} (-1)^k (\mathbb{1} - |f|^2)^k \right| \leq \varepsilon \mathbb{1}$$

almost everywhere for $n \geq n_0$ which yields the claim. □

Proof of Proposition 2.2.15. We check properties (i) - (iv) of Definition 2.2.9. As the closure of a linear subspace is again a linear subspace, we obtain (i). Property (ii) is clear by our assumption (iv) of Proposition 2.2.15. To obtain Definition 2.2.9 (iii), take $f \in F$ with $\|f\|_\infty \leq 1$, then $|f|^2 = f \cdot \overline{f} \in F$ by properties (i) and (ii). Since L^∞-convergence implies L^2-convergence, Lemma 2.2.16 shows $|f| \in \overline{F}$. If $f \in \overline{F}$, we find a sequence $(f_n)_{n\in\mathbb{N}}$ converging in L^2 to f, and this implies $|f| = \lim_{n\to\infty} |f_n| \in \overline{F}$. Finally, part (iv) of Definition 2.2.9 is an easy consequence of Proposition 2.2.15 (iv) and continuity of $U_{T_\gamma} \colon \mathrm{L}^2(X) \to \mathrm{L}^2(X)$ for $\gamma \in \Gamma$. □

To summarize this section, we can define subsystems of a measure-preserving system (X, T) in three different ways:

(i) via extensions $J \colon (Y, S) \to (X, T)$,

(ii) via invariant sub-σ-algebras $\Lambda \subseteq \Sigma(X)$, and

(iii) via invariant Markov sublattices $E \subseteq \mathrm{L}^2(X)$.

2.3 Comments and Further Reading

There is a plethora of textbooks on ergodic theory focusing on various different aspects. A few classical works related to the topics of this course are [Hal56], [Wal75], and [Fur14]. We also mention [Gla03], [EW11] and [VO16] for contemporary introductions, and [DNP87] and [EFHN15] for an operator theoretic approach to ergodic theory. A large part of the material of this course is based on the contents of these monographs and two previous editions of the Internet Seminar ([EFHN09] and [EF19]).

We remark that there are numerous further equivalent descriptions of subsystems of a given measure-preserving systems, see [EFHN15, Theorem 13.20]. Studying extensions of measure-preserving systems will play a crucial part later in this course.

2.4 Exercises

Exercise 2.1. For an index set $I \neq \emptyset$ let $\tau_i \colon X_i \to Y_i$ be measure-preserving maps between probability spaces for every $i \in I$. Show that the map

$$\prod_{i\in I} \tau_i \colon \prod_{i\in I} X_i \to \prod_{i\in I} Y_i, \quad (x_i)_{i\in I} \mapsto (\tau_i(x_i))_{i\in I}$$

between the product measure spaces is also measure-preserving.

Exercise 2.2. With the notation from Exercise 1.6 show that for a measure-preserving system (X, T) the following assertions are equivalent.

(a) (X, T) is ergodic.

(b) $X = \sup\{T_\gamma(A) \mid \gamma \in \Gamma\}$ in $\Sigma(X)$ for every $A \in \Sigma(X) \setminus \{\emptyset\}$.

Exercise 2.3. Show that for a measure-preserving map $\sigma \colon X \to X$ on a probability space X the following assertions are equivalent.

(a) $\sigma^*(A) = A$ for $A \in \Sigma(X)$ implies $\mu_X(A) \in \{0, 1\}$.

(b) $\sigma^{-1}(A) = A$ for $A \subseteq X$ measurable implies $\mu_X(A) \in \{0, 1\}$.

Hint: For a representative of A as in (a) consider $\bigcap_{n=0}^\infty \bigcup_{k=n}^\infty \sigma^{-k}(A)$.

Exercise 2.4. Assume that Γ is a finite abelian group and consider the Bernoulli shift (X^Γ, τ) from Example 2.1.8. Show that the system (X^Γ, τ^*) is ergodic precisely when $\Sigma(X) = \{\emptyset, X\}$.

Exercise 2.5. For $\Gamma = \mathbb{Z}$ consider the Bernoulli shift $(\{0, 1\}^\mathbb{Z}, \tau)$ where the measure on $\{0, 1\}$ is given by the probability vector $(\frac{1}{2}, \frac{1}{2})$, see Examples 1.1.3 (ii) and 2.1.8. Let further $([0, 1]^2, \sigma)$ be the measure-preserving system defined by the baker's transformation from Exercise 1.2 (cf. Remark 2.1.6). Then the map

$$q \colon \{0, 1\}^\mathbb{Z} \to [0, 1]^2, \quad (x_n)_{n\in\mathbb{Z}} \mapsto \left(\sum_{n=1}^\infty \frac{x_{n-1}}{2^n}, \sum_{n=1}^\infty \frac{x_{-n}}{2^n}\right)$$

defines an isomorphism between $(\{0, 1\}^\mathbb{Z}, \tau)$ and $([0, 1]^2, \sigma)$.

Hint: Recall that every $x \in [0, 1)$ has a base 2 expansion $x = \sum_{n=1}^\infty x_n 2^{-n}$ for some sequence $(x_n)_{n\in\mathbb{N}} \in \{0, 1\}^\mathbb{N}$. Moreover, this representation becomes unique if we exclude sequences $(x_n)_{n\in\mathbb{N}}$ such that there is $N \in \mathbb{N}$ with $x_n = 1$ for all $n \geq N$, see, e.g., [AE05, Topic 6 of Section II.7]).

Exercise 2.6. Show the following assertions.

(i) For every measure-preserving system (X, T) on a Lebesgue space X there is a unique concrete measure-preserving system (X, τ) with $(X, T) = (X, \tau^*)$.

(ii) If (X, τ) and (Y, σ) are concrete measure-preserving systems on Lebesgue spaces X and Y and $J\colon (Y, \sigma^*) \to (X, \tau^*)$ is an extension of measure-preserving systems, then there is a unique factor map $q\colon (X, \tau) \to (Y, \sigma)$ with $J = q^*$. Moreover, q is an isomorphism precisely when J is an isomorphism.

Exercise 2.7. Show the following assertions.

(i) Let X be a probability space and $U\colon \Gamma \to \mathrm{Aut}(\mathrm{L}^2(X))$ a representation of Γ as Markov automorphisms. Show that there is a unique measure-preserving system (X, T) with $U = U_T$.

(ii) Let (Y, S) and (X, T) be measure-preserving systems and $V\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ a Markov embedding with $VU_{S_\gamma} = U_{T_\gamma}V$ for all $\gamma \in \Gamma$. Then there is a unique extension $J\colon (Y, S) \to (X, T)$ with $V = U_J$. Moreover, J is an isomorphism precisely when V is bijective.

(iii) Let $J_1\colon (Y_1, S_1) \to (X, T)$ and $J_2\colon (Y_2, S_2) \to (X, T)$ be extensions of measure-preserving systems such that $U_{J_1}(\mathrm{L}^2(Y_1)) = U_{J_2}(\mathrm{L}^2(Y_2))$. Then there is a unique isomorphism $I\colon (Y_1, S_1) \to (Y_2, S_2)$ of measure-preserving systems with $J_2 = J_1 \circ I$.

Exercise 2.8. For probability spaces X and Y let $U\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ be a linear isometry with the following properties.

(i) $U(\mathrm{L}^\infty(Y)) \subseteq \mathrm{L}^\infty(X)$,

(ii) $U(f \cdot g) = Uf \cdot Ug$ for all $f, g \in \mathrm{L}^\infty(Y)$,

(iii) $\overline{Uf} = U\overline{f}$ for every $f \in \mathrm{L}^\infty(Y)$,

(iv) $U\mathbb{1} = \mathbb{1}$.

Show that U is a Markov embedding.

Lecture 3

In this lecture we prove a general version of von Neumann's famous mean ergodic theorem, an assertion about the long-term behavior of measure-preserving transformations. As an application we show Khintchin's refinement of Poincaré's recurrence theorem from the first lecture. In the second part of the lecture we then explore the connection between ergodic theory and topological dynamics.

3.1 Averaging and the Mean Ergodic Theorem

Mathematical ergodic theory originated in statistical mechanics. A measure-preserving map $\tau\colon X \to X$ can model how a state x of a physical dynamical system (describing, e.g., the position and impulse of all particles of a gas or fluid) is transformed, by the principles of physics, into a new state $\tau(x)$. From this perspective, a function $f\colon X \to \mathbb{C}$ can be seen as an "observable" giving some measurable data $f(x)$ about any given state x of the system. By making measurements "after each time step", we obtain a sequence of observables $f, f \circ \tau, f \circ \tau^2, f \circ \tau^3, \dots$ Von Neumann's famous mean ergodic theorem from 1931 provides information on the asymptotic behavior of this sequence.

Theorem 3.1.1 (von Neumann)**.** *Let $\tau\colon X \to X$ be a measure-preserving map. Then $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} U_\tau^n f$ exists in $\mathrm{L}^2(X)$ for every $f \in \mathrm{L}^2(X)$ and is an element of $\mathrm{fix}(U_\tau)$.*

We deduce Theorem 3.1.1 from a general operator theoretic result. The necessary notation, concepts and tools from basic Hilbert space theory are collected in Appendix A.2. We start from the following definition.

Definition 3.1.2. Let H be a Hilbert space. A family $\mathscr{S} \subseteq \mathscr{L}(H)$ is a **semigroup (of operators)** if $UV \in \mathscr{S}$ for all $U, V \in \mathscr{S}$. It is a **contraction semigroup** if, in addition, $\|U\| \leq 1$ for all $U \in \mathscr{S}$.

Generalizing the notion of the fixed space from Example 2.2.12, we introduce the following definition.

Definition 3.1.3. For a semigroup $\mathscr{S} \subseteq \mathscr{L}(H)$ on a Hilbert space H we call

$$\operatorname{fix}(\mathscr{S}) := \bigcap_{U\in\mathscr{S}} \operatorname{fix}(U) = \{f \in H \mid Uf = f \text{ for every } U \in \mathscr{S}\}$$

the **fixed space** of $\mathscr{S}$.

Remark 3.1.4. For every bounded linear operator $U \in \mathscr{L}(H)$ on a Hilbert space H we obtain a semigroup $\mathscr{S}_U := \{U^n \mid n \in \mathbb{N}_0\}$. Since the operator norm is submultiplicative, $\mathscr{S}_U$ is a contraction semigroup precisely when $\|U\| \leq 1$. Note further that $\operatorname{fix}(\mathscr{S}_U) = \operatorname{fix}(U)$.

As recalled in the Appendix (see Theorem A.2.2), every closed linear subspace $E \subseteq H$ of a Hilbert space can be orthogonally complemented to obtain a decomposition $H = E \oplus E^{\perp}$. For a contraction semigroup $\mathscr{S} \subseteq \mathscr{L}(H)$ on a Hilbert space H there is a nice description of the orthogonal complement of the fixed space $\operatorname{fix}(\mathscr{S})$ in terms of the ranges of the operators $\mathrm{Id}_H - U$ for $U \in \mathscr{S}$. Here and in the following, for a subset $A \subseteq H$ of a Hilbert space H,

(i) the **closed convex hull** $\overline{\mathrm{co}}\, A$ is the closure of the set of all convex combinations of elements of A, and

(ii) the **closed linear hull** $\overline{\mathrm{lin}}\, A$ is the closure of the set of all linear combinations of elements of A.

Theorem 3.1.5 (Abstract Mean Ergodic Theorem)**.** *For every contraction semigroup $\mathscr{S} \subseteq \mathscr{L}(H)$ on a Hilbert space H we have an orthogonal decomposition*

$$H = \operatorname{fix}(\mathscr{S}) \oplus \overline{\mathrm{lin}} \bigcup_{U\in\mathscr{S}} (\mathrm{Id}_H - U)(H).$$

Moreover, the orthogonal projection P onto $\operatorname{fix}(\mathscr{S})$ *has the following properties.*

(i) *$PU = UP = P$ for every $U \in \mathscr{S}$.*

(ii) *For every $f \in H$ the vector $Pf \in H$ is the unique element of $\overline{\mathrm{co}}\,\{Uf \mid U \in \mathscr{S}\}$ of minimal norm.*

The proof uses the following elementary observation on the adjoint operator of a contraction.

Lemma 3.1.6. *Let $U \in \mathscr{L}(H)$ with $\|U\| \leq 1$. Then* $\operatorname{fix}(U) = \operatorname{fix}(U^*)$.

Proof. Recall that the adjoint operator $U^* \in \mathscr{L}(H)$ satisfies $\|U^*\| = \|U\| \leq 1$. We first show the inclusion "$\subseteq$". By the "Pythagorean Theorem" we obtain for $f \in \operatorname{fix}(U)$

that

$$\begin{aligned}0 &\le \|U^*f - f\|^2 = \|U^*f\|^2 - 2\mathrm{Re}\,(U^*f|f) + \|f\|^2 \\ &= \|U^*f\|^2 - 2\mathrm{Re}\,(f|Uf) + \|f\|^2 = \|U^*f\|^2 - 2\|f\|^2 + \|f\|^2 = \|U^*f\|^2 - \|f\|^2 \\ &\le \|f\|^2 - \|f\|^2 = 0.\end{aligned}$$

This yields $U^*f = f$. Thus, $\mathrm{fix}(U) \subseteq \mathrm{fix}(U^*)$. Since $(U^*)^* = U$, the same argument applied to U^* shows $\mathrm{fix}(U^*) \subseteq \mathrm{fix}((U^*)^*) = \mathrm{fix}(U)$. □

Proof of Theorem 3.1.5. We prove (i) and (ii), before we show the claimed decomposition of H. For part (i), pick $U \in \mathscr{S}$. Since $PH = \mathrm{fix}(\mathscr{S})$, it is clear that $UP = P$ holds. By Lemma 3.1.6 we have $\mathrm{fix}(U) = \mathrm{fix}(U^*)$, and therefore also $U^*P = P$. Taking adjoints and using that $P^* = P$, we thus obtain

$$PU = ((PU)^*)^* = (U^*P^*)^* = (U^*P)^* = P^* = P.$$

For part (ii), take $f \in H$ and set $C := \overline{\mathrm{co}}\,\{Uf \mid U \in \mathscr{S}\}$. As a special case of Theorem A.2.1, every non-empty closed convex set of a Hilbert space contains a unique element of minimal norm. Let g be the unique element of minimal norm in C. We show that $g = Pf$.

For $U \in \mathscr{S}$ we have $UC \subseteq C$ since U is continuous and linear, and $\mathscr{S}$ is a semigroup. In particular, $Ug \in C$. Since $\|Ug\| \le \|g\|$, we must have $Ug = g$ by choice of g. This means $g \in \mathrm{fix}(\mathscr{S})$, and therefore $g = Pg \in PC$. But, using that P is continuous and linear, we have

$$PC \subseteq \overline{\mathrm{co}}\,\{PUf \mid U \in \mathscr{S}\} = \{Pf\}$$

by part (i). Hence, $Pf = g$ as claimed, and we have established (ii).

We finally prove the claimed orthogonal decomposition by showing that $\mathrm{fix}(\mathscr{S})^\perp = \overline{\mathrm{lin}}\bigcup_{U\in\mathscr{S}}(\mathrm{Id}_H - U)(H)$.

For the inclusion "$\supseteq$" observe that part (i) implies $P(\mathrm{Id}_H - U) = 0$ for every $U \in \mathscr{S}$. Since $P^{-1}(\{0\}) = \mathrm{fix}(\mathscr{S})^\perp$ is a closed linear subspace of H, we thus obtain $\overline{\mathrm{lin}}\bigcup_{U\in\mathscr{S}}(\mathrm{Id}_H - U)(H) \subseteq \mathrm{fix}(\mathscr{S})^\perp$.

For the converse inclusion "$\subseteq$" pick $f \in P^{-1}(\{0\})$. For every $\varepsilon > 0$ we find by (ii) elements $U_1, \dots, U_m \in \mathscr{S}$ and $c_1, \dots, c_m \in [0,1]$ with $\sum_{j=1}^m c_j = 1$ such that $\|\sum_{i=1}^m c_i U_i f\| \le \varepsilon$. But this means

$$\left\| f - \sum_{i=1}^m c_i(\mathrm{Id}_H - U_i)f \right\| \le \varepsilon.$$

Therefore f is in the closure of the subspace $\mathrm{lin}\bigcup_{U\in\mathscr{S}}(\mathrm{Id}_H - U)(H)$. □

The abstract (and general) Theorem 3.1.5 implies the convergence of many different "ergodic averages". To make this precise, we introduce the notion of a net, generalizing the classical concept of a sequence, see also [Sin19, Section 4.2] for more details.

A set $I \neq \emptyset$ with a relation $\leq$ is a **directed set** if

(i) $i \leq i$ for every $i \in I$,

(ii) $i_1 \leq i_2$ and $i_2 \leq i_3$ for $i_1, i_2, i_3 \in I$ implies $i_1 \leq i_3$, and

(iii) for all $i_1, i_2 \in I$ there is $i \in I$ with $i_1 \leq i$ and $i_2 \leq i$.

Basic examples are $\mathbb{N}$, $\mathbb{Z}$, or $\mathbb{R}$ with their natural ordering, or the power set $\mathcal{P}(X)$ of a set X equipped with set inclusion "$\subseteq$" as its relation.

A map $I \to \Omega$, $i \mapsto \omega_i$ from a directed set I into a set Ω is usually written as $(\omega_i)_{i\in I}$ and is called a **net** in Ω. If Ω is a topological space, $(\omega_i)_{i\in I}$ is a net in Ω, and $\omega \in \Omega$ is some point, then we say, just as for sequences, that

(i) $(\omega_i)_{i\in I}$ **converges to** ω if for every open set O with $\omega \in O$ there is some $i_0 \in I$ with $x_i \in O$ for all $i \geq i_0$. In this case, ω is called a **limit** of $(\omega_i)_{i\in I}$.

(ii) ω is an **accumulation point of** $(\omega_i)_{i\in I}$ if for every open set O with $\omega \in O$ and each $i_0 \in I$ there is an $i \geq i_0$ with $\omega_i \in O$.

In a Hausdorff space[1] Ω, a net $(\omega_i)_{i\in I}$ can have at most one limit ω (see [Sin19, Theorem 4.2.4], and we write $\lim_{i\in I} \omega_i := \omega$ in this case.

Many well-known characterizations of topological notions in metric spaces transfer over to topological spaces when one replaces the word "sequence" with "net". For example, a Hausdorff space Ω is compact[2] precisely when every net in Ω has an accumulation point (see, e.g., [Sin19, Theorems 4.2.10 and 5.1.17]).

The following definition uses the concept of nets in the context of the mean ergodic theorem.

Definition 3.1.7. Let $\mathscr{S} \subseteq \mathscr{L}(H)$ be a semigroup on a Hilbert space H. A net $(V_i)_{i\in I}$ in $\mathscr{L}(H)$ is an **ergodic net** for $\mathscr{S}$ if for every $f \in H$ we have

(i) $V_i f \in \overline{\text{co}}\,\{Uf \mid U \in \mathscr{S}\}$ for each $i \in I$, and

(ii) $\lim_{i\in I} V_i(f - Uf) = 0$ for every $U \in \mathscr{S}$.

We discuss some basic examples.

Examples 3.1.8. Let $U \in \mathscr{L}(H)$ with $\|U\| \leq 1$ and consider the induced contraction semigroup $\mathscr{S}_U = \{U^n \mid n \in \mathbb{N}_0\}$ of Remark 3.1.4.

---

[1]Recall that a topological space Ω is a **Hausdorff space** if for distinct $x, y \in \Omega$ we can always find open, disjoint subsets $O_x, O_y \subseteq \Omega$ with $x \in O_x$ and $y \in O_y$.

[2]In this course, the term *compact* will always include the Hausdorff property.

(i) The classical **Cesàro means** $V_N := \frac{1}{N}\sum_{n=0}^{N-1} U^n$ for $N \in \mathbb{N}$ define an ergodic net $(V_N)_{N\in\mathbb{N}}$ for $\mathscr{S}_U$: For $f \in H$ we obtain by telescopic summing,

$$\|V_N(f - Uf)\| = \left\| \frac{1}{N}\sum_{n=0}^{N-1} U^n f - \frac{1}{N}\sum_{n=1}^{N} U^n f \right\| = \frac{1}{N}\|f - U^N f\| \le \frac{2\|f\|}{N}$$

for every $N \in \mathbb{N}$, hence $\lim_{N\to\infty} \|V_N(f - Uf)\| = 0$. Again using telescopic summing this implies $\lim_{N\to\infty} \|V_N(f - U^n f)\| = 0$ for every $n \in \mathbb{N}$.

(ii) Turn $\mathbb{N}^2$ into a directed set by setting $(N_1, M_1) \le (N_2, M_2)$ if $M_1 \le M_2$ for $(N_1, M_1), (N_2, M_2) \in \mathbb{N}^2$. Similarly to (i), the means $V_{N,M} := \frac{1}{M}\sum_{n=N}^{N+M-1} U^n$ for $(N, M) \in \mathbb{N}^2$ define an ergodic net $(V_{N,M})_{(N,M)\in\mathbb{N}^2}$ for $\mathscr{S}_U$.

(iii) If we reverse the order on $(1, \infty)$, we obtain that the **Abel means** $V_r := (1 - 1/r)\sum_{n=0}^{\infty} r^{-n} U^n$ for $r \in (1, \infty)$ define an ergodic net $(V_r)_{r\in(1,\infty)}$ for $\mathscr{S}_U$, see Exercise 3.3.

A general construction of ergodic nets involves so-called *Følner nets* of groups. Here, for a subset $A \subseteq G$ of a group G and $x \in G$ we use the intuitive notation $Ax := \{yx \mid y \in A\}$. Analogously, we introduce xA and A^{-1} (which will appear at a later point).

Definition 3.1.9. For a group G call a net $(F_i)_{i\in I}$ of non-empty finite subsets of G a **(right) Følner net** if $\lim_{i\in I} \frac{|F_i \,\Delta\, F_i x|}{|F_i|} = 0$ for every $x \in G$.

Simple examples for the group $G = \mathbb{Z}$ are the sequences $(E_N)_{N\in\mathbb{N}}$ and $(F_N)_{N\in\mathbb{N}}$ with $E_N := \{0, \ldots, N-1\}$ and $F_N := \{-N+1, \ldots, N-1\}$ for $N \in \mathbb{N}$. We obtain the following general result.

Proposition 3.1.10. *Every abelian group G has a Følner net.*

We first prove a helpful lemma.

Lemma 3.1.11. *Let $E \subseteq G$ be a non-empty finite subset of an abelian group G and consider the sets $E^n := \{x_1 \cdots x_n \mid x_1, \ldots, x_n \in E\}$ for $n \in \mathbb{N}$. For every $r > 1$ there is some $n \in \mathbb{N}$ with $|E^{n+1}| \le r|E^n|$.*

Proof. We write $N := |E|$ for the number of elements of E. For $n \in \mathbb{N}$ there are

$$p(n) := \binom{n+N-1}{n} = \frac{(n+N-1)\cdots(n+1)}{(N-1)!}$$

many possibilities to chose n elements out of N distinct elements with replacement (see, e.g., [Bó06, Theorem 3.21]). This implies that the set E^n has at most $p(n)$ many elements for every $n \in \mathbb{N}$ (here we use that G is abelian!). Assuming that there is some $r > 1$ such that $|E^{n+1}| > r|E^n|$ for all $n \in \mathbb{N}$, we obtain that

$$r^{n-1}|E| < |E^n| \le p(n) \text{ for every } n \in \mathbb{N}.$$

But this contradicts the fact that power functions with base greater than 1 grow faster than any polynomial. □

Proof of Proposition 3.1.10. Let I be the set of pairs (E,m) of non-empty finite subsets $E \subseteq G$ and elements $m \in \mathbb{N}$. By setting

$$(E_1, m_1) \leq (E_2, m_2) \quad :\Leftrightarrow \quad E_1 \subseteq E_2 \text{ and } m_1 \leq m_2$$

for $(E_1, m_1), (E_2, m_2) \in I$, we turn I into a directed set. For every $(E,m) \in I$ we use Lemma 3.1.11 to find $n(E,m) \in \mathbb{N}$ with $|E^{n(E,m)+1}| \leq (1+\frac{1}{m})|E^{n(E,m)}|$ and set $F_{E,m} := E^{n(E,m)}$. We claim that $(F_{E,m})_{(E,m)\in I}$ is then a Følner net.

So let $x \in G$ and $n \in \mathbb{N}$. We take $(E,m) \in I$ with $(E,m) \geq (\{x, x^{-1}, 1\}, n)$ and show that $|F_{E,m} \,\Delta\, F_{E,m}x|/|F_{E,m}| \leq \frac{2}{n}$. Since $1, x \in E$, we have $F_{E,m}x = E^{n(E,m)}x \subseteq E^{n(E,m)+1}$ and $E^{n(E,m)} \subseteq E^{n(E,m)+1}$. This implies

$$\frac{|F_{E,m}x \setminus F_{E,m}|}{|F_{E,m}|} \leq \frac{|E^{n(E,m)+1} \setminus E^{n(E,m)}|}{|E^{n(E,m)}|} = \frac{|E^{n(E,m)+1}| - |E^{n(E,m)}|}{|E^{n(E,m)}|} \leq \frac{1}{m} \leq \frac{1}{n}.$$

But, since multiplication from the right with x^{-1} is a bijection, we also have

$$\frac{|F_{E,m} \setminus F_{E,m}x|}{|F_{E,m}|} = \frac{|(F_{E,m} \setminus F_{E,m}x)x^{-1}|}{|F_{E,m}|} = \frac{|(F_{E,m}x^{-1}) \setminus F_{E,m}|}{|F_{E,m}|} \leq \frac{1}{n}$$

by the same reasoning and the fact that $x^{-1} \in E$. Combining both estimates we obtain the claim. □

We now construct ergodic nets for certain group representations.

Definition 3.1.12. For a Hilbert space H and a group G we call a group homomorphism $U\colon G \to \mathscr{U}(H)$, $x \mapsto U_x$ to the group $\mathscr{U}(H)$ of unitary operators a **unitary representation** of G on H.

We are mostly interested in the following class of examples.

Example 3.1.13. Let (X,T) be a measure-preserving system. Then the Koopman representation $U_T\colon \Gamma \to \mathscr{U}(\mathrm{L}^2(X))$ is a unitary representation of the group Γ on the Hilbert space $\mathrm{L}^2(X)$.

Proposition 3.1.14. *Let $U\colon G \to \mathscr{U}(H)$ be a unitary representation of a group G. If $(F_i)_{i\in I}$ is a Følner net for G, then by setting $V_i := \frac{1}{|F_i|}\sum_{x\in F_i} U_x$ for $i \in I$ we obtain an ergodic net $(V_i)_{i\in I}$ for the image group $\{U_x \mid x \in G\}$.*

Proof. For $y \in G$ and $f \in H$ we have

$$\Big\|V_i(f - U_y f)\Big\| = \frac{1}{|F_i|}\Big\|\sum_{x\in F_i} U_x f - \sum_{x\in F_i y} U_x f\Big\| = \frac{1}{|F_i|}\Big\|\sum_{x\in F_i\setminus F_i y} U_x f - \sum_{x\in F_i y\setminus F_i} U_x f\Big\|$$
$$\leq \frac{1}{|F_i|}\sum_{x\in F_i\,\Delta\, F_i y} \|U_x f\| \leq \frac{|F_i\,\Delta\, F_i y|}{|F_i|}\cdot \|f\|$$

for all $i \in I$ which implies the claim. □

Thus, by Proposition 3.1.10, ergodic nets always exist for (the image of) a unitary representation of an abelian group. In Exercise 3.5 we even construct ergodic nets for any abelian contraction semigroup $\mathscr{S} \subseteq \mathscr{L}(H)$ on a Hilbert space H.

We now obtain the following consequence of Theorem 3.1.5.

Theorem 3.1.15. *Let $\mathscr{S} \subseteq \mathscr{L}(H)$ be a contraction semigroup on a Hilbert space H, and $P \in \mathscr{L}(H)$ the orthogonal projection onto* $\mathrm{fix}(\mathscr{S})$*. If $(V_i)_{i\in I}$ is an ergodic net for $\mathscr{S}$, then $\lim_{i\in I} V_i f = Pf$ for every $f \in H$.*

In the proof we use the following standard and very useful lemma of operator theory (cf. [Haa14, Exercise 9.10]). It is an easy exercise when using linearity and applying the triangle inequality.

Lemma 3.1.16. *Let E and F be normed spaces, and further $D \subseteq E$ such that the linear hull* $\mathrm{lin}\, D$ *is dense in E. For $(U_i)_{i\in I}$ a net in $\mathscr{L}(E, F)$ with $\sup_{i\in I} \|U_i\| < \infty$ and $U \in \mathscr{L}(E, F)$ the following assertions are equivalent.*

(i) *$(U_i f)_{i\in I}$ converges to Uf for every $f \in D$.*

(ii) *$(U_i f)_{i\in I}$ converges to Uf for every $f \in E$.*

Proof of Theorem 3.1.15. Take $f \in H$. Using the orthogonal decomposition from 3.1.5 we can reduce to the cases $f \in \mathrm{fix}(\mathscr{S})$ and $f \in \overline{\mathrm{lin}} \bigcup_{U\in\mathscr{S}} (\mathrm{Id}_H - U)(H)$.

In the first situation we have $Uf = f$ for every $U \in \mathscr{S}$. This implies $\overline{\mathrm{co}}\,\{Uf \mid U \in \mathscr{S}\} = \{f\}$, and therefore also $V_i f = f$ for every $i \in I$ by Definition 3.1.7 (i). This implies $\lim_{i\in I} V_i f = f$.

For the second case, observe that for $g \in H$, we obtain by Definition 3.1.7 (ii) that $\lim_{i\in I} V_i(\mathrm{Id}_H - U)g = 0$. By Lemma 3.1.16 we obtain that $\lim_{i\in I} V_i f = 0$ for every $f \in \overline{\mathrm{lin}} \bigcup_{U\in\mathscr{S}} (\mathrm{Id}_H - U)(H)$. □

In particular, by taking the Cesàro means (see Example 3.1.8 (ii)) of a Koopman operator, we recover von Neumann's Theorem 3.1.1. The following refinement of the Poincaré's recurrence theorem, Theorem 1.1.6, is an application.

Corollary 3.1.17 (Khintchin's refinement)**.** *Let $T\colon \Sigma(X) \to \Sigma(X)$ be a measure algebra homomorphism on a probability space X and $A \in \Sigma(X)$. Then*

$$\mu_X(A)^2 \leq \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu_X(A \cap T^n(A)).$$

In particular, if $\mu_X(A) > 0$, there are infinitely many $n \in \mathbb{N}$ with $\mu_X(A \cap T^n(A)) > 0$.

Proof. Write P for the orthogonal projection onto $\operatorname{fix}(U_T) = \operatorname{fix}(\mathscr{S}_{U_T})$ (cf. Remark 3.1.4). Since $\int_X U_T f = \int_X f$ for every $f \in \mathrm{L}^2(X)$ (see Lemma 1.3.3), we obtain

$$\mu_X(A) = \int_X \mathbb{1}_A = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X U_T^n \mathbb{1}_A = \int_X P\mathbb{1}_A = (P\mathbb{1}_A|\mathbb{1}).$$

By the Cauchy-Schwarz inequality we thus have

$$\begin{aligned}\mu_X(A)^2 &\leq \|P\mathbb{1}_A\|_2^2 \cdot \|\mathbb{1}\|_2^2 = (P\mathbb{1}_A|P\mathbb{1}_A) = (P^*P\mathbb{1}_A|\mathbb{1}_A) \\ &= (P\mathbb{1}_A|\mathbb{1}_A) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X U_T^n \mathbb{1}_A \cdot \mathbb{1}_A = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu_X(A \cap T^n(A)).\end{aligned}$$

□

3.2 Invariant Measures

In many examples the underlying maps of a concrete measure-preserving system are not only measurable, but even continuous. Ergodic theory is therefore closely related to the area of *topological dynamics*. We introduce its central objects. Here, for a compact space K, write $\operatorname{Homeo}(K)$ for the group of all homeomorphisms $\tau\colon K \to K$.

Definition 3.2.1. A **topological dynamical system** (K, τ) consists of a compact space $K \neq \emptyset$ and a group homomorphism $\tau\colon \Gamma \to \operatorname{Homeo}(K)$, $\gamma \mapsto \tau_\gamma$.

As an important example we discuss a topological version of the Bernoulli shift. For this, recall the product of topological spaces (see, e.g., [Sin19, Section 2.2]).

Definition 3.2.2. Let I be an index set and Ω_i a topological space for every $i \in I$. For a finite set $\{i_1, \dots, i_m\} \subseteq I$ and open sets $O_j \subseteq \Omega_{i_j}$ for each $j \in \{1, \dots, m\}$ we call

$$\mathrm{Z}(i_1, \dots, i_m; O_1, \dots, O_m) \coloneqq \left\{ (\omega_i)_{i\in I} \in \prod_{i\in I} \Omega_i \mid \omega_{i_j} \in O_j \text{ for all } j \in \{1, \dots, m\} \right\}$$

an **(open) cylinder set**. The topology on $\prod_{i\in I} \Omega_i$ generated by all such cylinder sets is the **product topology**.

Remark 3.2.3. The product topology is the smallest topology on the product $\prod_{i\in I}\Omega_i$ making all projection maps $\mathrm{pr}_j\colon \prod_{i\in I}\Omega_i \to \Omega_j$, $(\omega_i)_{i\in I}\mapsto \omega_j$ for $j\in I$ continuous (see [Sin19, Proposition 2.2.7]) A different way to think about this topology is via convergence of nets (and, in particular, sequences): A net converges in the product topology to some limit precisely when it converges in each component to that limit (see [Sin19, Theorem 4.2.7]).

From now on, we equip the product of topological spaces with the product topology. If we start with compact spaces, then we still obtain a compact space it that way (see, e.g., [Sin19, Theorem 5.1.14]):

Theorem 3.2.4 (Tychonoff). *Let I be an index set and K_i be a compact space for every $i\in I$. Then the product space $\prod_{i\in I}K_i$ is also compact.*

With the help of Tychonoff's Theorem we now obtain a topological version of Example 2.1.8.

Example 3.2.5. Let $K\neq\emptyset$ be a compact space, e.g., $K=\{0,\dots,k-1\}$ equipped with the discrete topology (i.e., every set is open). Then $\tau\colon\Gamma\to\mathrm{Homeo}(K^\Gamma)$, $\gamma\mapsto\tau_\gamma$ with $\tau_\gamma(x_\delta)_{\delta\in\Gamma}=(x_{\delta+\gamma})_{\delta\in\Gamma}$ for $(x_\delta)_{\delta\in\Gamma}\in K^\Gamma$ and $\gamma\in\Gamma$ defines a topological dynamical system (K^Γ,τ). In fact, with arguments similar as in Example 2.1.8, we obtain that $\tau_\gamma\colon K^\Gamma\to K^\Gamma$ is a homeomorphism for every $\gamma\in\Gamma$, and obviously $\tau_{\gamma_1+\gamma_2}=\tau_{\gamma_1}\circ\tau_{\gamma_2}$ for all $\gamma_1,\gamma_2\in\Gamma$.

We now transform topological dynamical systems into (concrete) measure-preserving systems by adding a suitable probability measure. Recall that a probability measure $\mu\colon\mathcal{B}(K)\to[0,1]$ on the Borel σ-algebra $\mathcal{B}(K)$ of a compact space K is **regular** if

$$\begin{aligned}\mu(A)&=\sup\{\mu(L)\mid L\subseteq K \text{ compact with } L\subseteq A\}\\&=\inf\{\mu(O)\mid O\subseteq K \text{ open with } A\subseteq O\}\end{aligned}$$

for all $A\in\mathcal{B}(K)$, see [Rud87, Paragraph 2.4]. Note that, since we only consider probability measures μ here, the second equality actually follows from the first by considering complements. The Lebesgue measure on $[0,1]$ is an important example of a regular Borel probability measure. We write $\mathrm{P}(K)$ for the set of all regular Borel probability measures on a compact space K.

There is an equivalent way to look at regular Borel probability measures based on the following observation: If $\mu\colon\mathcal{B}(K)\to[0,1]$ is a regular Borel probability measure on a compact space K, every continuous function $f\colon K\to\mathbb{C}$ is Borel measurable and bounded, hence integrable with respect to μ. We therefore obtain an "integration map"

$$\varphi_\mu\colon \mathrm{C}(K)\to\mathbb{C},\quad f\mapsto\int_K f\,\mathrm{d}\mu.$$

Moreover, $\varphi = \varphi_\mu$ is a linear map which is

(i) **positive**, i.e., $\varphi(f) \geq 0$ for all $f \in \mathrm{C}(K)$ with $f \geq 0$, and

(ii) **unital**, i.e., $\varphi(\mathbb{1}) = 1$ for the constant one-function $\mathbb{1}\colon K \to \mathbb{C}$.

The following representation result (see, e.g., [Rud87, Paragraph 2.3]) gives a converse.

Theorem 3.2.6 (Riesz–Markov–Kakutani)**.** *Let K be a compact space. For every unital, positive, linear map $\varphi\colon \mathrm{C}(K) \to \mathbb{C}$ there is a unique $\mu \in \mathrm{P}(K)$ with $\varphi = \varphi_\mu$.*

From now on, we identify for a compact space K each measure $\mu \in \mathrm{P}(K)$ with the corresponding linear form φ_μ. In particular, we write $\mu(f) \coloneqq \varphi_\mu(f) = \int_K f \,\mathrm{d}\mu$ for $f \in \mathrm{C}(K)$ and $\mu \in \mathrm{P}(K)$. In this way, every $\mu \in \mathrm{P}(K)$ becomes a linear map from $\mathrm{C}(K)$ to $\mathbb{C}$, and since $|\int_K f \,\mathrm{d}\mu| \leq \|f\|_\infty$ for every $f \in \mathrm{C}(K)$, it is bounded, hence an element of the dual Banach space $\mathrm{C}(K)'$ with norm $\|\mu\| \leq 1$. This allows us to topologize the set $\mathrm{P}(K)$.

On any normed space E the **weak* topology** is the smallest topology making all point evaluations

$$\mathrm{ev}_f\colon E' \to \mathbb{C}, \quad \varphi \mapsto \varphi(f)$$

for $f \in E$ continuous.[3] A net $(\varphi_i)_{i\in I}$ in E' converges to $\varphi \in E'$ with respect to the weak* topology precisely when the net $(\varphi_i(f))_{i\in I}$ converges to $\varphi(f)$ in $\mathbb{C}$ for every $f \in E$. We refer, e.g., to [Ped89, Sections 2.4 and 2.5] for more information on the weak* topology. The following is an important consequence of Tychonoff's theorem, Theorem 3.2.4 above (see, e.g., [Ped89, Theorem 2.5.2]).

Theorem 3.2.7 (Banach–Alaoglu)**.** *Let E be a normed space. Then the dual unit ball $\{\varphi \in E' \mid \|\varphi\| \leq 1\}$ is compact with respect to the weak* topology.*

In the context of regular Borel probability measures, this yields:

Proposition 3.2.8. *If K is a compact space, then $\mathrm{P}(K)$ is compact with respect to the weak* topology.*

Proof. In view of Theorem 3.2.7, since closed subsets of compact spaces are again compact, it suffices to show that $\mathrm{P}(K)$ is a closed subset of the dual unit ball $B \coloneqq \{\varphi \in \mathrm{C}(K)' \mid \|\varphi\| \leq 1\}$. However, we can write

$$\mathrm{P}(K) = B \cap \bigcap_{\substack{f\in \mathrm{C}(K)\\ f\geq 0}} \mathrm{ev}_f^{-1}([0,\infty)) \cap \mathrm{ev}_{\mathbb{1}}^{-1}(\{1\}),$$

and this implies that $\mathrm{P}(K)$ is closed as an intersection of closed subsets. □

[3] In other words, if E' is considered as a subset of $\mathbb{C}^E$, the space of all maps from E to $\mathbb{C}$, the weak* topology is simply the product topology.

To obtain measure-preserving systems from topological ones, we introduce the following definition.

Definition 3.2.9. Let (K, τ) be a topological dynamical system. A regular Borel probability measure $\mu\colon \mathcal{B}(K) \to [0,1]$ is **invariant** if τ_γ is measure-preserving with respect to μ for every $\gamma \in \Gamma$, i.e., $\mu(\tau_\gamma^{-1}(A)) = \mu(A)$ for all Borel sets $A \subseteq K$. We write $\mathrm{P}(K, \tau) \subseteq \mathrm{P}(K)$ for the set of all invariant probability measures.

Every invariant measure $\mu \in \mathrm{P}(K, \tau)$ of a topological dynamical system (K, τ) allows us to construct a concrete measure-preserving system $(K, \mathcal{B}(K), \mu, \tau)$ (see Definition 2.1.3), and then a measure-preserving system $(K, \mathcal{B}(K), \mu, \tau^*)$ in the sense of Definition 2.1.1.

For the following lemma notice that continuous maps between topological spaces are Borel measurable (by Proposition 1.1.4 (i)).

Lemma 3.2.10. *Let $\tau\colon K \to L$ be a continuous map between compact spaces and $\mu \in \mathrm{P}(K)$. Then the push-forward measure $\tau_*\mu\colon \mathcal{B}(L) \to [0,1]$, $A \mapsto \mu(\tau^{-1}(A))$ is a regular Borel probability measure with $\int_K f \circ \tau \,\mathrm{d}\mu = \int_L f \,\mathrm{d}\tau_*\mu$ for all $f \in \mathrm{C}(L)$.*

The proof is left as Exercise 3.7. This leads to the following characterization of invariant measures.

Proposition 3.2.11. *For a topological dynamical system (K, τ) and $\mu \in \mathrm{P}(K)$ the following assertions are equivalent.*

(a) *μ is invariant.*

(b) *$\int_K f \circ \tau_\gamma \,\mathrm{d}\mu = \int_K f \,\mathrm{d}\mu$ for all $f \in \mathrm{C}(K)$ and $\gamma \in \Gamma$.*

Does every topological dynamical system admit an invariant measure? To answer this question, we first introduce the following concept closely related to the notion of ergodic nets from the previous section.

Definition 3.2.12. For a topological dynamical system (K, τ) we call a net $(\mu_i)_{i \in I}$ in $\mathrm{P}(K)$ **asymptotically invariant** if $\lim_{i \in I} \int_K f - (f \circ \tau_\gamma) \,\mathrm{d}\mu_i = 0$ for every $f \in \mathrm{C}(K)$ and each $\gamma \in \Gamma$.

Example 3.2.13. Let $(F_i)_{i \in I}$ be a Følner net for the group Γ. Then for a topological dynamical system (K, τ) and any $\mu \in \mathrm{P}(K)$ (e.g., $\mu = \delta_x$ a point measure for some $x \in K$), set $\mu_i := \frac{1}{|F_i|} \sum_{\gamma \in F_i} (\tau_\gamma)_* \mu$ for $i \in I$. One can check, with similar arguments as in the proof of Proposition 3.1.14, that $(\mu_i)_{i \in I}$ is asymptotically invariant, see Exercise 3.8.

Proposition 3.2.14. *Let (K, τ) be a topological dynamical system. Every weak* accumulation point μ of an asymptotically invariant net $(\mu_i)_{i \in I}$ in $\mathrm{P}(K)$ is invariant.*

Proof. Take $\gamma \in \Gamma$ and $f \in \mathrm{C}(K)$. For $\varepsilon > 0$ we find some $i_0 \in I$ with $|\int_K f - (f \circ \tau_\gamma)\,\mathrm{d}\mu_i| < \varepsilon/3$ for each $i \geq i_0$. By the definition of the weak* topology, the set

$$O \coloneqq \{\nu \in \mathrm{P}(K) \mid |\mu(f) - \nu(f)| < \varepsilon/3\} \cap \{\nu \in \mathrm{P}(K) \mid |\mu(f \circ \tau_\gamma) - \nu(f \circ \tau_\gamma)| < \varepsilon/3\}$$

is an open set containing μ. By the definition of accumulation points, we thus find some $i \geq i_0$ with $\mu_i \in O$. But then $|\mu(f) - \mu(f \circ \tau_\gamma)| \leq \varepsilon$ by the triangle inequality. Thus, $\mu(f) = \mu(f \circ \tau_\gamma)$, and μ is invariant by Proposition 3.2.11. □

Corollary 3.2.15 (Krylov–Bogolyubov)**.** *Every topological dynamical system* (K, τ) *has an invariant measure* $\mu \in \mathrm{P}(K, \tau)$.

Proof. By Proposition 3.1.10 and Example 3.2.13 the system (K, τ) has an asymptotically invariant net $(\mu_i)_{i \in I}$ in $\mathrm{P}(K)$. By Proposition 3.2.8 the net $(\mu_i)_{i \in I}$ has a weak* accumulation point $\mu \in \mathrm{P}(K)$, and then μ is invariant by Proposition 3.2.14. □

3.3 Comments and Further Reading

Von Neumann proved his mean ergodic theorem in [vN32b], but in a completely different way than what is presented here. His work was a mathematical formalization of the *ergodic hypothesis* from statistical mechanics going back to Ludwig Boltzmann, see, e.g., [Bad06] for more information. Even though not part of the lecture, we also mention the pointwise ergodic theorem of George Birkhoff showing pointwise almost everywhere convergence of the Cesàro means $\frac{1}{N}\sum_{n=0}^{N-1} U_\tau^n f$ (instead of L^2-convergence), see [Bir31]. Both results had a tremendous impact on the further development of mathematical ergodic theory, but also caused a dispute between von Neumann and Birkhoff - more on this can be read, e.g., in [Ber04].

The "abstract mean ergodic theorem" of this lecture is a special case of a general result by Alaoglu and Birkhoff, see [AB40]. Our exposition is based on [EFHN15, Supplement of Chapter 8]. There are many works on mean ergodic theorems for operators and operator semigroups on Banach spaces, see, e.g., [Nag73], [Sat78], [Kre85, Chapter 2], [EFHN15, Chapter 8], [Sch13], and [Kre18].

Groups having a Følner net are called *amenable*. There are many equivalent characterizations of this property, and interesting examples, see, e.g., [Pat88], [Run02], [CSD21], and [Jus22]. Our proof of the fact that abelian groups are amenable relies on the property that these groups have "subexponential growth", see [CSD21, Corollary 14.7] or [Jus22, Subsection 2.6].

An introduction to topological dynamics is given in [Aus88], see also the classical monograph [GH55], as well as [Bro79] and [dV93]. Here we are mostly interested in constructing measure-preserving systems from topological ones. This will be a crucial procedure in the next lecture.

3.4 Exercises

Exercise 3.1. A sequence $(a_n)_{n\in\mathbb{N}}$ of complex numbers is *Cesàro convergent* to $a \in \mathbb{C}$ if $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} a_n = a$, and in this case a is the *Cesàro limit* of $(a_n)_{n\in\mathbb{N}}$.

(i) Show that for every sequence of real numbers $(a_n)_{n\in\mathbb{N}}$ the inequalities

$$\liminf_{n\to\infty} a_n \leq \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} a_n \leq \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} a_n \leq \limsup_{n\to\infty} a_n$$

hold.

(ii) Show that if a sequence $(a_n)_{n\in\mathbb{N}}$ of complex numbers converges to $a \in \mathbb{C}$, then it is also Cesàro convergent to a.

(iii) Let $a_n = \lambda^n$ for $n \in \mathbb{N}$, where $\lambda \in \mathbb{C}$. Determine (depending on λ), if $(a_n)_{n\in\mathbb{N}}$ is Cesàro convergent, and in this case compute the Cesàro limit.

(iv) Give an example of a bounded sequence in $\mathbb{C}$ which is not Cesàro convergent.

(v) Show that a sequence $(a_n)_{n\in\mathbb{N}}$ of complex numbers converges if and only if each subsequence of $(a_n)_{n\in\mathbb{N}}$ is Cesàro convergent.

Exercise 3.2. Give an example of a Banach space E, a bounded linear operator $U \in \mathscr{L}(E)$ with $\|U\| \leq 1$, and an element $f \in E$ such that the sequence $(\frac{1}{N}\sum_{n=0}^{N-1} U^n f)_{N\in\mathbb{N}}$ does not converge in E.

Exercise 3.3. Let $U \in \mathscr{L}(H)$ be a bounded linear operator on a Hilbert space H with $\|U\| \leq 1$ and $\mathscr{S}_U$ the generated contraction semigroup from Remark 3.1.4. Consider the reversed order on $(1,\infty)$ and define

$$V_r := (1 - {}^1\!/\!_r)\sum_{n=0}^{\infty} r^{-n}U^n = \lim_{N\to\infty} (1 - {}^1\!/\!_r)\sum_{n=0}^{N} r^{-n}U^n$$

for $r \in (1,\infty)$ where the limit exists with respect to the operator norm. Show that $(V_r)_{r\in(1,\infty)}$ is an ergodic net for $\mathscr{S}_U$.

Exercise 3.4. Let $(F_i)_{i\in I}$ be a Følner net for Γ. Show that for a measure-preserving system (X,T) the following assertions are equivalent.

(a) (X,T) is ergodic.

(b) For all $f \in \mathrm{L}^2(X)$ we have $\lim_\alpha \frac{1}{|F_\alpha|}\sum_{\gamma\in F_\alpha} U_{T_\gamma} f = (\int_X f)\cdot \mathbb{1}$ in $\mathrm{L}^2(X)$.

Exercise 3.5. Let $\emptyset \neq \mathscr{S} \subseteq \mathscr{L}(H)$ be an abelian contraction semigroup on a Hilbert space H, i.e., $UV = VU$ for all $U, V \in \mathscr{S}$. Equip the convex hull $I := \operatorname{co}\mathscr{S}$ with the relation $\leq$ defined by

$$U \leq V \quad :\Leftrightarrow \quad U = V \text{ or there is } W \in I \text{ with } V = WU$$

for $U, V \in I$. Show the following assertions.

(i) I is a directed set.

(ii) The map $I \to \mathscr{L}(H)$, $V \mapsto V$ is an ergodic net.
Hint: Note that $\frac{1}{N}\sum_{n=0}^{N-1} U^n \in I$ *for every* $N \in \mathbb{N}$ *and* $U \in \mathscr{S}$.

Exercise 3.6. A topological dynamical system (K, τ) is **topologically transitive** if for all non-empty open subsets $O_1, O_2 \subseteq K$ there is some $\gamma \in \Gamma$ such that $\tau_\gamma^{-1}(O_1) \cap O_2 \neq \emptyset$. Show that if Γ is an infinite abelian group, then for any non-empty compact space K the shift (K^Γ, τ) from Example 3.2.5 is topologically transitive.

Exercise 3.7. Prove Lemma 3.2.10.

Exercise 3.8. Consider a Følner net $(F_i)_{i \in I}$ for the group Γ, a topological dynamical system (K, τ), and $\mu \in \mathrm{P}(K)$. Show that the net $(\mu_i)_{i \in I}$ with $\mu_i \coloneqq \frac{1}{|F_i|}\sum_{\gamma \in F_i} (\tau_\gamma)_* \mu$ for $i \in I$ is asymptotically invariant.

Lecture 4

We continue the study of invariant measures started in the previous chapter and characterize so-called ergodic measures. With the established tools, we then build the bridge between ergodic theory and additive combinatorics alluded to in the first lecture. As a first application we prove a result on patterns in subsets of the integers.

4.1 Ergodic Measures

In the previous lecture we have seen that we can always turn a topological dynamical system into a measure-preserving one via an invariant measure. But can we also achieve that these systems are ergodic (see Definition 2.1.9)?

Definition 4.1.1. Let (K, τ) be a topological dynamical system. An invariant measure $\mu \in \mathrm{P}(K, \tau)$ is an **ergodic measure** if the induced measure-preserving system $(K, \mathcal{B}(K), \mu, \tau^*)$ is ergodic.

Before we address the problem of existence of such measures, we derive an important characterization. It relies on the following simple observation. Recall from Section 3.2 that for any compact space K we identify regular Borel probability measures $\mu \colon \mathcal{B}(K) \to [0, 1]$ with the induced positive, unital, linear functionals $\mathrm{C}(K) \to \mathbb{C}$, $f \mapsto \int_K f \, \mathrm{d}\mu$. In particular, we can view $\mathrm{P}(K)$ as a subset of the dual space $\mathrm{C}(K)'$.

Proposition 4.1.2. (i) *For any compact space K, the set of regular Borel probability measures $\mathrm{P}(K)$ is a convex and weak* compact subset of $\mathrm{C}(K)'$.*

(ii) *For any topological dynamical system (K, τ), the set of invariant regular Borel probability measures $\mathrm{P}(K, \tau)$ is a convex and weak* compact subset of $\mathrm{C}(K)'$.*

Proof. For part (i), recall that we have shown compactness with respect to the weak* topology in Proposition 3.2.8. To see that $\mathrm{P}(K)$ is convex, take $\mu_1, \mu_2 \in \mathrm{P}(K)$ and

$t \in [0,1]$. The linear functional $t\mu_1 + (1-t)\mu_2 \colon \mathrm{C}(K) \to \mathbb{C}$ satisfies

$$(t\mu_1 + (1-t)\mu_2)(f) = t\mu_1(f) + (1-t)\mu_2(f) \geq 0$$

for every $f \in \mathrm{C}(K)$ with $f \geq 0$. Moreover, $(t\mu_1 + (1-t)\mu_2)(\mathbb{1}) = 1$. Thus, $t\mu_1 + (1-t)\mu_2$ defines an element of $\mathrm{P}(K)$. Using Proposition 3.2.11, part (ii) readily follows from (i). □

Remark 4.1.3. For a compact space K and $\mu_1, \mu_2 \in \mathrm{P}(K)$, one can also define the convex combination $t\mu_1 + (1-t)\mu_2$ for $t \in [0,1]$ directly as a measure (instead of a linear functional $\mathrm{C}(K) \to \mathbb{C}$), see Exercise 4.1 below.

We are interested in elements of a convex set C which cannot be written as a non-trivial convex combination of two other points in C.

Definition 4.1.4. For a convex subset $C \subseteq E$ of a (real or complex) vector space E we call $v \in E$ an **extreme point** if the identity $tv_1 + (1-t)v_2 = v$ for $t \in (0,1)$ and $v_1, v_2 \in C$ implies $v = v_1 = v_2$. We write $\operatorname{ex} C$ for the set of all extreme points of the set C.

The definition is illustrated by the following two very basic examples.

Examples 4.1.5. (i) Consider the unit interval $C := [0,1] \subseteq \mathbb{C}$. Then the boundary points $0, 1$ are the extreme points of C.

(ii) Consider the triangle $C := \{r + s\mathrm{i} \mid r, s \geq 0 \text{ with } r + s \leq 1\} \subseteq \mathbb{C}$. One can check that the extreme points of C are precisely the corners $0, 1, \mathrm{i} \in C$.

We now prove a characterization of ergodic measures.

Proposition 4.1.6. *For a topological dynamical system (K, τ) and an invariant measure $\mu \in \mathrm{P}(K, \tau)$ the following assertions are equivalent.*

(a) *μ is an ergodic measure.*

(b) *μ is an extreme point of $\mathrm{P}(K, \tau)$.*

We use the following well-known fact for regular Borel probability measures (see, e.g., [Rud87, Theorem 3.14]).

Lemma 4.1.7. *For every compact space K, each $\mu \in \mathrm{P}(K)$, and any $p \in [1, \infty)$, the natural map $\mathrm{C}(K) \to \mathrm{L}^p(K, \mathcal{B}(K), \mu)$ sending a function $f \in \mathrm{C}(K)$ to its equivalence class in $\mathrm{L}^p(K, \mathcal{B}(K), \mu)$ has dense range.*

Proof of Proposition 4.1.6. We first prove the implication "(a) ⇒ (b)". So assume that μ is an ergodic measure, and recall that by Corollary 2.2.14 this means that the fixed space

$$\operatorname{fix}(U_{\tau^*}) = \{f \in \mathrm{L}^2(K, \mathcal{B}(K), \mu) \mid f \circ \tau_\gamma = f \text{ for every } \gamma \in \Gamma\}$$

only consists of (equivalence classes of) constant functions $c \cdot \mathbb{1}$ for $c \in \mathbb{C}$. If $\mu = t\mu_1 + (1-t)\mu_2$ for $\mu_1, \mu_2 \in \mathrm{P}(K, \tau)$ and $t \in (0,1)$, then

$$\left| \int_K f \, \mathrm{d}\mu_1 \right| \leq \int_K |f| \, \mathrm{d}\mu_1 \leq \frac{1}{t} \int_K |f| \, \mathrm{d}\mu = \frac{1}{t} \|f\|_{\mathrm{L}^1(K,\mathcal{B},\mu)} \leq \frac{1}{t} \|f\|_{\mathrm{L}^2(K,\mathcal{B},\mu)}$$

for every $f \in \mathrm{C}(K)$. In particular, if we write $I \colon \mathrm{C}(K) \to \mathrm{L}^2(K, \mathcal{B}(K), \mu)$ for the map from Lemma 4.1.7 for $p = 2$, we obtain the following implication: Whenever $If_1 = If_2$, hence $I(f_1 - f_2) = 0$, for $f_1, f_2 \in \mathrm{C}(K)$, we have $\int_K f_1 \, \mathrm{d}\mu_1 = \int_K f_2 \, \mathrm{d}\mu_1$. We therefore obtain a (well-defined!) linear functional

$$\varphi \colon I(\mathrm{C}(K)) \to \mathbb{C}, \quad If \mapsto \int_K f \, \mathrm{d}\mu_1.$$

Moreover, φ is bounded with $\|\varphi\| \leq \frac{1}{t}$. Since $I(\mathrm{C}(K))$ is dense in $\mathrm{L}^2(K, \mathcal{B}(K), \mu)$ by Lemma 4.1.7, we can apply Proposition A.1.1 to uniquely extend φ to a bounded linear functional $\varphi \colon \mathrm{L}^2(K, \mathcal{B}(K), \mu) \to \mathbb{C}$. By the Riesz-Fréchet representation theorem (see Theorem A.2.3), we therefore find a unique element $g \in \mathrm{L}^2(K, \mathcal{B}(K), \mu)$ with

$$\int_K f \, \mathrm{d}\mu_1 = \varphi(f) = (f|g)_{\mathrm{L}^2(K,\mathcal{B}(K),\mu)} = \int_K f \cdot \overline{g} \, \mathrm{d}\mu$$

for all $f \in \mathrm{C}(K)$.

We show that $g \in \mathrm{fix}(U_{\tau^*})$. Again using that $I(\mathrm{C}(K))$ is dense in $\mathrm{L}^2(K, \mathcal{B}(K), \mu)$, it suffices to show that $(f|g - g \circ \tau_\gamma) = 0$ for every $\gamma \in \Gamma$ and for each $f \in \mathrm{C}(K)$ since this implies $(g - g \circ \tau_\gamma | g - g \circ \tau_\gamma) = 0$ by approximation and continuity of the inner product. But, since both μ and μ_1 are invariant, we obtain from Proposition 1.1.4 that

$$\begin{aligned} (f|g - g \circ \tau_\gamma) &= \int_K f \cdot \overline{g} \, \mathrm{d}\mu - \int_K f \cdot \overline{g \circ \tau_\gamma} \, \mathrm{d}\mu = \int_K f \cdot \overline{g} \, \mathrm{d}\mu - \int_K (f \circ \tau_\gamma^{-1}) \cdot \overline{g} \, \mathrm{d}\mu \\ &= \int_K f \, \mathrm{d}\mu_1 - \int_K f \circ \tau_{-\gamma} \, \mathrm{d}\mu_1 = 0 \end{aligned}$$

for every $f \in \mathrm{C}(K)$ and $\gamma \in \Gamma$.

By ergodicity of μ we conclude that $g = c \cdot \mathbb{1}$ for some $c \in \mathbb{C}$. By choice of g, this means that $\int_K f \, \mathrm{d}\mu_1 = c \cdot \int_K f \, \mathrm{d}\mu$ for all $f \in \mathrm{C}(K)$. Taking $f = \mathbb{1}$, we see that $c = 1$, and hence $\mu = \mu_1$. But then $\mu = t\mu + (1-t)\mu_2$ automatically also yields $\mu = \mu_2$, finishing the proof of "(a) $\Rightarrow$ (b)".

We now show the converse implication "(b) $\Rightarrow$ (a)" by contraposition. If μ is not ergodic, we find some $A \in \mathcal{B}(K)$ with $0 < \mu(A) < 1$ such that $\mu(A \,\Delta\, \tau_\gamma^{-1}(A)) = 0$

for each $\gamma \in \Gamma$. Then

$$\mu_1 \colon \mathrm{C}(K) \to \mathbb{C}, \quad f \mapsto \frac{1}{\mu(A)} \int_A f \,\mathrm{d}\mu,$$
$$\mu_2 \colon \mathrm{C}(K) \to \mathbb{C}, \quad f \mapsto \frac{1}{\mu(K \setminus A)} \int_{K\setminus A} f \,\mathrm{d}\mu$$

are unital, positive, linear functionals and hence define elements of $\mathrm{P}(K)$. Since

$$\begin{aligned} \mu_1(f \circ \tau_\gamma) &= \frac{1}{\mu(A)} \int_K \mathbb{1}_A \cdot (f \circ \tau_\gamma) \,\mathrm{d}\mu = \frac{1}{\mu(A)} \int_K \mathbb{1}_{\tau_\gamma^{-1}(A)} \cdot (f \circ \tau_\gamma) \,\mathrm{d}\mu \\ &= \frac{1}{\mu(A)} \int_K (\mathbb{1}_A \cdot f) \circ \tau_\gamma \,\mathrm{d}\mu = \frac{1}{\mu(A)} \int_K \mathbb{1}_A \cdot f \,\mathrm{d}\mu = \mu_1(f) \end{aligned}$$

for every $f \in \mathrm{C}(K)$, we conclude that $\mu_1 \in \mathrm{P}(K, \tau)$ (see Proposition 3.2.11). Similarly, $\mu_2 \in \mathrm{P}(K, \tau)$. Moreover, for $t := \mu(A) \in (0, 1)$ we have $\mu = t\mu_1 + (1 - t)\mu_2$.

To finish the proof we check that $\mu \neq \mu_1$ which shows that μ is not an extreme point of $\mathrm{P}(K, \tau)$. Since $0 < \mu(A) < 1$ we have $\mu(A)\mathbb{1} \neq \mathbb{1}_A$ in $\mathrm{L}^2(K, \mathcal{B}(K), \mu)$. Again using Lemma 4.1.7, we find some $f \in \mathrm{C}(K)$ with $0 \neq (f|\mu(A)\mathbb{1} - \mathbb{1}_A) = \int_K (\mu(A)f - \mathbb{1}_A f)$. But this means $\mu(f) \neq \mu_1(f)$. □

Proposition 4.1.6 is particularly interesting when combined with the following special case of the Krein-Milman theorem from functional analysis, see, e.g., [Ped89, Theorem 2.5.4].

Theorem 4.1.8. *Let E be a normed space. If C is a weak* compact convex subset of the dual E', then the convex hull of* $\mathrm{ex}\, C$ *is weak* dense in C.*

We obtain the following consequence.

Proposition 4.1.9. *Every topological dynamical system (K, τ) has an ergodic measure $\mu \in \mathrm{P}(K, \tau)$.*

Proof. By Corollary 3.2.15 we know that $\mathrm{P}(K, \tau) \neq \emptyset$. Since $\mathrm{P}(K, \tau)$ is weak* compact and convex, we can apply Theorem 4.1.8 to see that, in particular, the set $\mathrm{P}(K, \tau)$ has some extreme point $\mu \in \mathrm{P}(K, \tau)$. By Proposition 4.1.6 we obtain that this μ is ergodic. □

4.2 Furstenberg's Correspondence Principle

With these concepts and tools at hand, we are now ready to establish a bridge between ergodic theory and combinatorial number theory. This will help us to find structure in "large" subsets of the integers, e.g., the following patterns.

Definition 4.2.1. For $k \in \mathbb{N}$ a set of the form $\{a, a+d, \ldots, a+(k-1)d\}$ for $a, d \in \mathbb{N}$ is called an **arithmetic progression of length** k.

One of the early questions on arithmetic progressions from so-called Ramsey Theory is the following: Assume that we color the natural numbers in two different colors, say, red and blue. Is there, for any $k \in \mathbb{N}$, an arithmetic progression of length k which is monochromatic, i.e., consists entirely of red or entirely of blue numbers? In some simple cases (e.g., there are only finitely many red numbers; or we color all odd numbers red, and even numbers blue), the answer is evidently positive. However, the general case is surprisingly difficult, and was first solved by Bartel Leendert van der Waerden in 1927, even in a stronger version:

Theorem 4.2.2 (van der Waerden)**.** *Assume that* $\mathbb{N} = A_1 \cup \cdots \cup A_m$ *for some* $m \in \mathbb{N}$. *Then there is* $j \in \{1, \ldots, m\}$ *such that* A_j *contains arithmetic progressions of arbitrary (finite) length.*

In 1975 Endre Szemerédi proved a stronger result for "asymptotically large" subsets. Before we recall his Theorem 1.1.8 from Lecture 1, we introduce the following important concept. Here, for a bounded net $(r_i)_{i\in I}$ of real numbers, we write $\limsup_{i\in I} r_i$ for the largest and $\liminf_{i\in I} r_i$ for the smallest accumulation point of $(r_i)_{i\in I}$ (which, as for sequences, always exist).

Definition 4.2.3. Let $(F_i)_{i\in I}$ be a Følner net (see Definition 3.1.9) for the abelian group Γ. For a subset $A \subseteq \Gamma$ we call

$$\overline{\mathrm{d}}_{(F_i)_{i\in I}}(A) := \limsup_{i\in I} \frac{|A \cap F_i|}{|F_i|} \quad \text{and} \quad \underline{\mathrm{d}}_{(F_i)_{i\in I}}(A) := \liminf_{i\in I} \frac{|A \cap F_i|}{|F_i|}$$

the **upper density** and **lower density** of A with respect to $(F_i)_{i\in I}$, respectively.

A concrete example is the following.

Example 4.2.4. Given a set $A \subseteq \mathbb{N}$, we usually consider the **natural upper** and **lower density**

$$\overline{\mathrm{d}}(A) := \limsup_{N\to\infty} \frac{|A \cap \{1, \ldots, N\}|}{N} \quad \text{and} \quad \underline{\mathrm{d}}(A) := \liminf_{N\to\infty} \frac{|A \cap \{1, \ldots, N\}|}{N}$$

with respect to the Følner net $(F_N)_{N\in\mathbb{N}}$ for $\Gamma = \mathbb{Z}$ given by $F_N := \{1, \ldots, N\}$ for $N \in \mathbb{N}$.

The following lemma collects some basic facts about upper and lower densities. The proof is left as Exercise 4.5.

Lemma 4.2.5. *Let* $(F_i)_{i\in I}$ *be a Følner net for the abelian group* Γ. *Then the following assertions hold for* $A, B \subseteq \Gamma$.

(i) $0 \leq \underline{\mathrm{d}}_{(F_i)_{i\in I}}(A) \leq \overline{\mathrm{d}}_{(F_i)_{i\in I}}(A) \leq 1$.

(ii) $\underline{\mathrm{d}}_{(F_i)_{i\in I}}(\emptyset) = \overline{\mathrm{d}}_{(F_i)_{i\in I}}(\emptyset) = 0$ *and* $\underline{\mathrm{d}}_{(F_i)_{i\in I}}(\Gamma) = \overline{\mathrm{d}}_{(F_i)_{i\in I}}(\Gamma) = 1$.

(iii) $\underline{\mathrm{d}}_{(F_i)_{i\in I}}(A) \le \underline{\mathrm{d}}_{(F_i)_{i\in I}}(B)$ *and* $\overline{\mathrm{d}}_{(F_i)_{i\in I}}(A) \le \overline{\mathrm{d}}_{(F_i)_{i\in I}}(B)$ *if* $A \subseteq B$.

(iv) $\overline{\mathrm{d}}_{(F_i)_{i\in I}}(A) + \underline{\mathrm{d}}_{(F_i)_{i\in I}}(\Gamma \setminus A) = 1$.

(v) $\overline{\mathrm{d}}_{(F_i)_{i\in I}}(A \cup B) \le \overline{\mathrm{d}}_{(F_i)_{i\in I}}(A) + \overline{\mathrm{d}}_{(F_i)_{i\in I}}(B)$.

(vi) *If* $\underline{\mathrm{d}}_{(F_i)_{i\in I}}(B) = 1$, *then*

$$\overline{\mathrm{d}}_{(F_i)_{i\in I}}(A \cap B) = \overline{\mathrm{d}}_{(F_i)_{i\in I}}(A) \quad \text{and} \quad \underline{\mathrm{d}}_{(F_i)_{i\in I}}(A \cap B) = \underline{\mathrm{d}}_{(F_i)_{i\in I}}(A).$$

(vii) *If* $\overline{\mathrm{d}}_{(F_i)_{i\in I}}(B) = 0$, *then*

$$\overline{\mathrm{d}}_{(F_i)_{i\in I}}(A \cup B) = \overline{\mathrm{d}}_{(F_i)_{i\in I}}(A) \quad \text{and} \quad \underline{\mathrm{d}}_{(F_i)_{i\in I}}(A \cup B) = \underline{\mathrm{d}}_{(F_i)_{i\in I}}(A).$$

We now restate Theorem 1.1.8 as follows.

Theorem 4.2.6 (Szemerédi)**.** *Let* $A \subseteq \mathbb{N}$ *with* $\overline{\mathrm{d}}(A) > 0$. *Then* A *contains arithmetic progressions of arbitrary (finite) length.*

Remark 4.2.7. Notice that if $\mathbb{N} = A_1 \cup \cdots \cup A_m$ for some $m \in \mathbb{N}$, then parts (ii) and (v) of Lemma 4.2.5 imply that there is some $j \in \{1, \ldots, m\}$ with $\overline{\mathrm{d}}(A_j) > 0$. Thus, Szemerédi's Theorem 4.2.6 is indeed stronger than van der Waerden's Theorem 4.2.2.

Let us now discuss how to approach these results via dynamical systems following ideas of Hillel Furstenberg from 1977. For any abelian group Γ (e.g., $\Gamma = \mathbb{Z}$) and the discrete space $\{0, 1\}$ consider the corresponding shift system $(\{0,1\}^\Gamma, \tau)$ from Example 3.2.5. Then any subset $A \subseteq \Gamma$ defines a point $\mathbf{1}_A := (\delta_{\gamma,A})_{\gamma\in\Gamma}$ of this system via

$$\delta_{\gamma,A} := \begin{cases} 1 & \text{for } \gamma \in A, \\ 0 & \text{for } \gamma \notin A. \end{cases}$$

We consider its "orbit closure" $K := \overline{\{\tau_\gamma(\mathbf{1}_A) \mid \gamma \in \Gamma\}} \subseteq \{0,1\}^\Gamma$ with the subspace topology. For each $\gamma \in \Gamma$ we have $\tau_\gamma(K) = K$, and hence τ_γ restricts to a homeomorphism $\sigma_\gamma := (\tau_\gamma)|_K : K \to K$. We therefore obtain a topological dynamical system (K, σ) via $\Gamma \to \mathrm{Homeo}(K)$, $\gamma \mapsto \sigma_\gamma$.

By definition of the product topology, the "cylinder set"

$$B := \{(a_\gamma)_{\gamma\in\Gamma} \in K \mid a_0 = 1\}$$

is both open and closed in K. We can use B to describe the elements of $A \subseteq \Gamma$: For $\gamma \in \Gamma$ we have $\gamma \in A$ precisely when $\delta_{\gamma,A} = 1$. But this means

$$\gamma \in A \quad \Leftrightarrow \quad \sigma_\gamma(\mathbf{1}_A) \in B.$$

Assume now that, for some $\gamma_1, \dots, \gamma_k \in \Gamma$, we knew that

$$\sigma_{\gamma_1}^{-1}(B) \cap \cdots \cap \sigma_{\gamma_k}^{-1}(B) \neq \emptyset.$$

Since this is an open set, and $\{\sigma_\gamma(\mathbf{1}_A) \mid \gamma \in \Gamma\}$ is dense in K, we then find some $\gamma \in \Gamma$ with $\sigma_\gamma(\mathbf{1}_A) \in \sigma_{\gamma_1}^{-1}(B) \cap \cdots \cap \sigma_{\gamma_k}^{-1}(B)$. But this means $\gamma + \gamma_1, \dots, \gamma + \gamma_k \in A$. In the special case of $\Gamma = \mathbb{Z}$ and $\gamma_j = (j-1)d$ for $j \in \{1, \dots, k\}$ and some $d \in \mathbb{N}$, we thus have $a, a+d, \cdots, a+(k-1)d \in A$ for $a \coloneqq \gamma$, and thus an arithmetic progression of length k in A.

So far we have not taken the upper density of the subset $A \subseteq \Gamma$ into account. We will do so now to construct a measure-preserving system.

Theorem 4.2.8 (Furstenberg's Correspondence Principle I). *Let $(F_i)_{i\in I}$ be a Følner net for the abelian group Γ, and let $A \subseteq \Gamma$ with $\overline{\mathrm{d}}_{(F_i)_{i\in I}}(A) > 0$. Then there is an ergodic measure-preserving system (X, T) and $B \in \Sigma(X)$ with $\mu_X(B) > 0$ having the following property: If $\mu_X(T_{\gamma_1}(B) \cap \cdots \cap T_{\gamma_k}(B)) > 0$ for some $\gamma_1, \dots, \gamma_k \in \Gamma$ and $k \in \mathbb{N}$, then $\gamma + \gamma_1, \dots, \gamma + \gamma_k \in A$ for some $\gamma \in \Gamma$.*

If we forego ergodicity of the measure-preserving system (X, T), we can even give an estimate on "how many" $\gamma \in \Gamma$ satisfy $\gamma + \gamma_1, \dots, \gamma + \gamma_k \in A$, or, equivalently, $\gamma \in A - \gamma_1, \dots, A - \gamma_k$.

Theorem 4.2.9 (Furstenberg's Correspondence Principle II). *Let $(F_i)_{i\in I}$ be a Følner net for the abelian group Γ and $A \subseteq \Gamma$. Then there is a measure-preserving system (X, T) and $B \in \Sigma(X)$ with $\mu_X(B) = \overline{\mathrm{d}}_{(F_i)_{i\in I}}(A)$ such that*

$$\mu_X(T_{\gamma_1}(B) \cap \cdots \cap T_{\gamma_k}(B)) \leq \overline{\mathrm{d}}_{(F_i)_{i\in I}}((A - \gamma_1) \cap \cdots \cap (A - \gamma_k))$$

for all $\gamma_1, \dots, \gamma_k \in \Gamma$ and $k \in \mathbb{N}$.

We prove both results at once.

Proof of Theorems 4.2.8 and 4.2.9. Take the topological dynamical system (K, σ) as well as the open and closed subset $B \subseteq K$ from above. Then the characteristic function $\mathbb{1}_B \colon K \to \mathbb{C}$ of B is continuous, and for $\gamma \in \Gamma$ we have

$$\gamma \in A \quad \Leftrightarrow \quad \sigma_\gamma(\mathbf{1}_A) \in B \quad \Leftrightarrow \quad (\mathbb{1}_B \circ \sigma_\gamma)(\mathbf{1}_A) = 1.$$

We can reformulate this condition once more by using pushforward measures of the point measure $\delta_{\mathbf{1}_A} \colon \mathrm{C}(K) \to \mathbb{C}$, $f \mapsto f(\mathbf{1}_A)$. For $\gamma \in \Gamma$ we have

$$\gamma \in A \quad \Leftrightarrow \quad ((\sigma_\gamma)_* \delta_{\mathbf{1}_A})(\mathbb{1}_B) = 1.$$

This allows us to write

$$\frac{|(A - \gamma_1) \cap \cdots \cap (A - \gamma_k) \cap F_i|}{|F_i|} = \frac{1}{|F_i|} \sum_{\gamma \in F_i} ((\sigma_\gamma)_* \delta_{\mathbf{1}_A})(\mathbb{1}_{\sigma_{\gamma_1}^{-1}(B) \cap \cdots \cap \sigma_{\gamma_k}^{-1}(B)})$$

for all $i \in I$ and $\gamma_1, \dots, \gamma_k \in \Gamma$. In particular, we obtain that

$$\overline{\mathrm{d}}_{(F_i)_{i\in I}}(A) = \limsup_{i\in I} \frac{1}{|F_i|} \sum_{\gamma\in F_i} ((\sigma_\gamma)_*\delta_{\mathbf{1}_A})(\mathbb{1}_B).$$

By the definition of accumulation points, we find for every $(i,n) \in I \times \mathbb{N}$ some $j(i,n) \in I$ with $j(i,n) \geq i$ and

$$\left| \overline{\mathrm{d}}_{(F_i)_{i\in I}}(A) - \frac{1}{|F_{j(i,n)}|} \sum_{\gamma\in F_{j(i,n)}} ((\sigma_\gamma)_*\delta_{\mathbf{1}_A})(\mathbb{1}_B) \right| < \frac{1}{n}.$$

Equip $I \times \mathbb{N}$ with the product direction, i.e., $(i_1, n_1) \leq (i_2, n_2)$ if $i_1 \leq i_2$ and $n_1 \leq n_2$ for $(i_1, n_1), (i_2, n_2) \in I \times \mathbb{N}$. Then $({}^1\!/\!{}_{|F_{j(i,n)}|} \sum_{\gamma\in F_{j(i,n)}} ((\sigma_\gamma)_*\delta_{\mathbf{1}_A}))_{(i,n)\in I\times\mathbb{N}}$ is a net[1] in $\mathrm{P}(K)$ with

$$\lim_{(i,n)\in I\times\mathbb{N}} \frac{1}{|F_{j(i,n)}|} \sum_{\gamma\in F_{j(i,n)}} ((\sigma_\gamma)_*\delta_{\mathbf{1}_A})(\mathbb{1}_B) = \overline{\mathrm{d}}_{(F_i)_{i\in I}}(A).$$

Using the assumption that $(F_i)_{i\in I}$ is a Følner net, we obtain that this net is asymptotically invariant (as in Example 3.2.13). By compactness of $\mathrm{P}(K)$ (see Proposition 3.2.8) it has a weak* accumulation point $\mu \in \mathrm{P}(K)$, and by Proposition 3.2.14 we obtain that μ is invariant. By definition of the weak* topology, this implies that also $\mu(f)$ is an accumulation point of the net of complex numbers $({}^1\!/\!{}_{|F_{j(i,n)}|} \sum_{\gamma\in F_{j(i,n)}} ((\sigma_\gamma)_*\delta_{\mathbf{1}_A})(f))_{(i,n)\in I\times\mathbb{N}}$ for every $f \in \mathrm{C}(K)$, and, in particular, we have

$$\mu(B) = \mu(\mathbb{1}_B) = \lim_{(i,n)\in I\times\mathbb{N}} \frac{1}{|F_{j(i,n)}|} \sum_{\gamma\in F_{j(i,n)}} ((\sigma_\gamma)_*\delta_{\mathbf{1}_A})(\mathbb{1}_B) = \overline{\mathrm{d}}_{(F_i)_{i\in I}}(A)$$

for $f = \mathbb{1}_B$. Again using the definition of accumulation points, for $f \in \mathrm{C}(K)$, $i_0 \in I$ and $\varepsilon > 0$ we find some index $(i,n) \in I \times \mathbb{N}$ satisfying $(i,n) \geq (i_0, 1)$ as well as $|{}^1\!/\!{}_{|F_{j(i,n)}|} \sum_{\gamma\in F_{j(i,n)}} ((\sigma_\gamma)_*\delta_{\mathbf{1}_A})(f) - \mu(f)| < \varepsilon$. But, as $j(i,n) \geq i \geq i_0$, this shows that $\mu(f)$ is also an accumulation point of the "original net" $({}^1\!/\!{}_{|F_i|} \sum_{\gamma\in F_i} ((\sigma_\gamma)_*\delta_{\mathbf{1}_A})(f))_{i\in I}$ for every $f \in \mathrm{C}(K)$. Taking $f = \mathbb{1}_{\sigma_{\gamma_1}^{-1}(B)\cap\cdots\cap\sigma_{\gamma_k}^{-1}(B)}$ for $\gamma_1, \dots, \gamma_k \in \Gamma$, we therefore have

$$\begin{aligned} \limsup_{i\in I} \frac{|(A-\gamma_1)\cap\cdots\cap(A-\gamma_k)\cap F_i|}{|F_i|} &= \limsup_{i\in I} \frac{1}{|F_i|} \sum_{\gamma\in F_i} ((\sigma_\gamma)_*\delta_{\mathbf{1}_A})(f) \\ &\geq \mu(f) = \int_K \mathbb{1}_{\sigma_{\gamma_1}^{-1}(B)\cap\cdots\cap\sigma_{\gamma_k}^{-1}(B)} \,\mathrm{d}\mu, \end{aligned}$$

[1] Readers familiar with the concept of nets will observe that we have constructed a *subnet*, which converges on $\mathbb{1}_B$ to the limit superior $\overline{\mathrm{d}}_{(F_i)_{i\in I}}(A)$. Working with the known relation between subnets and accumulation points, one can slightly shorten the proof.

and thus

$$\overline{\mathrm{d}}_{(F_i)_{i\in I}}((A-\gamma_1)\cap\cdots\cap(A-\gamma_k)) \geq \mu(\sigma_{\gamma_1}^{-1}(B)\cap\cdots\cap\sigma_{\gamma_k}^{-1}(B)).$$

With the induced measure-preserving system $(X,T) := (K,\mathcal{B}(K),\mu,\sigma^*)$ we thus obtain Theorem 4.2.9.

To construct an ergodic measure-preserving system as in Theorem 4.2.8, assume that $d := \overline{\mathrm{d}}_{(F_i)_{i\in I}}(A) > 0$. Then the set

$$O := \{\nu \in \mathrm{P}(K,\sigma) \mid \nu(\mathbb{1}_B) > {}^{d}\!/\!{}_{2}\}$$

is open in $\mathrm{P}(K,\sigma)$ with respect to the weak* topology and, since $\mu \in O$, also nonempty. By the Krein–Milman Theorem 4.1.8 and Proposition 4.1.6 we find a convex combination $\sum_{i=1}^r t_i\mu_i \in O$ of ergodic measures $\mu_i \in \mathrm{P}(K,\sigma)$ for $i \in \{1,\dots,r\}$, i.e., $t_i \in [0,1]$ for $i \in \{1,\dots,r\}$ with $\sum_{i=1}^r t_i = 1$. By the Pigeonhole principle, there has to be some $i \in \{1,\dots,r\}$ with $\mu_i(B) > \frac{d}{2} > 0$. We set $\nu := \mu_i$ and claim that $(X,T) := (K,\mathcal{B}(K),\nu,\sigma^*)$ gives us a measure-preserving system as in Theorem 4.2.8. Indeed, if for $\gamma_1,\dots,\gamma_k \in \Gamma$ we have $\nu(T_{\gamma_1}(B)\cap\cdots\cap T_{\gamma_k}(B)) > 0$, then, in particular, $\sigma_{\gamma_1}^{-1}(B)\cap\cdots\cap\sigma_{\gamma_k}^{-1}(B) = T_{\gamma_1}(B)\cap\cdots\cap T_{\gamma_k}(B) \neq \emptyset$. By the preliminary discussion (see page 57), we then find some $\gamma \in \Gamma$ with $\gamma+\gamma_1,\dots,\gamma+\gamma_k \in A$. □

Consequently, to prove Szemerédi's Theorem 4.2.6 it is sufficient to check that for every measure-preserving system (X,T) over $\Gamma = \mathbb{Z}$, every $B \in \Sigma(X)$ with $\mu_X(B) > 0$, and each $k \in \mathbb{N}$, we find some $n \in \mathbb{N}$ with

$$\mu_X(B \cap T^n(B) \cap \cdots \cap T^{kn}(B)) > 0,$$

cf. Theorem 1.1.7 from Lecture 1. The following result shows even more.

Theorem 4.2.10 (Furstenberg). *Let (X,T) be a measure-preserving system over $\Gamma = \mathbb{Z}$ and $f \in \mathrm{L}^\infty(X)$ with $f \geq 0, \int_X f\,\mathrm{d}\mu_X > 0$. For every $k \in \mathbb{N}$ we have*

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1}\int_X f\cdot U_T^n f\cdots U_T^{kn} f > 0.$$

Taking $f = \mathbb{1}_B$ gives, in particular, the desired multiple recurrence statement.

It will still take some time until we are in the position to give a proof of Theorem 4.2.10. We conclude the lecture with a different application of the correspondence principle instead.

Theorem 4.2.11 (Schur). *Assume that $\mathbb{N} = A_1 \cup \cdots \cup A_m$ for some $m \in \mathbb{N}$. Then there are $j \in \{1,\dots,m\}$ and $a,b \in \mathbb{N}$ such that $a,b,a+b \in A_j$.*

The proof discussed here is based on a "coloring trick" by Vitaly Bergelson and uses the following lemma.

Lemma 4.2.12. *Let $A \subseteq \mathbb{N}^d$ for some $d \in \mathbb{N}$ and consider the Følner net $(F_N)_{N\in\mathbb{N}^d}$ for $\mathbb{Z}^d$ given by $F_N := \{1,\dots,N_1\} \times \dots \times \{1,\dots,N_d\}$ for $N = (N_1,\dots,N_d) \in \mathbb{N}^d$ where $\mathbb{N}^d$ is equipped with the componentwise direction. Then*

$$\underline{\mathrm{d}}\big(\{n \in \mathbb{N} \mid \overline{\mathrm{d}}_{(F_N)_{N\in\mathbb{N}^d}}(A \cap (A - (n,\dots,n))) > 0\}\big) \geq \overline{\mathrm{d}}_{(F_N)_{N\in\mathbb{N}^d}}(A)^2.$$

Proof. We abbreviate $\vec{n} := (n,\dots,n) \in \mathbb{N}^d$ for $n \in \mathbb{N}$. For $\Gamma = \mathbb{Z}^d$ and the Følner net $(F_N)_{N\in\mathbb{N}^d}$ choose a measure-preserving system (X,T) as in Theorem 4.2.9. By Corollary 3.1.17 applied to the measure algebra homomorphism $T_{(1,\dots,1)}\colon \Sigma(X) \to \Sigma(X)$ we obtain

$$\begin{aligned}\overline{\mathrm{d}}_{(F_N)_{N\in\mathbb{N}^d}}(A)^2 = \mu_X(B)^2 &\leq \lim_{M\to\infty} \frac{1}{M}\sum_{n=1}^{M} \mu_X(B \cap T_{(n,\dots,n)}(B))\\ &\leq \liminf_{M\to\infty} \frac{1}{M}\sum_{n=1}^{M} \overline{\mathrm{d}}_{(F_N)_{N\in\mathbb{N}^d}}(A \cap (A - \vec{n}))\\ &\leq \liminf_{M\to\infty} \frac{|\{n \in \{1,\dots,M\} \mid \overline{\mathrm{d}}_{(F_N)_{N\in\mathbb{N}^d}}(A \cap (A-\vec{n})) > 0\}|}{M}.\end{aligned}$$

□

Proof of Theorem 4.2.11. By relabeling we assume that there is some $d \in \{1,\dots,m\}$ such that $\overline{\mathrm{d}}(A_i) > 0$ for $i \in \{1,\dots,d\}$ and $\overline{\mathrm{d}}(A_i) = 0$ for $i \in \{d+1,\dots,m\}$. Now let $(F_N)_{N\in\mathbb{N}^d}$ be the Følner net from Lemma 4.2.12. It is an easy exercise to check that the product set $A := A_1 \times \dots \times A_d \subseteq \mathbb{N}^d$ satisfies $\overline{\mathrm{d}}_{(F_N)_{N\in\mathbb{N}^d}}(A) = \overline{\mathrm{d}}(A_1)\cdots\overline{\mathrm{d}}(A_d) > 0$. In particular,

$$c := \underline{\mathrm{d}}\big(\{n \in \mathbb{N} \mid \overline{\mathrm{d}}_{(F_N)_{N\in\mathbb{N}^d}}(A \cap (A - (n,\dots,n))) > 0\}\big) > 0$$

by Lemma 4.2.12. On the other hand,

$$\underline{\mathrm{d}}(A_1 \cup \dots \cup A_d) = 1 - \overline{\mathrm{d}}(\mathbb{N} \setminus (A_1 \cup \dots \cup A_d)) \geq 1 - \overline{\mathrm{d}}(A_{d+1} \cup \dots \cup A_m) \geq 1$$

by Lemma 4.2.5 (iii), (iv), and (v). By Lemma 4.2.5 (vi) this implies

$$\underline{\mathrm{d}}\big((A_1 \cup \dots \cup A_d) \cap \{n \in \mathbb{N} \mid \overline{\mathrm{d}}_{(F_N)_{N\in\mathbb{N}^d}}(A \cap (A - (n,\dots,n))) > 0\}\big) = c > 0.$$

In particular, we find some $j \in \{1,\dots,d\}$ and $a \in A_j$ with

$$0 < \overline{\mathrm{d}}_{(F_N)_{N\in\mathbb{N}^d}}(A \cap (A - (a,\dots,a))) = \overline{\mathrm{d}}(A_1 \cap (A_1 - a))\cdots\overline{\mathrm{d}}(A_d \cap (A_d - a)).$$

But then $\overline{\mathrm{d}}(A_j \cap (A_j - a)) > 0$, and therefore we find $b \in A_j$ with $a + b \in A_j$. Hence $a, b, a+b \in A_j$, and we are done. □

Remark 4.2.13. Notice that Theorem 4.2.11 does not have a straightforward generalization to a "density version": If $A \subseteq \mathbb{N}$ is the set of all odd numbers, then $\overline{\mathrm{d}}(A) = \underline{\mathrm{d}}(A) = \frac{1}{2}$, but clearly $a + b \notin A$ for all $a, b \in A$.

4.3 Comments and Further Reading

The original publications of van der Waerden and Szemerédi concerning the topics of this lecture are [vdW27] and [Sze75], respectively. In his influential[2] article [Fur77], Furstenberg showed both his correspondence principle (in a slightly different form and with an alternative proof than what is presented here) and his multiple recurrence statement. We mention that one can also go the other direction and conclude the recurrence result from Szemerédi's theorem, see, e.g., [EFHN15, Section 20.2]. More on the impact of Furstenberg's work on ergodic theory, as well as many results related to the theorems of van der Waerden and Szemerédi (including "multidimensional versions"), can, e.g., be found in the recent article [BGW24].

Another famous theorem in this context is due to Ben Green and Terence Tao (see [GT08]): One can also find arithmetic progressions of arbitrary finite length within the set of primes (even though this set has upper density zero, see Exercise 4.8). Both Szemerédi's theorem and the Green–Tao theorem would be implied by a (still open) conjecture of Paul Erdős asserting that arithmetic progressions of arbitrary finite length[3] can be found in any subset $A \subseteq \mathbb{N}$ with $\sum_{n \in A} \frac{1}{n} = \infty$ (see [Gow13]).

Theorem 4.2.11 is based on Issai Schur's work [Sch17]. Our proof, using a technique of Bergelson from [Ber86], as well as the necessary version of Furstenberg's correspondence principle, is in essence from Joel Moreira's blog.[4]

[2] In 2020 Furstenberg received the Abel prize "for pioneering the use of methods from probability and dynamics in group theory, number theory and combinatorics."

[3] The case of 3-progressions was established in a recent breakthrough [BS20] by Bloom and Sisask.

[4] `https://joelmoreira.wordpress.com/2013/08/18/applications-of-the-coloring-trick/`

4.4 Exercises

Exercise 4.1. Let K be a compact space and $\mu_1, \mu_2 \in \mathrm{P}(K)$. Show that for $t \in [0,1]$

$$\nu\colon \mathcal{B}(K) \to [0,1], \quad A \mapsto t\mu_1(A) + (1-t)\mu_2(A)$$

is a regular Borel probability measure on K with

$$\int_K f \,\mathrm{d}\nu = t \int_K f \,\mathrm{d}\mu_1 + (1-t) \int_K f \,\mathrm{d}\mu_2$$

for all bounded Borel measurable functions $f\colon K \to \mathbb{C}$.

Exercise 4.2. For a compact space K and a regular Borel probability measure $\mu \in \mathrm{P}(K)$ the set

$$\mathrm{supp}(\mu) \coloneqq \{x \in K \mid \mu(O) > 0 \text{ for every open neighborhood } O \text{ of } x\}$$

is called the **support of** μ.

(i) Show that $\mathrm{supp}(\mu)$ is a closed subset of K.

(ii) Show that if $A \subseteq K$ is closed with $A \cap \mathrm{supp}(\mu) = \emptyset$, then $\mu(A) = 0$.

(iii) Use (ii) to conclude that $\mu(\mathrm{supp}(\mu)) = 1$. In particular, $\mathrm{supp}(\mu)$ is non-empty.

(iv) Show that $\mathrm{supp}(\mu)$ is the smallest closed subset $L \subseteq K$ with $\mu(L) = 1$.

(v) Show that for a point $x \in K$ we have $\mu = \delta_x$ precisely when $\mathrm{supp}(\mu) = \{x\}$.

(vi) Assume now that (K, τ) is a topological dynamical system and $\mu \in \mathrm{P}(K, \tau)$. Show that $\tau_\gamma(\mathrm{supp}(\mu)) = \mathrm{supp}(\mu)$ for every $\gamma \in \Gamma$.

Exercise 4.3. Let K be a compact space. Show that the extreme points of the weak* compact and convex subset $\mathrm{P}(K) \subseteq \mathrm{C}(K)'$ are precisely the point measures δ_x for $x \in K$.
Hint: Use Proposition 4.1.6 as well as Exercise 4.2.

Exercise 4.4. For $\Gamma = \mathbb{Z}$ consider the topological dynamical system $([0,1], \tau)$ given by $\tau\colon \mathbb{Z} \to \mathrm{Homeo}([0,1])$, $n \mapsto \sigma^n$ where $\sigma(x) \coloneqq x^2$ for $x \in [0,1]$. Describe all invariant measures $\mu \in \mathrm{P}([0,1], \tau)$. Which of these are ergodic?

Exercise 4.5. Prove Lemma 4.2.5.

Exercise 4.6. Show that for all $r, s \in [0,1]$ with $r \leq s$ there is a subset $A \subseteq \mathbb{N}$ with $\underline{\mathrm{d}}(A) = r$ and $\overline{\mathrm{d}}(A) = s$.
Hint: We may assume $0 < r \leq s < 1$. Set $M_0 \coloneqq 0$. Then recursively construct sequences $(N_k)_{k\in\mathbb{N}}$ and $(M_k)_{k\in\mathbb{N}}$. If $N_1, \dots, N_{k-1}, M_0, \dots, M_{k-1}$ have already been

constructed for some $k \in \mathbb{N}$, let N_k be the smallest $N > M_{k-1}$ with

$$\frac{1}{N}\left(\sum_{j=1}^{k-1}(N_j - M_{j-1}) + (N - M_{k-1})\right) > s,$$

and let M_k be the smallest $M > N_k$ with

$$\frac{1}{M}\sum_{j=1}^{k}(N_j - M_{j-1}) < r.$$

Then set $A := \bigcup_{j=1}^{\infty}\{M_{j-1}+1, \ldots, N_j\}$.

Exercise 4.7. A subset $A \subseteq \mathbb{N}$ is **syndetic** (or has **bounded gaps**) if there is $N \in \mathbb{N}$ such that $\{n, \ldots, n+N\} \cap A \neq \emptyset$ for every $n \in \mathbb{N}$.

(i) Show that if a subset $A \subseteq \mathbb{N}$ is syndetic, then $\underline{\mathrm{d}}(A) > 0$.

(ii) Find an example of a subset $A \subseteq \mathbb{N}$ with $\underline{\mathrm{d}}(A) > 0$ which is not syndetic.

Exercise 4.8. Let $\mathbb{P} \subseteq \mathbb{N}$ be the set of all prime numbers. Let further $\pi(n) := |\mathbb{P} \cap \{1, \ldots, n\}|$ for $n \in \mathbb{N}$ be the number of primes in $\{1, \ldots, n\}$. The prime number theorem (see, e.g., [Jam03, Section 1.1]) asserts that

$$\lim_{n\to\infty} \pi(n) \cdot \frac{\log(n)}{n} = 1.$$

(i) Use the prime number theorem to show that $\overline{\mathrm{d}}(\mathbb{P}) = 0$.

(ii) Show that $n^{\pi(2n)-\pi(n)} \leq \binom{2n}{n} \leq 4^n$ for every $n \in \mathbb{N}$.
Hint: All primes in $\{n+1, \ldots, 2n\}$ appear in the prime factorization of the number $\binom{2n}{n}$.

(iii) Use part (ii) to prove that $\pi(2^{n+1}) - \pi(2^n) \leq \frac{2^{n+1}}{n}$ for every $n \in \mathbb{N}$.

(iv) Use part (iii) to prove that $\pi(2^{2n}) - \pi(2) \leq 2^{n+1} + \frac{2^{2n+1}}{n}$ for every $n \in \mathbb{N}$.

(v) Use part (iv) to show that $\overline{\mathrm{d}}(\mathbb{P}) = 0$ without using the prime number theorem.

Lecture 5

In this chapter we prove a version of a famous decomposition result due to Jacobs, de Leeuw and Glicksberg, which has numerous applications in operator and ergodic theory. As a tool for its proof we investigate tensor products of Hilbert spaces and discuss their relations to so-called Hilbert–Schmidt operators.

5.1 The Discrete Spectrum Part

Assume that $U\colon G \to \mathscr{U}(H)$ is a unitary representation of a group G on a Hilbert space H (see Definition 3.1.12). If we apply the Abstract Mean Ergodic Theorem 3.1.5 to the image $U(G)$, we obtain an orthogonal decomposition of H into a subspace on which U acts in a very simple way (the fixed space $\mathrm{fix}(U(G))$) and a subspace on which $U(G)$ "goes to zero" in some sense (the closed linear hull $\overline{\mathrm{lin}} \bigcup_{x\in G}(\mathrm{Id}_H - U_x)(H)$).

We prove a second splitting result for U and begin with the "structured part" of the decomposition. Instead of vectors which remain fixed by the representation U, we now consider vectors $f \in H$ which "do not move too far" in the sense that the orbit $\{U_x f \mid x \in G\}$ is contained in a finite-dimensional subspace, or, equivalently, the linear hull $M_f := \mathrm{lin}\{U_x f \mid x \in G\}$ is finite-dimensional. If we call a subset $M \subseteq H$ **invariant** whenever $U_x f \in M$ for all $f \in M$ and $x \in G$, then each such M_f is an invariant finite-dimensional subspace. This motivates the following definition.

Definition 5.1.1. Let $U\colon G \to \mathscr{U}(H)$ be a unitary representation of a group G. Then the closure

$$H_{\mathrm{ds}} := \overline{\bigcup\{M \subseteq H \mid M \text{ invariant finite-dimensional subspace}\}} \subseteq H$$

is called the **discrete spectrum part** of U.

Notice that the sum $M_1 + M_2$ of invariant finite-dimensional subspaces M_1 and M_2 of a unitary representation $U\colon G \to \mathscr{U}(H)$ is again an invariant finite-dimensional subspace, and thus H_{ds} is always a (closed) linear subspace of H.

We introduce the following concept of "minimal" invariant finite-dimensional subspaces.

Definition 5.1.2. For a unitary representation $U\colon G \to \mathscr{U}(H)$ of a group G, we call an invariant finite-dimensional subspace $M \subseteq H$ **irreducible** if $\{0\}$ and M are the only invariant linear subspaces contained in M.

Every invariant finite-dimensional subspace splits into irreducible ones:

Proposition 5.1.3. *For every invariant finite-dimensional subspace M of a unitary representation $U\colon G \to \mathscr{U}(H)$ of a group G there are $d \in \mathbb{N}$ and irreducible invariant finite-dimensional subspaces $M_1, \dots, M_d \subseteq H$ such that $M = M_1 \oplus \cdots \oplus M_d$ orthogonally.*

Using the following lemma, the result follows readily by induction on the dimension of M (see Exercise 5.1).

Lemma 5.1.4. *Let $U\colon G \to \mathscr{U}(H)$ be a unitary representation of a group G. If $M \subseteq H$ is an invariant linear subspace, then so is the orthogonal complement $M^\perp$.*

Proof. Let $f \in M^\perp$ and $g \in M$. For $x \in G$ we have $(U_x f|g) = (f|U_{x^{-1}}g) = 0$ since $U_{x^{-1}}$ is unitary and M is invariant. But this means $U_x f \in M^\perp$. □

We obtain the following consequence of Proposition 5.1.3.

Corollary 5.1.5. *Let $U\colon G \to \mathscr{U}(H)$ be a unitary representation of a group G. Then*

$$H_{\mathrm{ds}} = \overline{\mathrm{lin}} \bigcup \{M \subseteq H \mid M \text{ irreducible invariant finite-dimensional subspace}\}.$$

For abelian groups (in particular, for our fixed abelian group Γ) this leads to a particularly simple description of the discrete spectrum part. Write $\mathbb{T} := \{z \in \mathbb{C} \mid |z| = 1\}$ for the multiplicative group of complex numbers of modulus one, and call this the **torus**.

Definition 5.1.6. For an abelian group G a group homomorphism $\chi\colon G \to \mathbb{T}$ is called a **character**. The **dual group** G^* of G is the set of all such characters equipped with the multiplication given by $(\chi_1\chi_2)(x) := \chi_1(x)\chi_2(x)$ for $x \in G$ and $\chi_1, \chi_2 \in G^*$.

Observe that the dual group G^* of an abelian group G is indeed an (abelian) group with the neutral element $\mathbb{1}\colon G \to \mathbb{T}$, $x \mapsto 1$ and inverse $\overline{\chi}\colon G \to \mathbb{T}$, $x \mapsto \overline{\chi(x)} = \chi(x)^{-1}$ for $\chi \in G^*$.

Example 5.1.7. For $G = \mathbb{Z}$ every $z \in \mathbb{T}$ defines a character $\chi_z\colon \mathbb{Z} \to \mathbb{T}$, $m \mapsto z^m$. One can check that $\mathbb{T} \to \mathbb{Z}^*$, $z \mapsto \chi_z$ is a group isomorphism. See Exercise 5.3 for this and further examples of representing dual groups.

We now extend the classical concept of eigenvectors of matrices and linear operators to unitary representations of abelian groups.

Definition 5.1.8. Let $U\colon G \to \mathscr{U}(H)$ be a unitary representation of an abelian group G. For a character $\chi \in G^*$ we call

$$\ker(\chi - U) \coloneqq \{f \in H \mid U_x f = \chi(x) f \text{ for every } x \in G\} = \bigcap_{x\in G} \ker(\chi(x)\mathrm{Id}_H - U_x)$$

the **eigenspace** of U with respect to χ, and elements $f \in \ker(\chi - U) \setminus \{0\}$ are called **eigenvectors** with respect to χ. The set $\sigma_{\mathrm{p}}(U) \coloneqq \{\chi \in G^* \mid \ker(\chi - U) \neq \{0\}\}$ is the **point spectrum** of U and its elements are called the **eigenvalues** of U.

Remark 5.1.9. Take a unitary operator $V \in \mathscr{U}(H)$ on a Hilbert space H. If $U\colon \mathbb{Z} \to \mathscr{U}(H)$, $m \mapsto V^m$ is the induced unitary representation of $G = \mathbb{Z}$, then, with the isomorphism $\mathbb{T} \to \mathbb{Z}^*$, $z \mapsto \chi_z$ from Example 5.1.7, $\ker(\chi_z - U) = \ker(z\mathrm{Id}_H - V)$ for every $z \in \mathbb{T}$. In this way, Definition 5.1.8 extends the known spectral theoretic concepts of eigenvectors, eigenvalues and eigenspaces from linear algebra and functional analysis.

Eigenspaces of unitary representations of abelian groups have the following properties.

Proposition 5.1.10. *Let $U\colon G \to \mathscr{U}(H)$ be a unitary representation of an abelian group G. Then the following assertions hold.*

(i) *The eigenspaces $\ker(\chi - U)$ for $\chi \in G^*$ are pairwise orthogonal.*

(ii) *For $M \subseteq H$ the following assertions are equivalent.*

  (a) *M is an irreducible invariant finite-dimensional subspace.*

  (b) *M is an invariant linear subspace which is at most one-dimensional.*

  (c) *There is $\chi \in G^*$ and $f \in \ker(\chi - U)$ such that $M = \mathbb{C} \cdot f$.*

Proof. For part (i) take $\chi_1, \chi_2 \in G^*$ with $\chi_1 \neq \chi_2$. Choose $x \in G$ with $\chi_1(x) \neq \chi_2(x)$. If $f \in \ker(\chi_1 - U)$ and $g \in \ker(\chi_2 - U)$, then

$$\chi_1(x)\overline{\chi_2}(x)(f|g) = (\chi_1(x)f|\chi_2(x)g) = (U_x f|U_x g) = (f|g).$$

Since $\chi_1(x)\overline{\chi_2}(x) \neq 1$, we obtain $(f|g) = 0$.

We now prove part (ii). The implications "(b) $\Rightarrow$ (a)" and "(c) $\Rightarrow$ (b)" are obvious. For "(a) $\Rightarrow$ (c)" take an irreducible invariant finite-dimensional subspace $M \subseteq H$. We may assume that $M \neq \{0\}$. For $x \in G$ the restriction $U_x|_M\colon M \to M$ of U_x to M is a unitary map on a finite-dimensional Hilbert space and thus has, by linear algebra, an eigenvector with respect to some eigenvalue $\chi(x) \in \mathbb{T}$, i.e., $\ker(\chi(x)\mathrm{Id}_M - U_x|_M) \neq$

$\{0\}$. Now if $f \in \ker(\chi(x)\mathrm{Id}_M - U_x|_M)$ and $y \in G$, then, since G is abelian[1],

$$U_x U_y f = U_{xy} f = U_{yx} f = U_y U_x f = U_y(\chi(x) f) = \chi(x) U_y f,$$

hence $U_y f \in \ker(\chi(x)\mathrm{Id}_M - U_x|_M)$. Thus, $\ker(\chi(x)\mathrm{Id}_M - U_x|_M)$ is an invariant subspace of H contained in M. Since M is irreducible, we obtain $M = \ker(\chi(x)\mathrm{Id}_M - U_x|_M)$, i.e., $U_x f = \chi(x) f$ for every $f \in M$.

Since U is a group homomorphism and there is some $f \in M \setminus \{0\}$, the map $\chi\colon G \to \mathbb{T}$, $x \mapsto \chi(x)$ is a character. By definition of χ we have $M \subseteq \ker(\chi - U)$. In particular, if we pick $f \in M \setminus \{0\}$, then $\mathbb{C} \cdot f$ is an invariant subspace contained in M. Thus, $M = \mathbb{C} \cdot f$ since M is irreducible. □

Combined with Corollary 5.1.5 we obtain:

Corollary 5.1.11. *Let $U\colon G \to \mathscr{U}(H)$ be a unitary representation of an abelian group G. Then $H_{\mathrm{ds}} = \overline{\mathrm{lin}} \bigcup_{\chi \in G^*} \ker(\chi - U)$.*

5.2 Interlude: Standard Constructions for Hilbert Spaces

Before we identify the orthogonal complement of the discrete spectrum part of a unitary representation, let us first review some standard constructions from Hilbert space theory.

Completions. Recall that we can "complete" any metric space (X, d_X) in the following way (see, e.g., [SV06, Section 1.5]): Let $\mathrm{CS}(X)$ be the set of all Cauchy sequences $(x_n)_{n\in\mathbb{N}}$ in X, and write $(x_n)_{n\in\mathbb{N}} \sim (y_n)_{n\in\mathbb{N}}$ for $(x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}} \in \mathrm{CS}(X)$ if $\lim_{n\to\infty} d_X(x_n, y_n) = 0$. Then $\sim$ is an equivalence relation on $\mathrm{CS}(X)$, and

$$d_{\overline{X}}\colon \overline{X} \times \overline{X} \to [0,\infty), \quad ([(x_n)_{n\in\mathbb{N}}], [(y_n)_{n\in\mathbb{N}}]) \mapsto \lim_{n\to\infty} \mathrm{d}_X(x_n, y_n)$$

defines a complete metric on the quotient space $\overline{X} := \mathrm{CS}(X)/{\sim}$. Mapping every $x \in X$ to the equivalence class of the constant sequence $(x, x, x, \dots) \in \mathrm{CS}(X)$ then yields an isometric (and, in particular, injective) map $i_X\colon X \to \overline{X}$, and it is common to identify x with its image $i_X(x) \in \overline{X}$. The space $(\overline{X}, d_{\overline{X}})$ is called the **completion** of (X, d_X).

Now if we start from an inner product $(\cdot|\cdot)$ on a vector space V and consider the induced metric given by $d_V(f,g) := \|f - g\| = \sqrt{(f-g|f-g)}$ for $f, g \in V$. Then

[1] Here we use multiplicative notation for the abelian group G (while our fixed group Γ is written additively).

one can check that the operations

$$\overline{V} \times \overline{V} \to \overline{V}, \quad ([(x_n)_{n\in\mathbb{N}}], [(y_n)_{n\in\mathbb{N}}]) \mapsto [(f_n + g_n)_{n\in\mathbb{N}}],$$
$$\mathbb{C} \times \overline{V} \to \overline{V}, \quad (c, [(f_n)_{n\in\mathbb{N}}]) \mapsto [(c \cdot f_n)_{n\in\mathbb{N}}]$$

turn $\overline{V}$ into a vector space, and

$$(\cdot|\cdot)\colon \overline{V} \times \overline{V} \to \mathbb{C}, \quad ([(f_n)_{n\in\mathbb{N}}], [(g_n)_{n\in\mathbb{N}}]) \mapsto \lim_{n\to\infty} (f_n|g_n)$$

is an inner product on $\overline{V}$ which induces the metric $d_{\overline{V}}$ (see, e.g., [Wei80, Theorem 4.11]). Thus, we can view V as a dense subspace of the Hilbert space $\overline{V}$. We call $\overline{V}$ the **completion** of V respect to the inner product $(\cdot|\cdot)$.

Dual Spaces. Recall that for any Hilbert space H, by the Riesz–Fréchet representation theorem (see Theorem A.2.3), the map $H \to H'$, $g \mapsto \overline{g}$ to its dual space H' given by $\overline{g}(f) \coloneqq (f|g)$ for $f, g \in H$ is a bijection. By setting $(\overline{f}|\overline{g}) \coloneqq (g|f)$ for $f, g \in H$ we then turn H' into a Hilbert space, which we call the **dual Hilbert space** of H[2].

We further note that any linear isometry $U\colon H \to K$ between Hilbert spaces also defines a linear isometry $\overline{U}\colon H' \to K'$, $\overline{f} \mapsto \overline{Uf}$ between the corresponding dual Hilbert spaces.

Tensor Products. Recall that for any (complex) vector spaces E_1 and E_2 we can consider their vector space tensor product $E_1 \otimes_{\mathrm{vect}} E_2$, see, e.g., [Hun74, Section IV.5] or [Lan02, Chapter 16]. Its elements are written as (finite) sums of simple tensors $f \otimes g$ for $f \in E_1$ and $g \in E_2$. The key feature of the tensor product is "linearizing" bilinear maps:

Proposition 5.2.1. *Let E_1, E_2 be vector spaces. Whenever $b\colon E_1 \times E_2 \to F$ is a bilinear map to a vector space F, then there is a unique linear map $l_b\colon E_1 \otimes_{\mathrm{vect}} E_2 \to F$ with $l_b(f \otimes g) = b(f, g)$ for all $f \in E_1$ and $g \in E_2$.*

We introduce a Hilbert space version of the tensor product using the following observation (see, e.g., [Wei80, Section 3.4]).

Proposition 5.2.2. *Let H_1 and H_2 be Hilbert spaces. Then there is a unique inner product $(\cdot|\cdot)$ on $H_1 \otimes_{\mathrm{vect}} H_2$ with $(f_1 \otimes f_2|g_1 \otimes g_2) = (f_1|g_1) \cdot (f_2|g_2)$ for all $f_1, g_1 \in H_1$ and $f_2, g_2 \in H_2$.*

Definition 5.2.3. For Hilbert spaces H_1 and H_2 their **Hilbert space tensor product** $H_1 \otimes H_2$ is the completion of the vector space tensor product $H_1 \otimes_{\mathrm{vect}} H_2$ with respect to the inner product from Proposition 5.2.2.

[2]Note that, since we work over the field of complex numbers here, the map $H \to H'$, $g \mapsto \overline{g}$ is antilinear, so we cannot directly identify H with H' as Hilbert spaces.

We need the following easy consequence of Propositions 5.2.1 and A.1.1.

Proposition 5.2.4. *Let $U_1\colon H_1 \to K_1$ and $U_2\colon H_2 \to K_2$ be linear isometries between Hilbert spaces. Then there is a unique linear isometry $U_1 \otimes U_2\colon H_1 \otimes H_2 \to K_1 \otimes K_2$ with $(U_1 \otimes U_2)(f_1 \otimes f_2) = U_1 f_1 \otimes U_2 f_2$ for all $f_1 \in H_1$ and $f_2 \in H_2$.*

5.3 The JdLG-Decomposition

With the previous constructions we now can consider for any group G

(i) the **dual representation**

$$\overline{U}\colon G \to \mathscr{U}(H'), \quad x \mapsto \overline{U_x}$$

of a unitary representation $U\colon G \to \mathscr{U}(H)$, and

(ii) the **tensor product representation**

$$U_1 \otimes U_2\colon G \to \mathscr{U}(H_1 \otimes H_2), \quad x \mapsto (U_1)_x \otimes (U_2)_x$$

of unitary representations $U_i\colon G \to \mathscr{U}(H_i)$ for $i = 1, 2$.

In particular, given any unitary representation $U\colon G \to \mathscr{U}(H)$ we can consider the tensor product $U \otimes \overline{U}\colon G \to \mathscr{U}(H \otimes H')$ with its dual representation. We use this to describe the orthogonal complement of the discrete spectrum part H_{ds}.

Let us assume, for the moment, that G is abelian. If $\chi \in G^*$ is a character and f is an element of the eigenspace $\ker(\chi - U)$, then

$$(U_x \otimes \overline{U_x})(f \otimes \overline{f}) = U_x f \otimes \overline{U_x f} = \chi(x) f \otimes (\overline{\chi(x) f}) = \chi(x) \cdot \overline{\chi(x)} \cdot f \otimes \overline{f} = f \otimes \overline{f},$$

for every $x \in G$ since $|\chi(x)|^2 = 1$. Thus, $f \otimes \overline{f} \in \mathrm{fix}(U \otimes \overline{U})$. We can therefore construct fixed vectors in the tensor product $H \otimes H'$ from eigenvectors in H!

If G is non-abelian, then the same idea still applies, but is slightly more complicated to write down. Assume that $\{e_1, \dots, e_n\}$ is an orthonormal basis (short: ONB) of an invariant finite-dimensional subspace $M \subseteq H$. For every $x \in G$ and $f \in M$ we can then write $U_x f = \sum_{i=1}^m (U_x f | e_i) e_i$, and thus

$$\begin{aligned}(U_x \otimes \overline{U_x}) \sum_{i=1}^n e_i \otimes \overline{e_i} &= \sum_{i=1}^n U_x e_i \otimes \overline{U_x e_i} = \sum_{i=1}^n \Big(\sum_{j=1}^n (U_x e_i | e_j) e_j\Big) \otimes \Big(\sum_{k=1}^n \overline{(U_x e_i | e_k)}\, \overline{e_k}\Big) \\ &= \sum_{j=1}^n \sum_{k=1}^n \Big(\sum_{i=1}^n (U_x e_i | e_j) \overline{(U_x e_i | e_k)}\Big) e_j \otimes \overline{e_k}.\end{aligned}$$

Since U_x is unitary, we obtain that the vectors $U_xe_1, \dots, U_xe_n$ also form an orthonormal basis of M, hence

$$\sum_{i=1}^{n}(U_xe_i|e_j)\overline{(U_xe_i|e_k)} = \left(\sum_{i=1}^{n}(e_k|U_xe_i)U_xe_i\middle|e_j\right) = (e_k|e_j) = \delta_{kj}.$$

This shows $(U_x \otimes \overline{U_x})\sum_{i=1}^{n} e_i \otimes \overline{e_i} = \sum_{i=1}^{n} e_i \otimes \overline{e_i}$ for every $x \in G$, and therefore $\sum_{i=1}^{n} e_i \otimes \overline{e_i} \in \operatorname{fix}(U \otimes \overline{U})$. In this way, orthonormal bases of invariant finite-dimensional subspaces of H give rise to fixed vectors in the tensor product $H \otimes H'$.

The key insight of this lecture is that these elements already generate the entire fixed space:

Theorem 5.3.1 (Key Lemma)**.** *Let $U\colon G \to \mathscr{U}(H)$ be a unitary representation of a group G. Then the linear hull*

$$\operatorname{lin}\left\{\sum_{i=1}^{n} e_i \otimes \overline{e_i} \,\middle|\, \{e_1, \dots e_n\} \subseteq H \text{ ONB of an invariant finite-dimensional subspace}\right\}$$

is dense in $\operatorname{fix}(U \otimes \overline{U})$.

We will prove the Key Lemma at the end of this lecture. For now, let us introduce the set of vectors which "do not contribute" to the fixed space of the tensor product.

Definition 5.3.2. Let $U\colon G \to \mathscr{U}(H)$ be a a unitary representation of a group G on a Hilbert space H. Then $H_{\mathrm{wm}} := \{f \in H \mid f \otimes \overline{f} \in \operatorname{fix}(U \otimes \overline{U})^{\perp}\}$ is the **weakly mixing part** of U.

If the group G has a Følner net (e.g., if it is abelian), then we obtain a more concrete description of the weakly mixing part.

Proposition 5.3.3. *Let $U\colon G \to \mathscr{U}(H)$ be a unitary representation of a group G with Følner net $(F_i)_{i\in I}$. Then*

$$\begin{aligned} H_{\mathrm{wm}} &= \left\{f \in H \,\middle|\, \lim_{i\in I} \sup_{\substack{g\in H\\ \|g\|\leq 1}} \frac{1}{|F_i|}\sum_{x\in F_i} |(U_xf|g)|^2 = 0\right\} \\ &= \left\{f \in H \,\middle|\, \lim_{i\in I} \frac{1}{|F_i|}\sum_{x\in F_i} |(U_xf|f)|^2 = 0\right\}. \end{aligned}$$

Proof. Write $P \in \mathscr{L}(H \otimes H')$ for the orthogonal projection onto $\operatorname{fix}(U \otimes \overline{U})$. For every $f \in H$ we obtain that

$$\lim_{i\in I} {}^{1}\!/\!{}_{|F_i|} \sum_{x\in F_i} U_xf \otimes \overline{U_xf} = P(f \otimes \overline{f})$$

by Proposition 3.1.14 and Theorem 3.1.15. Denote the sets on the above right-hand side by H_1 and H_2 (from top to bottom). For $f \in H$ and $g \in H$ with $\|g\| \leq 1$ we obtain

$$\frac{1}{|F_i|}\sum_{x\in F_i}|(U_x f|g)|^2 = \left(\frac{1}{|F_i|}\sum_{x\in F_i}U_x f\otimes\overline{U_x f}\,\middle|\, g\otimes\overline{g}\right) \leq \left\|\frac{1}{|F_i|}\sum_{x\in F_i}U_x f\otimes\overline{U_x f}\right\|$$

by the Cauchy–Schwarz inequality for every $i \in I$. This implies $H_{\mathrm{wm}} \subseteq H_1$. The inclusions $H_1 \subseteq H_2$ is obvious. Finally, if $f \in H_2$, then

$$\|P(f\otimes\overline{f})\|^2 = (P(f\otimes\overline{f})|f\otimes\overline{f}) = \lim_{i\in I}\frac{1}{|F_i|}\sum_{x\in F_i}(U_x f\otimes\overline{U_x f}|f\otimes\overline{f}) = 0.$$

□

We now prove our desired decomposition (assuming the Key Lemma), which is a version of a splitting result of Konrad Jacobs, Karl de Leeuw, and Irving Glicksberg. It will be of crucial importance in the next two lectures, but is also applied in Exercise 5.5 below to prove “weighted mean ergodic theorems”.

Theorem 5.3.4 (JdLG-decomposition)**.** *For every unitary representation $U\colon G \to \mathscr{U}(H)$ of a group G we have an orthogonal decomposition $H = H_{\mathrm{ds}} \oplus H_{\mathrm{wm}}$ into the invariant closed linear subspaces H_{ds} and H_{wm}.*

Proof. For $f \in H$ we have $f \in H_{\mathrm{ds}}^{\perp}$ precisely when $\sum_{i=1}^n |(f|e_i)|^2 = 0$ for each orthonormal basis $\{e_1, \dots, e_n\}$ of an invariant, finite-dimensional subspace $M \subseteq H$. But this means that $(f\otimes\overline{f}|\sum_{i=1}^n e_i\otimes\overline{e_i}) = 0$ for each such orthonormal basis. By Theorem 5.3.1 this is the case precisely when $f\otimes\overline{f} \in \mathrm{fix}(U\otimes\overline{U})^{\perp}$. □

5.4 Hilbert–Schmidt Operators

To prove the Key Lemma (Theorem 5.3.1), we take a different perspective on the tensor product $H \otimes H'$ of a Hilbert space H.

Definition 5.4.1. For a Hilbert space H call a linear map $A\colon H \to H$ a **Hilbert–Schmidt operator** if

$$\|A\|_{\mathrm{HS}} := \sup\left\{\left(\sum_{j=1}^n \|Ae_j\|^2\right)^{\frac{1}{2}}\,\middle|\,\{e_1,\dots,e_n\}\text{ orthonormal in }H\right\} < \infty.$$

Denote the set of Hilbert–Schmidt operators on H by $\mathrm{HS}(H)$.

We start with some elementary observations.

Lemma 5.4.2. *Let $A\colon H \to H$ be a linear map on a Hilbert space H.*

(i) *If $A \in \mathrm{HS}(H)$, then A is also a bounded operator with $\|A\| \le \|A\|_{\mathrm{HS}}$.*

(ii) *If $E \subseteq H$ is an orthonormal basis of H, then*

$$\|A\|_{\mathrm{HS}}^2 = \sum_{e\in E} \|Ae\|^2 := \sup\Big\{ \sum_{e\in F} \|Ae\|^2 \mid F \subseteq E \text{ finite} \Big\}.$$

(iii) *$A \in \mathrm{HS}(H)$ precisely when $A^* \in \mathrm{HS}(H)$. In this case $\|A\|_{\mathrm{HS}} = \|A^*\|_{\mathrm{HS}}$.*

Proof. For part (i) take $f \in H \setminus \{0\}$. Then the singleton set $\{f/\|f\|\}$ is orthonormal, hence $\|A(f/\|f\|)\| \le \|A\|_{\mathrm{HS}}$, and consequently $\|Af\| \le \|A\|_{\mathrm{HS}} \cdot \|f\|$. This implies (i).

For parts (ii) and (iii) let $E \subseteq H$ be an orthonormal basis of H and take a finite orthonormal subset $F \subseteq F$. By Parseval's identity (see Theorem A.2.5) we obtain

$$\sum_{f\in F} \|A^* f\|^2 = \sum_{f\in F}\sum_{e\in E} |(A^* f|e)|^2 = \sum_{e\in E}\sum_{f\in F} |(Ae|f)|^2 \le \sum_{e\in E} \|Ae\|^2$$

where the last inequality follows from Bessel's inequality (see A.2.6). This implies $\|A^*\|_{\mathrm{HS}}^2 \le \sum_{e\in E} \|Ae\|^2 \le \|A\|_{\mathrm{HS}}^2$. Since $(A^*)^* = A$, we conclude that

$$\|A\|_{\mathrm{HS}}^2 \le \sum_{e\in E} \|A^* e\|^2 \le \|A^*\|_{\mathrm{HS}}^2 \le \sum_{e\in E} \|Ae\|^2 \le \|A\|_{\mathrm{HS}}^2,$$

and hence all these inequalities are actually equalities. This shows (ii) and (iii). □

We need the following observation.

Proposition 5.4.3. *Let H be a Hilbert space. Then $\mathrm{HS}(H)$ is a linear subspace of $\mathscr{L}(H)$ and a Banach space when equipped with $\|\cdot\|_{\mathrm{HS}}$.*

Proof. Clearly, the zero operator is an element of $\mathrm{HS}(H)$, and it is easy to see that $cA \in \mathrm{HS}(H)$ with $\|cA\|_{\mathrm{HS}} = |c| \cdot \|A\|_{\mathrm{HS}}$ for each $c \in \mathbb{C}$ and $A \in \mathrm{HS}(H)$. Now if $A, B \in \mathrm{HS}(H)$, take a finite orthonormal subset $E = \{e_1, \ldots, e_n\} \subseteq H$. Then

$$\sum_{i=1}^n \|(A+B)e_i\|^2 = \sum_{i=1}^n \|Ae_i\|^2 + \sum_{i=1}^n 2 \cdot \operatorname{Re}(Ae_i|Be_i) + \sum_{i=1}^n \|Be_i\|^2$$

By the "Pythagorean theorem" and the Cauchy–Schwarz inequalities for H and $\mathbb{C}^n$ we thus obtain

$$\begin{aligned}\sum_{i=1}^n \|(A+B)e_i\|^2 &\le \sum_{i=1}^n \|Ae_i\|^2 + 2 \cdot \Big(\sum_{i=1}^n \|Ae_i\|^2\Big)^{\frac12} \Big(\sum_{i=1}^n \|Be_i\|^2\Big)^{\frac12} + \sum_{i=1}^n \|Be_i\|^2 \\ &= \bigg(\Big(\sum_{i=1}^n \|Ae_i\|^2\Big)^{\frac12} + \Big(\sum_{i=1}^n \|Be_i\|^2\Big)^{\frac12}\bigg)^2 \le (\|A\|_{\mathrm{HS}} + \|B\|_{\mathrm{HS}})^2.\end{aligned}$$

Thus, $A+B \in \mathrm{HS}(H)$ with $\|A+B\|_{\mathrm{HS}} \le \|A\|_{\mathrm{HS}} + \|B\|_{\mathrm{HS}}$. This shows that $\mathrm{HS}(H)$ is a linear subspace of $\mathscr{L}(H)$ and $\|\cdot\|_{\mathrm{HS}}$ is a seminorm on $\mathrm{HS}(H)$. By Lemma 5.4.2 (i) it is even a norm (since the operator norm is one).

To see that it is complete, take a Cauchy sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathrm{HS}(H)$ with respect to $\|\cdot\|_{\mathrm{HS}}$. By Lemma 5.4.2 (i) it is also a Cauchy sequence with respect to the operator norm $\|\cdot\|$, and thus there is a unique $A \in \mathscr{L}(H)$ with $\lim_{n\to\infty} \|A_n - A\| = 0$. For $\varepsilon > 0$ choose $n_0 \in \mathbb{N}$ such that $\|A_n - A_m\|_{\mathrm{HS}}^2 \le \varepsilon$ for all $n, m \ge n_0$. If $E \subseteq H$ is a finite orthonormal subset, we obtain

$$\sum_{e\in E} \|Ae - A_n e\|^2 = \lim_{m\to\infty} \sum_{e\in E} \|A_m e - A_n e\|^2 \le \limsup_{m\to\infty} \|A_n - A_m\|_{\mathrm{HS}}^2 \le \varepsilon$$

for every $n \ge n_0$. This readily implies $A \in \mathrm{HS}(H)$ with $\lim_{n\to\infty} \|A - A_n\|_{\mathrm{HS}} = 0$. □

One can even turn $\mathrm{HS}(H)$ into a Hilbert space in a canonical way (see Exercise 5.7).

The following are simple examples of Hilbert–Schmidt operators. More interesting ones are discussed in Exercise 5.6 below.

Example 5.4.4. Let H be a Hilbert space. For $g, h \in H$ consider the "rank-one" linear map $A_{g,h}\colon H \to H$ given by $A_{g,h}(f) := (f|h)g$ for $f \in H$. Then a short computation using the Fourier expansion (see Theorem A.2.5) shows that the map $A_{g,h}$ is a Hilbert–Schmidt operator with $\|A_{g,h}\|_{\mathrm{HS}} = \|g\| \cdot \|h\|$.

If $B\colon H \to H$ is any operator with finite-dimensional range and $\{e_1, \dots, e_n\}$ is an orthonormal basis for its range, then we can write

$$Bf = \sum_{i=1}^{n} (Bf|e_i)e_i = \sum_{i=1}^{n} A_{e_i, B^* e_i} f$$

for every $f \in H$, and hence B is also a Hilbert–Schmidt operator.

With the notation for rank-one operators from Example 5.4.4 we obtain the following identification.

Proposition 5.4.5. *For a Hilbert space H there is a unique isometric linear bijection $\Phi\colon H \otimes H' \to \mathrm{HS}(H)$ with $\Phi(g \otimes \overline{h}) = A_{g,h}$ for all $g, h \in H$.*

We first prove the following auxiliary result.

Lemma 5.4.6. *Let H be a Hilbert space. Then the linear hull* $\mathrm{lin}\{A_{g,h} \mid g, h \in H\}$ *is dense in* $\mathrm{HS}(H)$.

Proof. Pick an orthonormal basis $E \subseteq H$ of H and let $A \in \mathrm{HS}(H)$. Using Lemma 5.4.2 (ii) we choose for a given $\varepsilon > 0$ some finite orthonormal set $F \subseteq E$ with

$\|A\|_{\mathrm{HS}}^2 \leq \sum_{e\in F} \|Ae\|^2 + \varepsilon$ and consider $B := \sum_{e\in F} A_{Ae,e}$. For every finite subset $C \subseteq E$ we then obtain

$$\begin{aligned}\sum_{f\in C} \|Bf - Af\|^2 &= \sum_{f\in C\cap F} \Big\|\sum_{e\in F}(f|e)Ae - Af\Big\|^2 + \sum_{f\in C\setminus F} \Big\|\sum_{e\in F}(f|e)Ae - Af\Big\|^2 \\ &= 0 + \sum_{f\in C\setminus F} \|Af\|^2 \leq \|A\|_{\mathrm{HS}}^2 - \sum_{f\in F} \|Af\|^2 \leq \varepsilon.\end{aligned}$$

This implies $\|B - A\|_{\mathrm{HS}}^2 \leq \varepsilon$. □

Proof of Proposition 5.4.5. Observe that the map $H \times H' \to \mathrm{HS}(H)$, $(g, \overline{h}) \mapsto A_{g,h}$ is bilinear. Thus, it induces a unique linear map $\Phi\colon H \otimes_{\mathrm{vect}} H' \to \mathrm{HS}(H)$ with $\Phi(g \otimes \overline{h}) = A_{g,h}$ for all $g, h \in H$ (see Proposition 5.2.1). Moreover, if $E \subseteq H$ is an orthonormal basis for H we obtain by Lemma 5.4.2 (ii) and Theorem A.2.5,

$$\begin{aligned}\Big\|\Phi\Big(\sum_{i=1}^n g_i \otimes \overline{h_i}\Big)\Big\|_{\mathrm{HS}}^2 &= \sum_{e\in E}\Big\|\sum_{i=1}^n (e|h_i)g_i\Big\|^2 = \sum_{i=1}^n\sum_{j=1}^n\sum_{e\in E}(e|h_i)\overline{(e|h_j)}(g_i|g_j) \\ &= \sum_{i=1}^n\sum_{j=1}^n\Big(\sum_{e\in E}(h_j|e)e\Big|h_i\Big)(g_i|g_j) = \sum_{i=1}^n\sum_{j=1}^n(h_j|h_i)(g_i|g_j) \\ &= \sum_{i=1}^n\sum_{j=1}^n(g_i\otimes\overline{h_i}|g_j\otimes\overline{h_j}) = \Big\|\sum_{i=1}^n g_i\otimes\overline{h_i}\Big\|^2\end{aligned}$$

for all $g_1, \ldots, g_n, h_1, \ldots, h_n \in H$. By Lemma A.1.1, Φ uniquely extends to a linear isometry $\Phi\colon H \otimes H' \to \mathrm{HS}(H)$. Since the image of linear isometries between Banach spaces is complete, hence closed, we obtain that the range of $\Phi(H \otimes H')$ is a closed linear subspace of $\mathrm{HS}(H)$. Since it contains all operators $A_{g,h}$ for $g, h \in H$, we conclude from Lemma 5.4.6 that Φ is surjective, and hence an isometric linear bijection. □

We use the identification from Proposition 5.4.5 to translate the Key Lemma to a statement about Hilbert–Schmidt operators. The proof will then be a consequence of the following version of the spectral theorem for self-adjoint Hilbert–Schmidt operators, which represents such an operator as a series of finite rank operators.

Theorem 5.4.7. *Let $A \in \mathrm{HS}(H)$ be a self-adjoint Hilbert–Schmidt operator on a Hilbert space H. Then there is a unique sequence $(\lambda_n)_{n\in\mathbb{N}}$ in $[0,\infty)$ and unique sequences $(P_n^+)_{n\in\mathbb{N}}$ and $(P_n^-)_{n\in\mathbb{N}}$ of orthogonal projections in $\mathscr{L}(H)$ with*

(i) *$\lambda_{n+1} \leq \lambda_n$ with equality only if $\lambda_n = 0$ for $n \in \mathbb{N}$,*

(ii) *$\lambda_n = 0$ precisely when $P_n^+ = P_n^- = 0$ for $n \in \mathbb{N}$,*

(iii) $\inf_{n\in\mathbb{N}} \lambda_n = 0$,

(iv) $P_n^+(H)$ *and* $P_n^-(H)$ *are finite-dimensional for every* $n \in \mathbb{N}$,

(v) $P_n^+(H) \perp P_m^+(H)$ *and* $P_n^-(H) \perp P_m^-(H)$ *for* $n \neq m$,

(vi) $P_n^+(H) \perp P_n^-(H)$ *for every* $n \in \mathbb{N}$,

such that $A = \sum_{n=1}^{\infty} \lambda_n (P_n^+ - P_n^-)$ *with respect to the Hilbert–Schmidt norm.*

This is a consequence of the spectral theorem for compact self-adjoint operators, which most readers of the course should be familiar with. However, we include a complete proof of Theorem 5.4.7 as a supplement to this lecture. Let us now use it to finally show the Key Lemma, and therefore finish the proof our decomposition result Theorem 5.3.4.

Proof of Theorem 5.3.1. For every $x \in G$ we obtain an isometric linear bijection $\mathcal{U}_x \colon \mathrm{HS}(H) \to \mathrm{HS}(H)$ via $\mathcal{U}_x A := U_x A U_x^{-1}$ for $A \in \mathrm{HS}(H)$ (note that $\mathcal{U}_x A$ is indeed again a Hilbert–Schmidt operator with $\|\mathcal{U}_x A\|_{\mathrm{HS}} = \|A\|_{\mathrm{HS}}$). For $g, h \in H$ we obtain

$$((\Phi \circ (U_x \otimes \overline{U_x}))g \otimes \overline{h})(f) = (f|U_x h)U_x g = (U_x^{-1} f|h)U_x g = ((\mathcal{U}_x \circ \Phi)g \otimes \overline{h})(f)$$

for every $f \in H$. By linearity and continuity, this implies that $\Phi \circ (U_x \otimes \overline{U_x}) = \mathcal{U}_x \circ \Phi$ for every $x \in G$. This shows that

$$\Phi(\mathrm{fix}(U \otimes U')) = \mathrm{fix}(\mathcal{U}) := \{A \in \mathrm{HS}(H) \mid U_x A = A U_x \text{ for every } x \in G\}.$$

Since Φ is a isometric linear bijection, the claim of the key lemma is equivalent to the following: The linear hull L of all operators $\Phi(\sum_{i=1}^{n} e_i \otimes \overline{e_i})$, where $\{e_1, \ldots, e_n\}$ is an orthonormal basis of an invariant finite-dimensional subspace of H, is dense in $\mathrm{fix}(\mathcal{U}) \subseteq \mathrm{HS}(H)$ with respect to the Hilbert–Schmidt norm $\|\cdot\|_{\mathrm{HS}}$.

So take $A \in \mathrm{fix}(\mathcal{U})$ and show that A is in the closure of L. First observe that we can write $A = \frac{A+A^*}{2} + \mathrm{i}\frac{A-A^*}{2\mathrm{i}}$ and a moment's thought reveals that $\frac{A+A^*}{2}, \frac{A-A^*}{2\mathrm{i}}$ are self-adjoint elements of $\mathrm{fix}(\mathcal{U})$. We may therefore assume that A is self-adjoint and take the "spectral decomposition" $A = \sum_{n=1}^{\infty} \lambda_n (P_n^+ - P_n^-)$ from Theorem 5.4.7. For $x \in G$ we obtain

$$A = U_x A U_x^{-1} = \sum_{n=1}^{\infty} \lambda_n (U_x P_n^+ U_x^{-1} - U_x P_n^- U_x^{-1}),$$

and uniqueness of the representation implies that $U_x P_n^+ U_x^{-1} = P_n^+$ and $U_x P_n^- U_x^{-1} = P_n^-$, hence $P_n^+, P_n^- \in \mathrm{fix}(\mathcal{U})$ for every $n \in \mathbb{N}$. In particular, the orthogonal projections P in $\mathrm{fix}(\mathcal{U})$ with finite-dimensional range span a dense linear subspace of $\mathrm{fix}(\mathcal{U})$. We may therefore reduce to the case that $A = P$ is such a projection. Since $U_x P = P U_x$ for every $x \in G$, the range $M := \mathrm{ran}(P)$ is then an invariant finite-dimensional subspace of H. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis for M. Then $\Phi(\sum_{i=1}^{n} e_i \otimes \overline{e_i}) = \sum_{i=1}^{n} A_{e_i, e_i} = P$, see Example 5.4.4. This shows the claim. □

5.5 Supplement: The Spectral Theorem for Self-Adjoint Hilbert–Schmidt Operators

We now present a rather elementary[3] proof of the spectral theorem for self-adjoint Hilbert–Schmidt operators (Theorem 5.4.7) based on [Con85, Paragraph II.5] and [EHK24, Section 4.1]. We start with some preliminary observations on self-adjoint operators.

Lemma 5.5.1. *Let $A \in \mathscr{L}(H)$ be a self-adjoint operator on a Hilbert space H.*

(i) $\ker(\alpha \cdot \mathrm{Id}_H - A) \perp \ker(\beta \cdot \mathrm{Id}_H - A)$ *for all $\alpha, \beta \in \mathbb{R}$ with $\alpha \neq \beta$.*

(ii) $\|A\| = \sup\{|(Af|f)| \mid f \in H$ *with* $\|f\| = 1\}$.

Proof. The proof of part (i) is similar to the one of Proposition 5.1.10 (i): Take $f \in \ker(\alpha \cdot \mathrm{Id}_H - A)$ and $g \in \ker(\beta \cdot \mathrm{Id}_H - A)$. Then $\alpha(f|g) = (Af|g) = (f|Ag) = \beta(f|g)$, and hence $(f|g) = 0$.

For part (ii) denote the right hand side by c. The inequality $c \leq \|A\|$ follows from the Cauchy–Schwarz inequality. A short computation (using that $A = A^*$) shows the identities

$$(A(f \pm g)|f \pm g) = (Af|f) \pm 2\mathrm{Re}(Af|g) + (Ag|g),$$

which in turn yield

$$(A(f+g)|f+g) - (A(f-g)|f-g) = 4 \cdot \mathrm{Re}(Af|g)$$

for all $f, g \in H$. The definition of c and rescaling gives us $|(Ah|h)| \leq c\|h\|^2$ for all $h \in H$. Since in a Hilbert space H the norm satisfies the "parallelogram law"

$$\|(f+g)\|^2 + \|f-g\|^2 = 2(\|f\|^2 + \|g\|^2),$$

we therefore obtain $4 \cdot \mathrm{Re}(Af|g) \leq 2c \cdot (\|f\|^2 + \|g\|^2)$ for all $f, g \in H$. Now for $f \in H$ with $\|f\| \leq 1$ and $Af \neq 0$ set $g := {}^{Af}/_{\|Af\|}$. Then $4\|Af\| = 4(Af|g) \leq 2c(1+1) = 4c$. This shows $\|A\| \leq c$. □

The following is the key ingredient in the proof of Theorem 5.4.7.

Lemma 5.5.2. *Let $A \in \mathrm{HS}(H)$ be a self-adjoint Hilbert–Schmidt operator on a Hilbert space $H \neq \{0\}$. Then* $\ker(\|A\| \cdot \mathrm{Id}_H - A) \neq \{0\}$ *or* $\ker(\|A\| \cdot \mathrm{Id}_H + A) \neq \{0\}$.

Proof. We may assume that $A \neq 0$ and, by rescaling, that $\|A\| = 1$. By Lemma 5.5.1 (ii) we find a sequence $(f_n)_{n\in\mathbb{N}}$ in H with $\|f_n\| = 1$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} |(Af_n|f_n)| = 1$. Since A is self-adjoint, $(Af_n|f_n) = \overline{(Af_n|f_n)} \in \mathbb{R}$ for

[3] With some knowledge of spectral theory shorter proofs are available.

every $n \in \mathbb{N}$, and thus we may assume (after passing to a subsequence) that $\lim_{n\to\infty}(Af_n|f_n) = 1$ or $\lim_{n\to\infty}(Af_n|f_n) = -1$. We only treat the first case (the second one is similar). Since

$$\|f_n - Af_n\|^2 = \|f_n\|^2 - 2(Af_n|f_n) + \|Af_n\|^2 \leq 2 - 2(Af_n|f_n)$$

for all $n \in \mathbb{N}$, we obtain $\lim_{n\to\infty} f_n - Af_n = 0$.

By Lemma 5.4.6 we find a sequence $(A_k)_{k\in\mathbb{N}}$ of operators on H with finite-dimensional range with $\lim_{n\to\infty} \|A - A_k\|_{\mathrm{HS}} = 0$. In particular, we have $\lim_{n\to\infty} \|A - A_k\| = 0$ by Lemma 5.4.2 (i). Since $(A_k f_n)_{n\in\mathbb{N}}$ is a bounded sequence in a finite-dimensional normed space, it has a convergent subsequence. By using a diagonal sequence argument we may assume (passing to a subsequence) that $(A_k f_n)_{n\in\mathbb{N}}$ converges for every $k \in \mathbb{N}$. Combining this with the fact that $\lim_{n\to\infty} \|A - A_k\| = 0$, an application of the triangle inequality reveals that $(Af_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in H, hence convergent to some $g \in H$. But then also

$$\lim_{n\to\infty} f_n = \lim_{n\to\infty} (f_n - Af_n) + \lim_{n\to\infty} Af_n = g$$

and, in particular, $\|g\| = \lim_{n\to\infty} \|f_n\| = 1$. Since A is continuous, we have $Ag = \lim_{n\to\infty} Af_n = g$, hence $g \in \ker(\mathrm{Id}_H - A) \neq \{0\}$. □

We also need the following observation.

Lemma 5.5.3. *Let $A \in \mathrm{HS}(H)$ be a self-adjoint Hilbert–Schmidt operator on a Hilbert space H and $\alpha \in \mathbb{R} \setminus \{0\}$. Then* $\dim\ker(\alpha \cdot \mathrm{Id}_H - A) \leq \|A\|_{\mathrm{HS}}^2/\alpha^2 < \infty$.

Proof. For any finite orthonormal subset E in $\ker(\alpha \cdot \mathrm{Id}_H - A)$ we have

$$|E| = \sum_{e\in E} \|e\|^2 = \frac{1}{\alpha^2} \sum_{e\in E} \|Ae\|^2 \leq \frac{\|A\|_{\mathrm{HS}}^2}{\alpha^2}.$$

By the Gram–Schmidt process, this implies the claim. □

Let us now restate and prove the Spectral Theorem 5.4.7.

Theorem. *Let $A \in \mathrm{HS}(H)$ be a self-adjoint Hilbert–Schmidt operator on a Hilbert space H. Then there is a unique sequence $(\lambda_n)_{n\in\mathbb{N}}$ in $[0,\infty)$ and unique sequences $(P_n^+)_{n\in\mathbb{N}}$ and $(P_n^-)_{n\in\mathbb{N}}$ of orthogonal projections in $\mathscr{L}(H)$ with*

(i) *$\lambda_{n+1} \leq \lambda_n$ with equality only if $\lambda_n = 0$ for $n \in \mathbb{N}$,*

(ii) *$\lambda_n = 0$ precisely when $P_n^+ = P_n^- = 0$ for $n \in \mathbb{N}$,*

(iii) *$\inf_{n\in\mathbb{N}} \lambda_n = 0$,*

(iv) *$P_n^+(H)$ and $P_n^-(H)$ are finite-dimensional for every $n \in \mathbb{N}$,*

(v) *$P_n^+(H) \perp P_m^+(H)$ and $P_n^-(H) \perp P_m^-(H)$ for $n \neq m$,*

(vi) *$P_n^+(H) \perp P_n^-(H)$ for every $n \in \mathbb{N}$,*

such that $A = \sum_{n=1}^{\infty} \lambda_n (P_n^+ - P_n^-)$ with respect to the Hilbert–Schmidt norm.

Proof. First prove existence. We recursively build sequences of finite rank orthogonal projections $(P_n^+)_{n\in\mathbb{N}}$ and $(P_n^-)_{n\in\mathbb{N}}$ as well as an auxiliary sequence $(A_n)_{n\in\mathbb{N}}$ of self-adjoint Hilbert–Schmidt operators with $A_1 = A$ such that with $\lambda_n := \|A_n\|$ for $n \in \mathbb{N}$ the following assertions hold:

(1) $A = A_n + \sum_{j=1}^{n-1} \lambda_j (P_j^+ - P_j^-)$,

(2) $A_n P_m^{\pm} = P_m^{\pm} A_n = 0$ for $m < n$,

(3) $P_n^+(H) \perp P_n^-(H)$,

(4) $P_n^+(H) \perp P_m^+(H)$ and $P_n^-(H) \perp P_m^-(H)$ for $m < n$,

(5) $\lambda_n = 0$ if and only if $P_n^+ = P_n^- = 0$,

(6) $\lambda_n \leq \lambda_{n-1}$ with equality only if $\lambda_{n-1} = 0$,

for every $n \in \mathbb{N}$.

So assume that $A_1, \dots, A_{n-1}$, $P_1^+, \dots, P_{n-1}^+$, and $P_1^-, \dots, P_{n-1}^-$ have already been defined for some $n \in \mathbb{N}$ (for $n = 1$ this is a trivial assumption). We let

$$A_n := A_{n-1} - \lambda_{n-1}(P_{n-1}^+ - P_{n-1}^-) \in \mathrm{HS}(H).$$

If $A_n = 0$, we set $P_n^+ := 0$ and $P_n^+ := 0$. Otherwise, we let P_n^+ and P_n^- be the orthogonal projections onto the eigenspaces $\ker(\|A_n\| \cdot \mathrm{Id}_H - A_n)$ and $\ker(\|A_n\| \cdot \mathrm{Id}_H + A_n)$, respectively. By Lemma 5.5.3 the projections P_n^+ and P_n^- have finite-dimensional range. We check the list of properties above.

(1) This holds by construction and property (1) for $n - 1$.

(2) For $m \in \{1, \dots, n-1\}$ we have

$$P_m^{\pm} A_n = P_m^{\pm} A - \sum_{j=1}^{n-1} \lambda_j P_m^{\pm}(P_j^+ - P_j^-) = \pm\lambda_m P_m^{\pm} - (\pm\lambda_m P_m^{\pm}) = 0.$$

Similarly (or by taking adjoints), we obtain $A_n P_m^{\pm} = 0$.

(3) This holds by Lemma 5.5.1 (i).

(4) Since $P_n^{\pm}(H) \subseteq A_n(H)$, this follows from (2).

(5) If $\lambda_n = \|A_n\| = 0$, then $P_n^+ = P_n^- = 0$ by definition. If, conversely, $P_n^+ = P_n^- = 0$, then $\lambda_n = \|A_n\| = 0$ by Lemma 5.5.2.

(6) We have $A_{n-1} = A_n + \lambda_{n-1}(P_{n-1}^+ - P_{n-1}^-)$ by definition. Since $A_n(H) \perp P_m^{\pm}(H)$

for every $m < n$ by (2) we obtain

$$\|A_{n-1}f\|^2 = \|A_n f\|^2 + \lambda_{n-1}^2 \|(P_{n-1}^+ - P_{n-1}^-)f\|^2 \geq \|A_n f\|^2.$$

for every $f \in H$. This implies $\lambda_n = \|A_n\| \leq \|A_{n-1}\| = \lambda_{n-1}$.

Assume that $\lambda_n = \lambda_{n-1} \neq 0$. By Lemma 5.5.2 we find $f \in H \setminus \{0\}$ with $\|f\| = 1$ and $A_n f = \|A_n\| f$ or $A_n f = -\|A_n\| f$. In particular, $f \in \operatorname{ran}(A_n)$ which implies $P_{n-1}^{\pm} f = 0$ by part (2). But then

$$A_{n-1}f = A_{n-1}(\mathrm{Id}_H - (P_{n-1}^+ - P_{n-1}^-))f = A_n f \in \{\lambda_{n-1}f, -\lambda_{n-1}f\}$$

by construction of A_n and part (2). But then $f \in P_{n-1}^+(H)$ or $f \in P_{n-1}^-(H)$, contradicting $P_{n-1}^{\pm} f = 0$.

One can check that if $B, C \in \mathrm{HS}(H)$ are self-adjoint with $BC = CB = 0$, then $\|B + C\|_{\mathrm{HS}}^2 = \|B\|_{\mathrm{HS}}^2 + \|C\|_{\mathrm{HS}}^2$. For the constructed sequences we therefore readily obtain that

$$\|A\|_{\mathrm{HS}}^2 = \|A_n\|_{\mathrm{HS}}^2 + \sum_{j=1}^{n-1} \lambda_j^2 (\|P_j^+\|_{\mathrm{HS}}^2 + \|P_j^-)\|_{\mathrm{HS}}^2) \geq \sum_{j=1}^{n-1} \lambda_j^2 (\|P_j^+\|_{\mathrm{HS}}^2 + \|P_j^-)\|_{\mathrm{HS}}^2)$$

for every $n \in \mathbb{N}$ by properties (1) – (4).

In particular, $\sum_{j=1}^{\infty} \lambda_j^2 (\|P_j^+\|_{\mathrm{HS}}^2 + \|P_j^-)\|_{\mathrm{HS}}^2) < \infty$, which implies that $(\sum_{j=1}^{n-1} \lambda_j (P_j^+ - P_j^-))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathrm{HS}(H)$, and hence the limit $\sum_{j=1}^{\infty} \lambda_j (P_j^+ - P_j^-) = \lim_{n\to\infty} \sum_{j=1}^{n-1} \lambda_j (P_j^+ - P_j^-)$ exists with respect to the Hilbert–Schmidt norm (and, in particular, with respect to the operator norm).

We obtain by (5) that $\sum_{j=1}^{\infty} \lambda_j^2 \leq \sum_{j=1}^{\infty} \lambda_j^2 (\|P_j^+\|_{\mathrm{HS}}^2 + \|P_j^-)\|_{\mathrm{HS}}^2) < \infty$ which implies $\lim_{n\to\infty} \|A_n\| = \lim_{n\to\infty} \lambda_n = 0$. Thus,

$$\left\| A - \sum_{j=1}^{\infty} \lambda_j (P_j^+ - P_j^-) \right\| = \lim_{n\to\infty} \left\| A - \sum_{j=1}^{n-1} \lambda_j (P_j^+ - P_j^-) \right\| = \lim_{n\to\infty} \|A_n\| = 0.$$

This shows $A = \sum_{j=1}^{\infty} \lambda_j (P_j^+ - P_j^-)$ as desired.

For uniqueness take any representation $A = \sum_{n=1}^{\infty} \lambda_n (P_n^+ - P_n^-)$ as in Theorem 5.4.7. One can then readily check that

$$\{\lambda_n \mid n \in \mathbb{N}\} \setminus \{0\} = \{|\lambda| \mid \lambda \in \mathbb{R} \text{ eigenvalue of } A\}, \text{ and}$$
$$\operatorname{ran}(P_n^{\pm}) = \ker(\pm\lambda_n \mathrm{Id}_H - A) \text{ for every } n \in \mathbb{N} \text{ with } \lambda_n \neq 0.$$

We leave the details to the reader. In view of properties (i) and (ii) this establishes the desired uniqueness. □

5.6 Comments and Further Reading

The original work [Jac56] of Jacobs from 1956 establishes a decomposition for bounded abelian semigroups of operators on reflexive Banach spaces. A more general version of the splitting result was later established by de Leeuw and Glicksberg in [dLG61]. There are now several different versions of this decomposition and several techniques of proof (see, e.g., [EFHN15, Chapter 16]). The term *discrete spectrum* (also known as *pure point spectrum*) is related to an approach via a decomposition of spectral measures into a discrete and continuous part (see, e.g., [EFHN15, Remark 18.21]).[4] Our proof here follows the ideas discussed in the introduction of [EHK24] which are, however, already present in many earlier texts on ergodic theory.

[4]The terminological origin of *weak mixing* will become clear in Lecture 7.

5.7 Exercises

Exercise 5.1. Prove Proposition 5.1.3.

Exercise 5.2. For the symmetric group $\mathrm{S}_3 = \{\mathrm{id}, (1\,2), (1\,3), (2\,3), (1\,2\,3), (1\,3\,2)\}$ of degree 3 show that $U_\sigma(v_1, v_2, v_3) = (v_{\sigma^{-1}(1)}, v_{\sigma^{-1}(2)}, v_{\sigma^{-1}(3)})$ for $(v_1, v_2, v_3) \in \mathbb{C}^3$ and $\sigma \in \mathrm{S}_3$ defines a unitary representation $U\colon \mathrm{S}_3 \to \mathscr{U}(\mathbb{C}^3)$, $\sigma \mapsto U_\sigma$. Determine a decomposition of $\mathbb{C}^3 = M_1 \oplus M_2$ into irreducible invariant finite-dimensional subspaces as in Proposition 5.1.3.

Exercise 5.3. Show that the following maps are group isomorphisms.

(i) $\mathbb{T} \to \mathbb{Z}^*$, $z \mapsto \chi_z$ where $\chi_z(m) \coloneqq z^m$ for $m \in \mathbb{Z}$ and $z \in \mathbb{T}$.

(ii) $\mathbb{Z}/n\mathbb{Z} \to (\mathbb{Z}/n\mathbb{Z})^*$, $k + n\mathbb{Z} \mapsto \chi_{k+n\mathbb{Z}}$ with $\chi_{k+n\mathbb{Z}}(l + n\mathbb{Z}) = \mathrm{e}^{2\pi \mathrm{i} \frac{lk}{n}}$ for $k, l \in \mathbb{Z}$, where $n \in \mathbb{Z}$.

(iii) $G_1^* \times G_2^* \to (G_1 \times G_2)^*$, $(\chi, \varrho) \mapsto \chi \otimes \varrho$ with $(\chi \otimes \varrho)(x, y) = \chi(x)\varrho(y)$ for $(x, y) \in G_1 \times G_2$, $\chi \in G_1^*$ and $\varrho \in G_2^*$, where G_1, G_2 are any abelian groups.

Exercise 5.4. Equip $\mathbb{C}^n$ for $n \in \mathbb{N}$ with the standard inner product.

(i) Show that for $n \in \mathbb{N}$ the map $\mathbb{C}^n \to (\mathbb{C}^n)', v \mapsto \varphi_v$ with $\varphi_v(u) \coloneqq \sum_{i=1}^n u_i v_i$ for $u = (u_1, \ldots, u_n), v = (v_1, \ldots, v_n) \in \mathbb{C}^n$ is a unitary operator.

(ii) Show that for $n, m \in \mathbb{N}$ there is a unique unitary operator $\Psi\colon \mathbb{C}^n \otimes \mathbb{C}^m \to \mathbb{C}^{nm}$ with

$$\Psi(u \otimes v) = (u_1 v_1, \ldots, u_1 v_m, u_2 v_1, \ldots\ldots\ldots, u_n v_1, \ldots, u_n v_m)$$

for all $u = (u_1, \ldots, u_n) \in \mathbb{C}^n$ and $\mathrm{v} = (v_1, \ldots, v_m) \in \mathbb{C}^m$.

Exercise 5.5. Let $V \in \mathscr{U}(H)$ be a unitary operator on a Hilbert space H and $(a_n)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathbb{C}$ such that the limit $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} a_n z^n$ exists for every $z \in \mathbb{T}$. Show that for each $f \in H$ the limit $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} a_n V^n f$ exists. *Hint: Use Proposition 5.3.3 to show that* $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} a_n V^n f = 0$ *for every* $f \in H_{\mathrm{wm}}$.

Exercise 5.6. We consider the Hilbert space ℓ^2 of square-summable sequences $(a_n)_{n\in\mathbb{N}}$ in $\mathbb{C}$ with the inner product given by $((a_n)_{n\in\mathbb{N}} | (b_n)_{n\in\mathbb{N}}) \coloneqq \sum_{n=1}^\infty a_n \overline{b_n}$ for $(a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \in \ell^2$.

(i) Show that for a double sequence $c = (c_{n,m})_{n,m\in\mathbb{N}}$ of complex numbers $\mathbb{C}$ with $\|c\|_2 \coloneqq \sum_{n=1}^\infty \sum_{m=1}^\infty |c_{n,m}|^2 < \infty$ the map

$$A_c\colon \ell^2 \to \ell^2, \quad (a_n)_{n\in\mathbb{N}} \mapsto \left(\sum_{m=1}^\infty c_{n,m} a_m \right)_{n\in\mathbb{N}}$$

is a Hilbert–Schmidt operator with $\|A_c\|_{\mathrm{HS}} = \|c\|_2$.

(ii) Show that $(A_c)^* = A_{c^*}$ for each double sequence $c = (c_{n,m})_{n,m\in\mathbb{N}}$ of complex numbers with $\|c\|_2 < \infty$, where $(c^*)_{n,m} := \overline{c_{m,n}}$ for $n, m \in \mathbb{N}$. In particular, A_c is self-adjoint precisely when $c_{n,m} = \overline{c_{m,n}}$ for all $n, m \in \mathbb{N}$.

(iii) Show that for every Hilbert-Schmidt operator $A\colon \mathrm{L}^2 \to \ell^2$ there is a unique double sequence c of complex numbers with $\|c\|_2 < \infty$ such that $A = A_c$.

Exercise 5.7. Let H be a Hilbert space and consider the space $\mathrm{HS}(H)$ of Hilbert–Schmidt operators. Let E be an orthonormal basis for H and consider the set $\mathcal{F}$ of all finite subsets of E directed by set inclusion. Then

$$(A|B)_{\mathrm{HS}} = \sum_{e\in E}(Ae|Be) := \lim_{F\in\mathcal{F}} \sum_{e\in F}(Ae|Be)$$

for all for $A, B \in \mathrm{HS}(H)$ defines an inner product on $\mathrm{HS}(H)$ with $\|A\|_{\mathrm{HS}} = \sqrt{(A|A)_{\mathrm{HS}}}$ for $A \in \mathrm{HS}(H)$.[5]

[5]You may use the following fact (or prove it as a bonus exercise): If $(x_i)_{i\in I}$ is a **Cauchy net** in a complete metric space (X, d_X), i.e., for every $\varepsilon > 0$ there is some $i_0 \in I$ with $d_X(x_i, x_j) \le \varepsilon$ for all $i, j \ge i_0$, then $(x_i)_{i\in I}$ converges to some $x \in X$.

Lecture 6

In the first part of this lecture we introduce compact groups and show that their strongly continuous unitary representations always have *discrete spectrum*, i.e., the weakly mixing part of the JdLG-decomposition is trivial. In the second part of the lecture we then prove a famous representation and classification result due to Halmos and von Neumann for ergodic measure-preserving systems with discrete spectrum.

6.1 Compact Groups and Discrete Spectrum

In the previous lecture we have seen that for any unitary representation $U\colon G \to \mathscr{U}(H)$ of a group G we obtain a decomposition $H = H_{\mathrm{ds}} \oplus H_{\mathrm{wm}}$ of the Hilbert space H into the discrete spectrum part H_{ds} and the weakly mixing part H_{wm}. We now study situations when the weakly mixing part is trivial.

Definition 6.1.1. A unitary representation $U\colon G \to \mathscr{U}(H)$ of a group G **has discrete spectrum** if

$$H = H_{\mathrm{ds}} = \overline{\bigcup \{M \subseteq H \mid M \text{ invariant finite-dimensional subspace}\}}.$$

To obtain examples for this situation, we consider groups G endowed with additional structure.

Definition 6.1.2. A group G equipped with a topology is called a **topological group** if the multiplication and inversion maps

$$\begin{aligned} \cdot\colon G \times G \to G, &\quad (x,y) \mapsto xy \\ {}^{-1}\colon G \to G, &\quad x \mapsto x^{-1} \end{aligned}$$

are both continuous (where $G \times G$ is endowed with the product topology). It is a **compact group** if, in addition, the topology on G is compact[1].

[1]Recall that this includes the Hausdorff property.

One can readily check that for a topological group G the left and right rotations $l_x, r_x \colon G \to G$ with $l_x(y) := xy$ and $r_x(y) = yx$ for $y \in G$ are homeomorphisms.

Basic examples of compact groups include every finite group with the discrete topology, as well as the ***d*-torus** $\mathbb{T}^d$ with componentwise multiplication, the groups $\mathscr{O}(d)$ of orthogonal $d \times d$-matrices and $\mathscr{U}(d)$ of unitary $d \times d$-matrices (with the natural Euclidean topology) for $d \in \mathbb{N}$. Compact groups admit a rich structure theory (see, e.g., [HM20]) with many applications in mathematics and physics. The following is one of their important features (see, e.g., [Fol15, Section 2.2]).

Theorem and Definition 6.1.3. *For a compact group G there is a unique regular Borel probability measure $\mathrm{m}_G \in \mathrm{P}(G)$ such that*

(i) $\mathrm{m}_G(xA) = \mathrm{m}_G(A)$ *for every Borel set $A \subseteq G$ and $x \in G$.*

*This is called the **Haar measure** of G and also has the following properties.*

(ii) $\mathrm{m}_G(Ax) = \mathrm{m}_G(A)$ *for every Borel set $A \subseteq G$ and $x \in G$.*

(iii) $\mathrm{m}_G(A^{-1}) = \mathrm{m}_G(A)$ *for every Borel set $A \subseteq G$.*

(iv) $\mathrm{m}_G(O) > 0$ *for every non-empty open set $O \subseteq G$.*

A proof of Theorem 6.1.3 in the special case of compact abelian groups is discussed in Exercise 6.2. From now on, we equip every compact group G with the Borel σ-algebra $\mathcal{B}(G)$ and the Haar measure m_G.

Remark 6.1.4. Via Proposition 1.1.4 (ii) properties (i) - (iii) of Theorem 6.1.3 translate to the following assertions: For any integrable function $f \colon G \to \mathbb{C}$ also the functions $f \circ l_x, f \circ r_x, \colon G \to \mathbb{C}$ for $x \in G$ and $f \circ {}^{-1} \colon G \to \mathbb{C}$ are integrable with

$$\int_G f(y)\,\mathrm{dm}_G(y) = \int_G f(xy)\,\mathrm{dm}_G(y) = \int_G f(yx)\,\mathrm{dm}_G(y) = \int_G f(y^{-1})\,\mathrm{dm}_G(y).$$

Moreover, Theorem 6.1.3 (iv) implies that $\int |f|\,\mathrm{dm}_G > 0$ for every $f \in \mathrm{C}(G) \setminus \{0\}$. Thus, the canonical map $\mathrm{C}(G) \to \mathrm{L}^2(G)$ from Lemma 4.1.7 is injective, and we identify $\mathrm{C}(G)$ with a dense linear subspace of $\mathrm{L}^2(G)$.

We demonstrate this situation in a concrete case.

Example 6.1.5. Consider the torus $G = \mathbb{T}$. Then the unital, positive linear map $\mathrm{C}(\mathbb{T}) \to \mathbb{C}$, $f \mapsto \int_0^1 f(e^{2\pi i t})\,\mathrm{d}t$ defines a regular Borel probability measure $\mu \in \mathrm{P}(\mathbb{T})$. For $f \in \mathrm{C}(\mathbb{T})$ and an element $b = \mathrm{e}^{2\pi \mathrm{i} s} \in \mathbb{T}$ with $s \in [0,1)$ we obtain

$$\begin{aligned}\int_0^1 f(be^{2\pi \mathrm{i}t})\,\mathrm{d}t &= \int_s^{1+s} f(e^{2\pi \mathrm{i}t})\,\mathrm{d}t = \int_s^1 f(e^{2\pi \mathrm{i}t})\,\mathrm{d}t + \int_1^{1+s} f(e^{2\pi \mathrm{i}t})\,\mathrm{d}t \\ &= \int_s^1 f(e^{2\pi \mathrm{i}t})\,\mathrm{d}t + \int_0^s f(e^{2\pi \mathrm{i}(t+1)})\,\mathrm{d}t = \int_0^1 f(e^{2\pi \mathrm{i}t})\,\mathrm{d}t\end{aligned}$$

since the exponential function is 2πi-periodic. By Lemma 3.2.10 this shows that $\mu = \mathrm{m}_{\mathbb{T}}$ is the Haar measure of $\mathbb{T}$.

If we consider unitary representations of topological (in particular, compact) groups, it is natural to demand some kind of continuity.

Definition 6.1.6. A unitary representation $U\colon G \to \mathscr{U}(H)$ of a topological group G is **strongly continuous** if $G \to H$, $x \mapsto U_x f$ is continuous for every $f \in H$.

Strong continuity is preserved by the two important constructions from Section 5.3:

Lemma 6.1.7. *Let G be a topological group.*

(i) *If $U\colon G \to \mathscr{U}(H)$ is a strongly continuous unitary representation of G, then its dual representation $\overline{U}\colon G \to \mathscr{U}(H')$ is also strongly continuous.*

(ii) *If $U_i\colon G \to \mathscr{U}(H_i)$ are strongly continuous unitary representations of G for $i = 1, 2$, then their tensor product representation $U_1 \otimes U_2\colon G \to \mathscr{U}(H_1 \otimes H_2)$ is also strongly continuous.*

Proof. Part (i) follows directly from the definition. For part (ii) observe that by Exercise 6.3 it suffices to show "weak continuity", i.e., we have to prove that each map $G \to \mathbb{C}, x \mapsto ((U_1 \otimes U_2)_x f|g)$ for $f, g \in H_1 \otimes H_2$ is continuous. Using linearity and an approximation argument one can readily reduce the assertion to the case of simple tensors $f = f_1 \otimes f_2$ and $g = g_1 \otimes g_2$ for $f_1, g_1 \in H_1$ and $f_2, g_2 \in H_2$. But then

$$G \to \mathbb{C}, \quad x \mapsto ((U_1 \otimes U_2)_x f|g) = ((U_1)_x f_1|g_1) \cdot ((U_2)_x f_2|g_2)$$

is continuous since U_1 and U_2 are strongly (and in particular weakly) continuous. □

The following are important examples for strongly continuous unitary representations.

Proposition and Definition 6.1.8. *For every compact group G the maps*

$$L\colon G \to \mathscr{U}(\mathrm{L}^2(G)), \quad x \mapsto L_x$$
$$R\colon G \to \mathscr{U}(\mathrm{L}^2(G)), \quad x \mapsto R_x$$

with $L_x f := f \circ l_{x^{-1}}$ and $R_x f := f \circ r_x$ for $f \in \mathrm{L}^2(G)$ and $x \in G$ are strongly continuous unitary representations. We call L and R the ***left*** *and* ***right regular representation*** *of G, respectively.*

For the proof we use the following topological lemma, see Exercise 6.4.

Lemma 6.1.9. *Let Ω be a topological space, and K and L be compact spaces. Assume further that $\varphi\colon \Omega \times K \to L$, $(\omega, x) \mapsto \varphi_\omega(x)$ is continuous with respect to the product topology. Then for every $f \in \mathrm{C}(L)$ the map $\Omega \to \mathrm{C}(K)$, $\omega \mapsto f \circ \varphi_\omega$ is continuous with respect to the supremum norm on $\mathrm{C}(K)$.*

Proof of Proposition 6.1.8. With Remark 6.1.4 it is easy to see that L and R are well-defined unitary representations. We now show that L is strongly continuous (the proof for R is similar). First pick a continuous function $f \in \mathrm{C}(G)$. Let $x \in G$ and $\varepsilon > 0$. Since the map $G \times G \to G$, $(y, z) \mapsto y^{-1}z = l_{y^{-1}}(z)$ is continuous, we can apply Lemma 6.1.9 to find a neighborhood O of $x \in G$ with $\|f \circ l_{y^{-1}} - f \circ l_{x^{-1}}\|_\infty \leq \varepsilon$ for every $y \in O$. But then

$$\|L_y f - L_x f\|_2 = \|f \circ l_{y^{-1}} - f \circ l_{x^{-1}}\|_2 \leq \|f \circ l_{y^{-1}} - f \circ l_{x^{-1}}\|_\infty \leq \varepsilon$$

for every $y \in O$.

For general $f \in \mathrm{L}^2(G)$ the result now follows by approximation: For $\varepsilon > 0$ we find by Lemma 4.1.7 some $g \in \mathrm{C}(G)$ with $\|g - f\|_2 \leq \varepsilon$. Since U_y is unitary, we also have $\|U_y g - U_y f\|_2 \leq \varepsilon$ for every $y \in G$. The result therefore follows by the triangle inequality. □

Remark 6.1.10. With slight adjustments we can also cover representations on homogeneous spaces.

(1) If G is a compact group and $W \subseteq G$ is a closed subgroup, consider the quotient space G/W with the quotient topology (see, e.g., [Sin19, Section 6.1]), which is again a compact (Hausdorff!) space (see [Sin19, Proposition 12.3.2]).

(2) We can equip G/W with the Borel σ-algebra and the pushforward measure $\mathrm{m}_{G/W} := q_*\mathrm{m}_G$ with respect to the quotient map $q\colon G \to G/W, x \mapsto xW$. Then $\mathrm{m}_{G/W}(xA) = \mathrm{m}_{G/W}(A)$ for every $x \in G$ and every Borel set $A \subseteq G/W$.

(3) Then $U\colon G \to \mathscr{U}(\mathrm{L}^2(G/W))$, $f \mapsto U_x f$ with $U_x f(yW) := f(x^{-1}yW)$ for $yW \in G/W$, $f \in \mathrm{L}^2(G/W)$ and $x \in G$ is also a strongly continuous unitary representation.

We now prove the following important result.

Proposition 6.1.11. *If $U\colon G \to \mathscr{U}(H)$ is a strongly continuous unitary representation of a compact group G, then U has discrete spectrum.*

For the proof we need the following observation.

Lemma 6.1.12. *Assume that $U\colon G \to \mathscr{U}(H)$ is a strongly continuous unitary representation of a compact group G. Let further P be the orthogonal projection onto the fixed space* $\mathrm{fix}(U(G))$. *Then* $(Pf|g) = \int_G (U_x f|g)\,\mathrm{d}\mathrm{m}_G(x)$ *for all $f, g \in H$.*

Proof. Take $f, g \in H$. For every $x \in G$ we have $U_x P = P$ and this implies $(Pf|g) = \int_G (U_x Pf|g)\,\mathrm{d}\mathrm{m}_G(x)$ (since m_G is a probability measure). We choose a sequence $(\sum_{i=1}^{k_n} \lambda_{n,i} U_{x_{n,i}} f)_{n\in\mathbb{N}}$ in the convex hull $\mathrm{co}\{U_x f \mid x \in G\}$ converging to Pf (see Theorem 3.1.5 (ii)). Then, by Lebesgue's theorem and invariance of the Haar measure (cf. Remark 6.1.4),

$$
\begin{aligned}
(Pf|g) &= \int_G (PU_x f|g)\,\mathrm{dm}_G(x) = \lim_{n\to\infty}\sum_{i=1}^{k_n}\lambda_{n,i}\int_G (PU_{xx_{n,i}}f\mid g)\,\mathrm{dm}_G(x)\\
&= \lim_{n\to\infty}\sum_{i=1}^{k_n}\lambda_{n,i}\int_G (U_x f|g)\,\mathrm{dm}_G(x) = \int_G (U_x f|g)\,\mathrm{dm}_G(x).
\end{aligned}
$$

□

Proof of Proposition 6.1.11. The representation $U\otimes\overline{U}\colon G\to\mathscr{U}(H\otimes H')$ is strongly continuous by Lemma 6.1.7. For $f\in H_{\mathrm{wm}}$, we have $f\otimes\overline{f}\in\operatorname{fix}(U\otimes\overline{U})^{\perp}$ and thus

$$0 = \int_G (U_x f\otimes\overline{U_x f}|f\otimes\overline{f})\,\mathrm{dm}_G(x) = \int_G |(U_x f|f)|^2\,\mathrm{dm}_G(x)$$

by Lemma 6.1.12. By the last part of Remark 6.1.4 we have $(U_x f|f)=0$ for every $x\in G$, in particular for $x=1$. Thus, $(f|f)=0$ which yields $f=0$. □

Proposition 6.1.11 has interesting consequences for compact abelian groups. We first introduce the following notion. Recall here from Definition 5.1.6 that G^* denotes the group of all characters (i.e., group homomorphisms $\chi\colon G\to\mathbb{T}$) of a group G.

Definition 6.1.13. Let G be a topological group. We call the subgroup

$$G' := \{\chi\colon G\to\mathbb{T}\mid \chi \text{ continuous group homomorphism}\}$$

of G^* the **Pontryagin dual** of G.

We obtain the following consequence of Proposition 6.1.11.

Proposition 6.1.14. *Let $U\colon G\to\mathscr{U}(H)$ be a strongly continuous unitary representation of a compact abelian group G. Then $H=\overline{\operatorname{lin}}\bigcup_{\chi\in G'}\ker(\chi-U)$.*

Proof. By Proposition 6.1.11 and Corollary 5.1.11 we have $H=\overline{\operatorname{lin}}\bigcup_{\chi\in G^*}\ker(\chi-U)$. However, if $f\in\ker(\chi-U)\setminus\{0\}$ is an eigenvector with respect to some character $\chi\in G^*$, then χ is automatically continuous, since $U_x f=\chi(x)f$ for all $x\in G$ and U is strongly continuous. □

We combine this with the following fundamental[2] theorem about compact spaces (see, e.g., [Sin19, Proposition 5.1.9 and Theorem 8.2.11] for a proof).

[2]In fact, we have already used the result implicitly at several points, e.g., in the proofs of Theorem 3.2.6 and Lemma 4.1.7.

Theorem 6.1.15 (Tietze). *Let K be a compact space and $A \subseteq K$ a closed subset. Then any continuous function $f\colon A \to [0,1]$ can be extended to a continuous function $f\colon K \to [0,1]$.*

In particular, Tietze's theorem implies that for any compact space K, the continuous functions $f\colon K \to \mathbb{C}$ "separate points", i.e., for $x, y \in K$ with $x \neq y$ there is $f \in \mathrm{C}(K)$ with $f(x) \neq f(y)$. We use this fact to prove the following.

Proposition 6.1.16. *Let G be a compact abelian group. Then G' separates points, i.e., for $x, y \in G$ with $x \neq y$ there is a continuous character $\chi\colon G \to \mathbb{T}$ with $\chi(x) \neq \chi(y)$.*

Proof. Take $x, y \in G$ with $x \neq y$. By Tietze's theorem we find $f \in \mathrm{C}(G)$ with $f(x) \neq f(y)$, hence $(f \circ l_x)(1) \neq (f \circ l_y)(1)$. This implies $L_{x^{-1}} \neq L_{y^{-1}}$. Since the left regular representation L has discrete spectrum, we find by Proposition 6.1.14 some $\chi \in G'$ and an eigenvector $f \in \mathrm{L}^2(X)$ with respect to χ such that $L_{x^{-1}}f \neq L_{y^{-1}}f$, hence $\chi(x^{-1})f \neq \chi(y^{-1})f$, which implies $\chi(x) \neq \chi(y)$. □

Corollary 6.1.17. *Let G be a compact abelian group. For a subgroup $W \subseteq G$ the following assertions are equivalent.*

(a) *W is dense in G.*

(b) *If $\chi \in G'$ satisfies $\chi(x) = 1$ for every $x \in W$, then $\chi = \mathbb{1}$.*

Proof. The implication "(a) ⇒ (b)" follows by continuity of the elements $\chi \in G'$. We prove the converse implication "(b) ⇒ (a)" by contraposition. So assume that W is not dense in G. Since the closure of W is again a subgroup (see Exercise 6.5), we may assume that W is closed. We further pick some $y \in G \setminus W$. Consider the quotient space G/W which is compact (see Remark 6.1.10) and an abelian group. In fact, it is again a compact abelian group (see again Exercise 6.5). Since $yW \neq 1W$ we find by Proposition 6.1.16 some $\varrho \in (G/W)'$ with $\varrho(yW) \neq 1 = \varrho(1W)$. The composition $\chi \coloneqq \varrho \circ q\colon G \to \mathbb{T}$ for the quotient map $q\colon G \to G/W$ thus yields a continuous character $\chi \in G'$ with $\chi(x) = 1$ for all $x \in W$ but $\chi \neq \mathbb{1}$. □

Proposition 6.1.16 is particularly interesting in combination with the following famous approximation result (see, e.g., [Ped89, Theorem 4.3.4]).

Theorem 6.1.18 (Stone–Weierstraß). *Let K be a compact space. Assume that $A \subseteq \mathrm{C}(K)$ is a linear subpace with the following properties.*

(i) *$f \cdot g \in A$ for all $f, g \in A$.*

(ii) *$\overline{f} \in A$ for every $f \in A$.*

(iii) *$\mathbb{1} \in A$.*

(iv) *For all $x, y \in K$ with $x \neq y$ there is $f \in A$ with $f(x) \neq f(y)$.*

Then A is dense in $\mathrm{C}(K)$.

This allows us to prove that the continuous characters form an orthonormal basis. More generally, we have the following.

Proposition 6.1.19. *Let G be a compact abelian group with Pontryagin dual G'. For a subgroup $W \subseteq G'$ the following assertions are equivalent.*

(a) *W separates the points of G.*

(b) *W is an orthonormal basis of* $\mathrm{L}^2(G)$.

(c) *$W = G'$.*

We start with the following lemma.

Lemma 6.1.20. *The Pontryagin dual G' of a compact abelian group G defines an orthonormal set in* $\mathrm{L}^2(G)$.

Proof. If $\chi \in G'$, the left regular representation satisfies $L_x\chi(y) = \chi(x^{-1}y) = \overline{\chi}(x)\chi(y)$ for all $x, y \in G$. Thus, $\chi \in \ker(\overline{\chi} - L) \setminus \{0\}$ is an eigenvector with respect to $\overline{\chi}$. The claim thus follows from Proposition 5.1.10 (i). □

Proof of Proposition 6.1.19. The implication "(c) ⇒ (a)" is Proposition 6.1.16. For "(a) ⇒ (b)" observe that if W separates the points of G, then the linear hull $A := \operatorname{lin} W$ satisfies all conditions of Theorem 6.1.18, hence is dense in $\mathrm{C}(G)$ with respect to the supremum norm, and in particular with respect to the L^2-norm. Since $\mathrm{C}(G)$ defines a dense subspace of $\mathrm{L}^2(G)$ by Lemma 4.1.7, we obtain that $\operatorname{lin} W$ is a dense subset of $\mathrm{L}^2(G)$. Combined with Lemma 6.1.20 this shows that W is an orthonormal basis, hence (b) follows. Finally, since G' defines an orthonormal set, the implication "(b) ⇒ (c)" holds by the definition of an orthonormal basis as a maximal orthonormal subset. □

Example 6.1.21. Consider the torus $\mathbb{T}$ as a compact group. For every $m \in \mathbb{Z}$ the map $\chi_m \colon \mathbb{T} \to \mathbb{T}$, $z \mapsto z^m$ is a continuous character. Since $W := \{\chi_m \mid m \in \mathbb{Z}\}$ is a subgroup of $\mathbb{T}'$ and separates the points of $\mathbb{T}$ (all we need is χ_1), we conclude from Proposition 6.1.19 that $W = \mathbb{T}'$. This readily yields that the map $\mathbb{Z} \to \mathbb{T}'$, $m \mapsto \chi_m$ is a group isomorphism.

We finish this section with an elegant proof of the following classical result.

Theorem 6.1.22 (Kronecker)**.** *Assume that $a \in \mathbb{T}$ is not a root of unity, i.e., $a^m \neq 1$ for all $m \in \mathbb{Z} \setminus \{0\}$. Then $\{a^n \mid n \in \mathbb{Z}\}$ is dense in $\mathbb{T}$.*

Proof. Take $a \in \mathbb{T}$ such that $W := \{a^n \mid n \in \mathbb{Z}\}$ is not dense in $\mathbb{T}$. By Corollary 6.1.17 we find some character $\chi \in \mathbb{T}'$ with $\chi \neq \mathbb{1}$ and $\chi(a^n) = 1$ for every $n \in \mathbb{Z}$. By Example 6.1.21 there is a unique $m \in \mathbb{Z} \setminus \{0\}$ with $\chi = \chi_m$. But this implies $a^m = \chi_m(a) = 1$. □

6.2 Systems with Discrete Spectrum

We now return to measure-preserving systems over our fixed abelian group Γ.

Definition 6.2.1. A measure-preserving system (X, T) has **discrete spectrum** if the induced Koopman representation $U_T\colon \Gamma \to \mathscr{U}(\mathrm{L}^2(X))$ has discrete spectrum.

It is clear from the definition that every trivial system (X, Id) (see Example 2.1.2) has discrete spectrum. Here is a more interesting example.

Example 6.2.2. For $a \in \mathbb{T}$ the rotation $l_a\colon \mathbb{T} \to \mathbb{T}$, $z \mapsto az$ induces a topological dynamical system $(\mathbb{T}, \tau_a)$ over $\mathbb{Z}$ via $\tau_a\colon \mathbb{Z} \to \mathrm{Homeo}(\mathbb{T})$, $k \mapsto l_a^k = l_{a^k}$. The Haar measure $\mathrm{m}_{\mathbb{T}}$ is invariant, and hence we obtain a measure-preserving system $(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mathrm{m}_{\mathbb{T}}, (\tau_a)^*)$. The left regular representation $L\colon \mathbb{T} \to \mathscr{U}(\mathrm{L}^2(\mathbb{T}))$ of $\mathbb{T}$ has discrete spectrum by Propositions 6.1.8 and 6.1.11. Since

$$U_{\tau_a^*}(\mathbb{Z}) = \{U_{l_a}^k \mid k \in \mathbb{Z}\} \subseteq \{L_b \mid b \in \mathbb{T}\} = L(\mathbb{T}),$$

every invariant linear subspace with respect to L is also invariant with respect to the Koopman representation $U_{\tau_a^*}$. Thus, $(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mathrm{m}_{\mathbb{T}}, (\tau_a)^*)$ has discrete spectrum.

We even obtain the following more general class of examples.

Example 6.2.3. Let $c\colon \Gamma \to G$ be a group homomorphism from the abelian group Γ to any compact abelian group G. This yields a topological dynamical system (G, τ_c) via $\tau_c\colon \Gamma \to \mathrm{Homeo}(G)$, $\gamma \mapsto l_{c(\gamma)}$. By the same reasoning as in Example 6.2.2 this gives rise to a measure-preserving system $(G, \mathcal{B}(G), \mathrm{m}_G, (\tau_c)^*)$ with discrete spectrum. We call this a **rotation system**.

There is a nice characterization of ergodicity for these systems. Recall that for a topological dynamical system (K, τ) we denote by $\mathrm{P}(K, \tau)$ the set of its invariant regular Borel probability measures (see Definition 3.2.9).

Proposition 6.2.4. *For a group homomorphism $c\colon \Gamma \to G$ to a compact abelian group G the following assertions are equivalent.*

(a) *The map c has dense range.*

(b) $\mathrm{P}(G, \tau_c) = \{\mathrm{m}_G\}$.

(c) *The system $(G, \mathcal{B}(G), \mathrm{m}_G, (\tau_c)^*)$ is ergodic.*

Proof. We first prove the implication "(a) $\Rightarrow$ (b)". So assume that $c(\Gamma)$ is dense in G and take an invariant measure $\mu \in \mathrm{P}_{\tau_c}(G)$. For $f \in \mathrm{C}(G)$ we obtain from Lemma 6.1.9 that the map $G \to \mathrm{C}(G)$, $x \mapsto f \circ l_x$ is continuous, hence also

$$h\colon G \to \mathbb{C}, \quad x \mapsto \int_G f \circ l_x \,\mathrm{d}\mu$$

is continuous. For every $x \in c(\Gamma)$ we have $h(x) = \int_G f \, \mathrm{d}\mu$. Since $c(\Gamma)$ is dense in G and h is continuous, we even obtain $\int f \circ l_x \, \mathrm{d}\mu = h(x) = \int_G f \, \mathrm{d}\mu$ for every $x \in G$. By Lemma 3.2.10 this means that $\mu(xA) = \mu(A)$ for every $x \in G$ and every Borel set $A \subseteq G$, and hence μ is the Haar measure m_G.

The implication "(b) $\Rightarrow$ (c)" follows directly from Proposition 4.1.9. We finally prove "(c) $\Rightarrow$ (a)". So assume that $(G, \mathcal{B}(G), \mathrm{m}_G, (\tau_c)^*)$ is ergodic. In view of Corollary 6.1.17 it suffices to show that $\chi \in G'$ with $\chi(c(\gamma)) = 1$ for every $\gamma \in \Gamma$ implies $\chi = \mathbb{1}$. So take such $\chi \in G'$ and observe that

$$U_{(\tau_c)_\gamma} \chi(x) = \chi(c(\gamma)x) = \chi(c(\gamma))\chi(x) = \chi(x)$$

for all $x \in G$ and $\gamma \in \Gamma$. Thus, $\chi \in \operatorname{fix}(U_{(\tau_c)^*})$, and by ergodicity we find $d \in \mathbb{C}$ with $\chi = d\mathbb{1}$ (see Corollary 2.2.14). Since $|\chi| = \mathbb{1}$ we have $|d| = 1$, and then $d = (d\mathbb{1}|\mathbb{1}) = (\chi|\mathbb{1}) \in \{0, 1\}$ by Lemma 6.1.20, which implies $d = 1$. Thus, $\chi = \mathbb{1}$. □

Example 6.2.5. For $a \in \mathbb{T}$ consider the rotation $(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mathrm{m}_\mathbb{T}, (\tau_a)^*)$ on the torus from Example 6.2.2. By Theorem 6.1.22 and Proposition 6.2.4 this system is ergodic precisely when a is not a root of unity.

It turns out that (up to an isomorphism) ergodic rotation systems are the only ergodic measure-preserving systems with discrete spectrum.

Theorem 6.2.6 (Halmos–von Neumann Representation Theorem)**.** *Let (X, T) be an ergodic measure-preserving system with discrete spectrum. Then there is a group homomorphism $c\colon \Gamma \to G$ with dense range to a compact abelian group G such that (X, T) is isomorphic to the rotation system $(G, \mathcal{B}(G), \mathrm{m}_G, (\tau_c)^*)$.*

The following related result shows that an ergodic system (X, T) with discrete spectrum is (again up to an isomorphism) determined by the point spectrum $\sigma_\mathrm{p}(U_T) \subseteq \Gamma^*$ (cf. Definition 5.1.8) of its Koopman representation $U_T \colon \Gamma \to \mathscr{U}(\mathrm{L}^2(X))$.

Theorem 6.2.7 (Halmos–von Neumann Uniqueness Theorem)**.** *Two ergodic measure-preserving systems (X, T) and (Y, S) with discrete spectrum are isomorphic if and only if $\sigma_\mathrm{p}(U_T) = \sigma_\mathrm{p}(U_S)$.*

There is also a third aspect of the Halmos–von Neumann classification of ergodic systems with discrete spectrum:

Theorem 6.2.8 (Halmos–von Neumann Realization Theorem)**.** *A subset $W \subseteq \Gamma^*$ of the dual group Γ^* is a subgroup if and only if there is an ergodic measure-preserving system (X, T) with discrete spectrum such that $\sigma_p(U_T) = W$.*

With these three aspects, ergodic systems with discrete spectrum are completely understood: They are all isomorphic to rotation systems, and are (up to isomorphism) in a one-to-one correspondence with the subgroups of the dual group Γ^*.

As a first step towards the proof of these results, we establish the following description of the point spectrum of rotation systems.

Proposition 6.2.9. *Let $c\colon \Gamma \to G$ be a group homomorphism to a compact abelian group G and consider the corresponding rotation system $(G, \mathcal{B}(G), \mathrm{m}_G, (\tau_c)^*)$. Then the map $c^*\colon G' \to \sigma_{\mathrm{p}}(U_{(\tau_c)^*})$, $\chi \mapsto \chi \circ c$ is a surjective group homomorphism. It is an isomorphism if and only if c has dense range.*

Proof. The map c^* is well-defined since $\chi \in \ker(\chi \circ c - U_{(\tau_c)^*}) \setminus \{0\}$ for every $\chi \in G'$. Moreover, it is clear that c^* is a group homomorphism. To see that it is surjective, observe that G' is an orthonormal system of eigenvectors of $U_{(\tau_c)^*}$ with respect to the eigenvalues $\chi \circ c$ (where $\chi \in G'$). But by Proposition 6.1.19 it is an orthonormal basis and hence by Proposition 5.1.10 there can be no further eigenvalues.

If c has dense range, then the map c^* injective (since two continuous functions agreeing on a dense subset have to be identical). On the other hand, if $c(\Gamma)$ is not dense in G, we find by Corollary 6.1.17 some $\chi \in G' \setminus \{\mathbb{1}\}$ with $\chi(c(\gamma)) = 1$ for all $\gamma \in \Gamma$, hence $c^*(\chi) = c^*(\mathbb{1})$. □

Thus, for ergodic rotations $(G, \mathcal{B}(G), \mathrm{m}_G, (\tau_c)^*)$ we can identify the point spectrum of the Koopman representation with the Pontryagin dual G'. The next auxiliary result is purely group theoretic. We give a proof (even in a slightly more general version) as a supplement at the end of this lecture.

Lemma 6.2.10. *Let G be an abelian group and $W \subseteq G$ a subgroup. Every group homomorphism $\alpha\colon W \to \mathbb{T}$ can be extended to a group homomorphism $\alpha\colon G \to \mathbb{T}$.*

The remaining lemmas needed for our proof of Theorems 6.2.6, 6.2.7, and 6.2.8 establish some "spectral theoretic properties" of measure-preserving systems.

Lemma 6.2.11. *Let (X, T) be a measure-preserving system. If $f \in \ker(\chi - U_T)$ for some $\chi \in \Gamma^*$, then $|f| \in \mathrm{fix}(U_T)$.*

Proof. For $f \in \ker(\chi - U_T)$ we have $U_{T_\gamma}|f| = |U_{T_\gamma} f| = |\chi(\gamma) f| = |f|$ for every $\gamma \in \Gamma$. Thus, $|f| \in \mathrm{fix}(U_T)$ as claimed. □

Lemma 6.2.12. *For every ergodic measure-preserving system (X, T) the following assertions hold.*

(i) *Each eigenspace $\ker(\chi - U_T)$ for $\chi \in \Gamma^*$ is contained in $\mathrm{L}^\infty(\mathrm{X})$ and is at most one-dimensional.*

(ii) *The point spectrum $\sigma_{\mathrm{p}}(U_T)$ is a subgroup of the dual Γ^*.*

Proof. Note first that for $\chi \in G^*$ and $f \in \ker(\chi - U_T) \setminus \{0\}$ we have $|f| = c\mathbb{1}$ for some $c \in (0, \infty)$ by Lemma 6.2.11 and ergodicity (see Corollary 2.2.14). In particular, $\ker(\chi - U_T) \subseteq \mathrm{L}^\infty(X)$.

For eigenvectors $f \in \ker(\chi - U_T) \setminus \{0\}$ and $g \in \ker(\varrho - U_T) \setminus \{0\}$ where $\chi, \varrho \in \Gamma^*$ we have $f\overline{g} \in \ker(\chi\overline{\varrho} - U_T) \setminus \{0\}$. Indeed for such f, g, χ, ϱ write $|f| = c\mathbb{1}$ and $|g| = d\mathbb{1}$ for some $c, d \in (0, \infty)$. Then $U_{T_\gamma}(f\overline{g}) = \chi(\gamma)\overline{\varrho}(\gamma)f\overline{g}$ for every $\gamma \in \Gamma$ by Proposition 1.3.3 and Corollary 1.3.8. Moreover, $|f\overline{g}| = cd$, hence $f\overline{g} \neq 0$.

We prove (i) by applying this observation to $\chi = \varrho$. Assume that $f, g \in \ker(\chi - U_T) \setminus \{0\}$ and write $|g| = d\mathbb{1}$ for some $d \in (0, \infty)$ as above. Then $f\overline{g} \in \ker(\mathbb{1} - U_T) = \mathrm{fix}(U_T)$. Thus, by ergodicity, we have $f\overline{g} = a\mathbb{1}$ for some $a \in \mathbb{C} \setminus \{0\}$, hence $fd^2 = (f\overline{g})g = ag$. This implies that $\dim\ker(\chi - U_T) \leq 1$.

For (ii) notice that the above observation shows $\sigma_{\mathrm{p}}(U_T) \cdot \sigma_{\mathrm{p}}(U_T)^{-1} \subseteq \sigma_{\mathrm{p}}(U_T)$. Since $U_{T_\gamma}\mathbb{1} = \mathbb{1}$ for every $\gamma \in \Gamma$, the constant one-character $\mathbb{1}\colon \Gamma \to \mathbb{T}, \gamma \mapsto 1$ is an element of $\sigma_{\mathrm{p}}(U_T)$. Thus, $\sigma_{\mathrm{p}}(U_T)$ is indeed a subgroup of Γ^*. □

Lemma 6.2.13. *Let* (X, T) *be an ergodic system with discrete spectrum. Then there is an orthonormal basis* $\{f_\chi \mid \chi \in \sigma_{\mathrm{p}}(U_T)\}$ *of* $\mathrm{L}^2(X)$ *contained in* $\mathrm{L}^\infty(X)$ *such that*

(i) $f_\chi \in \ker(\chi - U_T)$ *for every* $\chi \in \sigma_{\mathrm{p}}(U_T)$*, and*

(ii) $f_{\chi\chi'} = f_\chi f_{\chi'}$ *for all* $\chi, \chi' \in \sigma_{\mathrm{p}}(U_T)$*.*

Proof. Using Lemma 6.2.11 we choose for every eigenvalue $\chi \in \sigma_{\mathrm{p}}(U_T)$ some $g_\chi \in \ker(\chi - U_T)$ with $|g_\chi| = \mathbb{1}$ (and, in particular $\|g_\chi\|_2 = 1$). By Lemma 6.2.12 (i) and Corollary 5.1.11 the linear hull $E \coloneqq \mathrm{lin}\{g_\chi \mid \chi \in \sigma_{\mathrm{p}}(U_T)\}$ is dense in $\mathrm{L}^2(X)$, and thus, by Proposition 5.1.10 (i), the set $\{g_\chi \mid \chi \in \sigma_{\mathrm{p}}(U_T)\}$ is an orthonormal basis. Again using Lemma 6.2.12 (i) we find for $\chi, \chi' \in \sigma_{\mathrm{p}}(U_T)$ some $r(\chi, \chi') \in \mathbb{C}$ with $g_\chi g_{\chi'} = r(\chi, \chi')g_{\chi\chi'}$ (since $g_\chi g_{\chi'} \in \ker(\chi\chi' - U_T)$). Since $|g_\chi| = |g_{\chi'}| = |g_{\chi\chi'}| = \mathbb{1}$, we have $r(\chi, \chi') \in \mathbb{T}$.

Now consider the group $\mathrm{L}^\infty(\mathrm{X}, \mathbb{T}) \coloneqq \{f \in \mathrm{L}^\infty(X) \mid |f| = \mathbb{1}\}$ (with respect to multiplication) and the subgroup $\mathbb{T} \cdot \mathbb{1} \coloneqq \{c\mathbb{1} \mid c \in \mathbb{T}\} \subseteq \mathrm{L}^\infty(\mathrm{X}, \mathbb{T})$. By Lemma 6.2.10, the group homomorphism $\alpha\colon \mathbb{T} \cdot \mathbb{1} \to \mathbb{T}, c\mathbb{1} \mapsto c$ extends to a group homomorphism $\alpha\colon \mathrm{L}^\infty(\mathrm{X}, \mathbb{T}) \to \mathbb{T}$. Then $\alpha(g_\chi)\alpha(g_{\chi'}) = r(\chi, \chi')\alpha(g_{\chi\chi'})$ for all $\chi, \chi' \in \sigma_{\mathrm{p}}(U_T)$. Setting $f_\chi \coloneqq \alpha(g_\chi)^{-1}g_\chi$ for $\chi \in \sigma_{\mathrm{p}}(U_T)$ we therefore obtain

$$f_\chi f_{\chi'} = \alpha(g_\chi)^{-1}\alpha(g_{\chi'})^{-1}r(\chi, \chi')\alpha(g_{\chi\chi'})f_{\chi\chi'} = f_{\chi\chi'}$$

for all $\chi, \chi' \in \sigma_{\mathrm{p}}(U_T)$. Then $\{f_\chi \mid \chi \in \sigma_{\mathrm{p}}(U_T)\}$ is the desired orthonormal basis. □

Proof of Theorems 6.2.6, 6.2.7, and 6.2.8. We start by showing that for every subgroup $W \subseteq \Gamma^*$ we find an ergodic rotation $(G, \mathcal{B}(G), \mathrm{m}_G, (\tau_c)^*)$ with point spectrum $\sigma_{\mathrm{p}}(U_{(\tau_c)^*}) = W$.

To do so we "dualize": The group $G \coloneqq W^*$ consists of all maps $\varrho\colon W \to \mathbb{T}$ with $\varrho(\chi\chi') = \varrho(\chi)\varrho(\chi')$, and hence defines a closed subset of the (by Tychonoff's Theorem 3.2.4 compact) product space $\mathbb{T}^W$, the set of all maps $W \to \mathbb{T}$. With the subspace topology the group G becomes a compact abelian group. Moreover, the "point evaluation map" $c\colon \Gamma \to G$, $\gamma \mapsto \delta_\gamma$ given by $\delta_\gamma(\varrho) := \varrho(\gamma)$ for $\varrho \in G = W^*$ and $\gamma \in \Gamma$ is a group homomorphism.

What is the Pontryagin dual G' of the compact group G? If $\chi \in W$, then the point evaluation $\varepsilon_\chi\colon G \to \mathbb{T}$, $\varrho \mapsto \varrho(\chi)$ is a continuous group homomorphism, hence an element of G'. In this way we obtain a subgroup $\{\varepsilon_\chi \mid \chi \in W\}$ of G' which clearly separates the points of G. By Proposition 6.1.19 this means $G' = \{\varepsilon_\chi \mid \chi \in W\}$.

By Corollary 6.1.17 this implies that the map c has dense range: If $\chi \in W$ satisfies $\varepsilon_\chi(c(\gamma)) = 1$ for all $\gamma \in \Gamma$, then $\chi(\gamma) = \delta_\gamma(\chi) = \varepsilon_\chi(\delta_\gamma) = 1$ for every $\gamma \in \Gamma$, hence $\chi = \mathbb{1}$ (and therefore also $\varepsilon_\chi = \mathbb{1} \in G'$). Hence $c(\Gamma)$ is dense in G which means that the rotation $(G, \mathcal{B}(G), \mathrm{m}_G, (\tau_c)^*)$ is ergodic (cf. Proposition 6.2.4). We further conclude from Proposition 6.2.9 that $\sigma_\mathrm{p}(U_{\tau_c}) = \{\varepsilon_\chi \circ c \mid \chi \in W\} = W$ as desired.

We finally prove the three aspects of the Halmos-von Neumann classification result.

Realization: It is clear from Lemma 6.2.12 (ii) that the point spectrum $\sigma_\mathrm{p}(U_T)$ of an ergodic system (X, T) (with discrete spectrum) is a subgroup of the dual group Γ^*. Conversely, if we start from a subgroup $W \subseteq \Gamma^*$, then by the above observation we find an ergodic rotation $(G, \mathcal{B}(G), \mathrm{m}_G, (\tau_c)^*)$ with point spectrum $\sigma_\mathrm{p}(U_{(\tau_c)^*}) = W$. This has discrete spectrum by Example 6.2.3.

Uniqueness: It is clear that isomorphic systems give rise to the same point spectrum of their Koopman representations. Conversely, let (X, T) and (Y, S) be ergodic measure-preserving systems with discrete spectrum such that $W := \sigma_\mathrm{p}(U_T) = \sigma_\mathrm{p}(U_S)$. Let $(e_\chi)_{\chi\in\sigma}$ and $(f_\chi)_{\chi\in\sigma}$ be the corresponding orthonormal bases of $\mathrm{L}^2(X)$ and $\mathrm{L}^2(Y)$ as in Lemma 6.2.13. By basic linear algebra and Proposition A.1.1 there is a unique unitary operator $V\colon \mathrm{L}^2(X) \to \mathrm{L}^2(Y)$ with $Ve_\chi = f_\chi$ for every $\chi \in W$. By choice of the orthonormal bases, we obtain $V(fg) = (Vf)(Vg)$ for all $f, g \in E$, $V\overline{f} = \overline{Vf}$ for all $f \in E$, and $V\mathbb{1} = Ve_{\mathbb{1}} = f_{\mathbb{1}} = \mathbb{1}$. We obtain from Lemma 2.2.16 that $V|f| = |Vf|$ for every $f \in \mathrm{lin}\{e_\chi \mid \chi \in W\}$, and then even for every $f \in \mathrm{L}^2(X)$ by a density argument. Thus V is a bijective Markov embedding (cf. Exercise 2.8). We show that V also "intertwines the dynamics": For $\gamma \in \Gamma$ we have

$$U_{S_\gamma} V e_\chi = U_{S_\gamma} f_\chi = \chi(\gamma) f_\chi = V(\chi(\gamma) e_\chi) = V U_{T_\gamma} e_\chi$$

for every $\chi \in W$. By linearity and continuity we obtain $U_{S_\gamma} V = V U_{T_\gamma}$ for every $\gamma \in \Gamma$. But this means that the systems (X, T) and (Y, T) are isomorphic, see Exercise 2.7.

Representation: If (X, T) is an ergodic system with discrete spectrum and $W := \sigma_\mathrm{p}(X, T)$ is its point spectrum, we find by the preliminary observation an ergodic rotation $(G, \mathcal{B}(G), \mathrm{m}_G, (\tau_c)^*)$ with $\sigma_\mathrm{p}(U_{(\tau_c)^*}) = W = \sigma_\mathrm{p}(U_T)$. By the uniqueness part, the systems (X, T) and $(G, \mathcal{B}(G), \mathrm{m}_G, (\tau_c)^*)$ are isomorphic. □

6.3 Supplement: Extension of Characters

We follow the arguments of [DE09, Lemma 4.2.13] to show Lemma 6.2.10 in a more general version. Here an abelian group Z is called **divisible** if for every $x \in Z$ and $n \in \mathbb{N}$ there is an "nth root" $y \in Z$, i.e., $y^n = x$. Clearly the torus $\mathbb{T}$ is a divisible group.

Lemma 6.3.1. *Let G be an abelian group, $W \subseteq G$ a subgroup and Z a divisible group. Then every group homomorphism $\alpha\colon W \to Z$ can be extended to a group homomorphism $\alpha\colon G \to Z$.*

Proof. The proof uses Zorn's lemma. Let $\mathcal{M}$ be the set of pairs (H, β) where $H \subseteq G$ is a subgroup containing W and $\beta\colon H \to G$ is a group homomorphism extending H. We obtain an order on $\mathcal{M}$ by setting $(H_1, \beta_1) \leq (H_2, \beta_2)$ for such pairs $(H_1, \beta_1), (H_2, \beta_2) \in \mathcal{M}$ if $H_1 \subseteq H_2$ and $\beta_2|_{H_1} = \beta_1$. Clearly $(W, \alpha) \in \mathcal{M}$, so $\mathcal{M}$ is non-empty. Moreover, if $\mathcal{C}$ is a totally ordered subset of $\mathcal{M}$, then a moment's thought reveals that we obtain an upper bound (A, δ) via $A \coloneqq \bigcup_{(H,\alpha)\in\mathcal{C}} H$ and $\delta(x) \coloneqq \alpha(x)$ for $x \in H$ and $(H, \alpha) \in \mathcal{C}$.

By an application of Zorn's lemma we obtain a maximal element (H, α') of $\mathcal{M}$. We claim that $H = G$, which then finishes the proof. Assuming the contrary, pick $x \in G \setminus H$.

If $x^m \notin H$ for every $m \in \mathbb{N}$, setting $\alpha'(yx^n) \coloneqq \alpha'(y)$ for $y \in H$ and $n \in \mathbb{Z}$ yields an extension of α' to a (well-defined) group homomorphism on the subgroup $\{yx^n \mid y \in H, n \in \mathbb{Z}\}$ which is impossible.

If $x^m \in H$ for some $m \in \mathbb{N}$ we take the smallest $m \in \mathbb{N}$ with this property. One can then easily check that $x^n \in H$ can only hold for multiplies $n = km$ wheren $k \in \mathbb{Z}$. Since Z is divisible, we find some $z \in Z$ with $z^m = \alpha'(x^m)$. Then, if $y_1 x^{n_1} = y_2 x^{n_2}$ for $y_1, y_2 \in H$ and $n_1, n_2 \in \mathbb{Z}$, we have $x^{n_1 - n_2} = y_1^{-1} y_2 \in H$, hence $n_1 - n_2 = km$ for some $k \in \mathbb{Z}$. This implies $\alpha'(y_1^{-1} y_2) = z^{km} = z^{n_1 - n_2}$, and consequently $\alpha'(y_1) z^{n_1} = \alpha'(y_2) z^{n_2}$. Therefore setting $\alpha'(yx^n) \coloneqq \alpha'(y) z^n$ for $y \in H$ and $n \in \mathbb{Z}$ again would give us a well-defined extension to a group homomorphism on the subgroup $\{yx^n \mid y \in H, n \in \mathbb{Z}\}$. $\square$

6.4 Comments and Further Reading

The first part of the lecture treats a (very small) part of abstract harmonic analysis for locally compact (and in particular, compact) groups (see, e.g., [DE09], [Fol15], [HR79] and [HR02]). Our proof that strongly continuous representations of compact groups have discrete spectrum using the JdLG-decomposition is based on [EHK24, Notes and Comments to Part II]. It is closely related to the usual proof of the so-called Peter–Weyl theorem from abstract harmonic analysis (see, e.g., [Fol15, Chapter 5]) which gives a precise description of the left and right regular representation of a compact group in terms of matrix coefficients of irreducible unitary representations.

One of the key features of locally compact abelian groups G is that, by suitably topologizing the dual G', there is a canonical isomorphism between G and its double dual G'' (see, e.g., [DE09, Chapter 3]). The results on compact abelian groups at the end of Section 6.1 can be inferred from this general and powerful result, known as the Pontryagin duality theorem. We implicitly use a special case of this duality in the proof of Theorems 6.2.6, 6.2.7, and 6.2.8.

The classification of ergodic systems with discrete spectrum is one of the corner stones of ergodic structure theory. It was established by Paul Halmos and John von Neumann in [HvN42] for $\Gamma = \mathbb{Z}$, and then later generalized in different directions (see, e.g., the introduction of [HK23] for more information). Here we basically follow the proof by Halmos given in his book, see [Hal56, Page 46–48]. We will apply the representation part of the Halmos–von Neumann classification in Lecture 8.

6.5 Exercises

Exercise 6.1. Show that for a unitary representation $U\colon G \to \mathscr{U}(H)$ of an abelian group G the following assertions are equivalent.

(a) U has discrete spectrum.

(b) H has an orthonormal basis consisting of eigenvectors of U.

Exercise 6.2. Let G be a compact abelian group.

(i) Show that there is a unique Borel probability measure $\mathrm{m}_G \in \mathrm{P}(G)$ with $\mathrm{m}_G(xA) = \mathrm{m}_G(A)$ for every Borel set $A \subseteq G$ and every $x \in G$.
Hint: Use Fubini's theorem to prove uniqueness.

(ii) Show that $\mathrm{m}_G(A) = \mathrm{m}_G(A^{-1})$ for every Borel set $A \subseteq G$.

(iii) Show that $\mathrm{m}_G(O) > 0$ for every non-empty open set $O \subseteq G$.
Hint: Cover G with the sets xO for $x \in G$.

Exercise 6.3. Let $U\colon G \to \mathscr{U}(H)$ be a unitary representation of a topological group G. Show that the following assertions are equivalent.

(a) U is strongly continuous.

(b) U is weakly continuous, i.e., the map $G \to \mathbb{C}$, $x \mapsto (U_x f|g)$ is continuous for all $f, g \in H$.

Exercise 6.4. Prove Lemma 6.1.9.

Exercise 6.5. Let G be a topological group and $W \subseteq G$ a subgroup. Show the following assertions.

(i) The closure $\overline{W}$ is also a subgroup.

(ii) If W is a normal subgroup, then also $\overline{W}$ is a normal subgroup.

(iii) The quotient map $q\colon G \to G/W$ is open, i.e., $q(O)$ is open in G/W for every open subset $O \subseteq G$.

(iv) If W is a normal subgroup, then G/W is a topological group.

Exercise 6.6. Take a unitary representation $U\colon G \to \mathscr{U}(H)$ of a group G. Show that the following assertions are equivalent.

(a) U has discrete spectrum.

(b) There is a subset $D \subseteq H$ such that the linear hull $\mathrm{lin}\, D$ is dense in H and $\overline{\{U_x f \mid x \in G\}}$ is compact for every $f \in D$.

(c) $\overline{\{U_x f \mid x \in G\}}$ is compact for every $f \in H$.

(d) The closure of the image $U(G)$ in $\mathscr{L}(H)$ respect to the strong operator topology is a compact subgroup of $\mathscr{U}(H)$.[3]

You may use the following basic facts about nets in topological spaces (see, e.g., [Sin19, Chapters 4 and 5] for these and further properties).

(i) For a subset $A \subseteq \Omega$ of a topological space Ω and $\omega \in \Omega$ we have $\omega \in \overline{A}$ precisely when there is a net $(\omega_i)_{i\in I}$ in A converging to ω.

(ii) A map $\psi\colon \Omega_1 \to \Omega_2$ between topological spaces is continuous if and only if for every net $(\omega_i)_{i\in I}$ in Ω_1 converging to some $\omega \in \Omega_1$ the net $(\psi(\omega_i))_{i\in I}$ converges to $\psi(\omega)$.

(iii) Every net $(\omega_i)_{i\in I}$ in a compact space K has an accumulation point.

Hint: For "(b) ⇒ (d)" write $\mathscr{S}$ for the closure of $U(G)$ and observe that $\mathscr{S}$ consists of linear isometries. Show that the map

$$\mathscr{S} \to \prod_{f\in D} \overline{\{U_x f \mid x \in G\}}, \quad V \mapsto (Vf)_{f\in D}$$

has closed range and is a homeomorphism onto its range. Use Tychonoff's Theorem 3.2.4 to prove compactness of $\mathscr{S}$, and then use this to show that every element of $\mathscr{S}$ is invertible, hence $\mathscr{S} \subseteq \mathscr{U}(H)$. Finally, apply Exercise 6.5 (i).

Exercise 6.7. Equip $X = [0,1)$ with the Borel σ-algebra and the Lebesgue measure. For $\alpha \in [0,1)$ consider the measure-preserving map $\tau_\alpha\colon X \to X$, $x \mapsto x + \alpha$ from Example 1.1.3 (iii)). Let further $a := \mathrm{e}^{2\pi\mathrm{i}\alpha}$ and consider the rotation map $l_a\colon \mathbb{T} \to \mathbb{T}$, $z \mapsto az$ from Example 6.2.2. Show that $q\colon [0,1) \to \mathbb{T}$, $x \mapsto \mathrm{e}^{2\pi\mathrm{i}x}$ defines an isomorphism between the concrete measure-preserving systems over $\mathbb{Z}$ induced by τ_α and l_a, respectively (cf. Remark 2.1.6).

[3] Recall that for a Hilbert space H the **strong operator topology** on $\mathscr{L}(H)$ is the subspace topology when viewing $\mathscr{L}(H)$ as a subset of H^H (the space of all maps $H \to H$), i.e., it is the smallest topology on $\mathscr{L}(H)$ such that all evaluation maps $\mathscr{L}(H) \to H$, $V \mapsto Vf$ for $f \in H$ become continuous. One can readily check that the group $\mathscr{U}(H)$ of unitary operators on H is a Hausdorff topological group with respect to the strong operator topology.

Lecture 7

In this lecture we first study weakly mixing systems as a counterpart to systems with discrete spectrum from the previous section. In the second part, we then introduce uniquely ergodic topological dynamical systems. This will help us to deduce a famous equidistribution result due to Weyl.

7.1 Weakly Mixing Systems

Recall that for any measure-preserving system (X, T) we obtain, as a special case of Theorem 5.3.4, a decomposition $\mathrm{L}^2(X) = \mathrm{L}^2(X)_{\mathrm{ds}} \oplus \mathrm{L}^2(X)_{\mathrm{wm}}$ for the induced Koopman representation $U_T \colon \Gamma \to \mathscr{U}(\mathrm{L}^2(X))$ of the abelian group Γ. We have already studied systems where the weakly mixing part $\mathrm{L}^2(X)_{\mathrm{wm}}$ of this splitting vanishes. We now consider the other extreme.

Definition 7.1.1. A measure-preserving system (X, T) is called **weakly mixing** if $\mathrm{L}^2(X)_{\mathrm{ds}} = \mathbb{C} \cdot \mathbb{1}$.

Remark 7.1.2. Since the fixed space $\mathrm{fix}(U_T)$ is contained in the discrete spectrum part $\mathrm{L}^2(X)_{\mathrm{ds}}$, every weakly mixing system (X, T) is ergodic by Corollary 2.2.14.

One can deduce the following equivalent descriptions of weak mixing. The proof is left as Exercise 7.2.

Proposition 7.1.3. *Assume that $(F_i)_{i \in I}$ is a Følner net for the abelian group Γ. Then for a measure-preserving system (X, T) the following assertions are equivalent for each $p \in [1, \infty)$.*

(a) (X, T) *is weakly mixing.*

(b) $\lim_i \sup_{\|g\|_2 \leq 1} \frac{1}{|F_i|} \sum_{\gamma \in F_i} |\int_X (U_\gamma f) g - (\int_X f) \cdot (\int_X g)|^p = 0$ *for every* $f \in \mathrm{L}^2(X)$.

(c) $\lim_i \frac{1}{|F_i|} \sum_{\gamma \in F_i} |\int_X (U_\gamma f) g - (\int_X f) \cdot (\int_X g)|^p = 0$ *for all* $f, g \in \mathrm{L}^2(X)$.

(d) $\lim_i \frac{1}{|F_i|} \sum_{\gamma \in F_i} |\int_X (U_\gamma f) \overline{f} - |\int_X f|^2|^p = 0$ *for every* $f \in \mathrm{L}^2(X)$.

(e) $\lim_i \frac{1}{|F_i|} \sum_{\gamma\in F_i} |\mu_X(T_\gamma(A)\cap B) - \mu_X(A)\cdot\mu_X(B)|^p = 0$ *for all* $A, B \in \Sigma(X)$.

Using condition (e) of Proposition 7.1.3 we explaining the terminology. If (X,T) is a weakly mixing system and $A \in \Sigma(X)$, then "asymptotically"[1] we have $\frac{\mu_X(T_\gamma(A)\cap B)}{\mu_X(B)} \approx \mu_X(A)$ for every $B \in \Sigma(X)$ with $\mu_X(B) > 0$. So in the long run the proportion of $T_\gamma(A)$ within any such B is approximately the same, just as, e.g., you find the same proportion of an ingredient in each part of the glass of a well mixed milk shake.

To deduce yet another characterization, we need a concrete description of the Hilbert space tensor product $H\otimes H'$ from Lecture 5 in the case $H = \mathrm{L}^2(X)$ for a probability space X. We start with the dual space H'. For $f \in \mathrm{L}^2(X)$ the notation $\overline{f}$ has two different meanings: We use it for the complex conjugate, but also for the linear functional $\mathrm{L}^2(X) \to \mathbb{C}$, $g \mapsto (g|f)$. The following result gives some justification for this.

Proposition 7.1.4. *Let X be a probability space. Then there is a unique unitary operator $\Phi\colon \mathrm{L}^2(X) \to \mathrm{L}^2(X)'$ sending the complex conjugate $\overline{f}$ to the functional $\overline{f}\colon \mathrm{L}^2(X) \to \mathbb{C}$ for each $f \in \mathrm{L}^2(X)$.*

Proof. Write, for the moment, $\mathrm{C}(f)$ for the complex conjugate of $f \in \mathrm{L}^2(X)$, and $\varphi_f\colon \mathrm{L}^2(X) \to \mathbb{C}$, $g \mapsto (g|f)$ for the induced linear functional. A moment's thought reveals that $\Phi\colon \mathrm{L}^2(X) \to \mathrm{L}^2(X)'$, $f \mapsto \varphi_{\mathrm{C}(f)}$ is the desired unitary map. □

We now investigate the Hilbert space tensor product $\mathrm{L}^2(X_1)\otimes \mathrm{L}^2(X_2)$ for probability spaces X_1 and X_2. For $f_1 \in \mathrm{L}^2(X_1)$ and $f_2 \in \mathrm{L}^2(X_2)$ we obtain an element $f_1 \odot f_2 \in \mathrm{L}^2(X_1\times X_2)$ via $(f_1\odot f_2)(x_1,x_2) := f_1(x_1)f_2(x_2)$ for $x_1 \in X_1$ and $x_2 \in X_2$.

Proposition 7.1.5. *Let X_1 and X_2 be probability spaces. Then there is a unique unitary operator $\Psi\colon \mathrm{L}^2(X_1)\otimes \mathrm{L}^2(X_2) \to \mathrm{L}^2(X_1\times X_2)$ with $\Psi(f_1\otimes f_2) = f_1\odot f_2$ for all $f_1 \in \mathrm{L}^2(X_1)$ and $f_2 \in \mathrm{L}^2(X_2)$.*

The proof of Proposition 7.1.5 is based on the following measure theoretic fact.

Lemma 7.1.6. *The linear hull $\mathrm{lin}\{\mathbb{1}_{A_1\times A_2} \mid A_1 \subseteq X_1, A_2 \subseteq X_2$ measurable$\}$ is dense in $\mathrm{L}^p(X_1\times X_2)$ for all $p \in [1,\infty)$ and all probability spaces X_1 and X_2.*

Since finite disjoint unions of measurable cylinder sets form an algebra over the product space (see, e.g., [HS65, Theorem 21.3]), Lemma 7.1.6 is an easy consequence of Lemmas 1.3.6 and 2.1.12.

Proof of Proposition 7.1.5. By Proposition 5.2.1 we obtain a unique linear map $\Psi\colon \mathrm{L}^2(X_1)\otimes_{\mathrm{vect}} \mathrm{L}^2(X_2) \to \mathrm{L}^2(X_1\times X_2)$ on the vector space tensor product with $\Psi(f_1\otimes f_2) = f_1\odot f_2$ for all $f_1 \in \mathrm{L}^2(X_1)$ and $f_2 \in \mathrm{L}^2(X_2)$. A short computation

[1] One can make this idea precise in other ways obtaining different concepts like *strong* or *mild mixing*.

(similar to the one in the proof of Proposition 5.4.5) shows that Ψ is isometric with respect to the tensor product norm and the L^2-norm. Since $\mathrm{L}^2(X_1) \otimes_{\mathrm{vect}} \mathrm{L}^2(X_2)$ is dense in $\mathrm{L}^2(X_1) \otimes \mathrm{L}^2(X_2)$ by construction and Ψ has dense range by Lemma 7.1.6, an application of Proposition A.1.1 yields the claim. □

We obtain the following consequence.

Corollary 7.1.7. *Consider homomorphisms $T_1 \colon \Sigma(Y_1) \to \Sigma(X_1)$ and $T_2 \colon \Sigma(Y_2) \to \Sigma(X_2)$ between measure algebras of probability spaces. Then there is a unique measure algebra homomorphism $T_1 \times T_2 \colon \Sigma(Y_1 \times Y_2) \to \Sigma(X_1 \times X_2)$ with*

$$(T_1 \times T_2)(A_1 \times A_2) = T_1(A_1) \times T_2(A_2) \text{ for all } A_1 \in \Sigma(Y_1) \text{ and } A_2 \in \Sigma(Y_2).$$

Moreover, $U_{T_1 \times T_2}(f_1 \odot f_2) = U_{T_1} f_1 \odot U_{T_2} f_2$ for all $f_1 \in \mathrm{L}^2(Y_1)$ and $f_2 \in \mathrm{L}^2(Y_2)$.

Proof. Consider the induced Markov embeddings $U_{T_i} \colon \mathrm{L}^2(Y_i) \to \mathrm{L}^2(X_i)$ for $i = 1, 2$. By Propositions 5.2.4 and 7.1.5 there is a unique linear isometry $U \colon \mathrm{L}^2(Y_1 \times Y_2) \to \mathrm{L}^2(X_1 \times X_2)$ with $U(f_1 \odot f_2) = U_{T_1} f_1 \odot U_{T_2} f_2$ for all $f_1 \in \mathrm{L}^2(Y_1)$ and $f_2 \in \mathrm{L}^2(Y_2)$.

We show that U is a Markov embedding. Clearly, $U(\mathbb{1} \odot \mathbb{1}) = \mathbb{1} \odot \mathbb{1}$. Moreover, if $f_1, g_1 \in \mathrm{L}^\infty(Y_1)$ and $f_2, g_2 \in \mathrm{L}^\infty(Y_2)$, then

$$U((f_1 \odot f_2) \cdot (g_1 \odot g_2)) = U_{T_1}(f_1 g_1) \odot U_{T_2}(f_2 g_2) = U(f_1 \odot f_2) \cdot U(g_1 \odot g_2)$$

by Corollary 1.3.8. Similarly to the proof of Theorem 6.2.7 we can use linearity and Lemma 2.2.16 to see that $|Uf| = U|f|$ for all $f \in \mathrm{lin}\{f_1 \odot f_2 \mid f_1 \in \mathrm{L}^\infty(Y_1), f_2 \in \mathrm{L}^\infty(Y_2)\}$. Then use Lemma 7.1.6 to still obtain this equality for all $f \in \mathrm{L}^2(Y_1 \times Y_2)$. Thus U is indeed a Markov embedding.

By Theorem 1.3.7 we find a unique measure algebra homomorphism $T_1 \times T_2 \colon \Sigma(Y_1 \times Y_2) \to \Sigma(X_1 \times X_2)$ with $U = U_{T_1 \times T_2}$. By definition of U we have $U_{T_1 \times T_2}(f_1 \odot f_2) = U_{T_1} f_1 \odot U_{T_2} f_2$ for all $f_1 \in \mathrm{L}^2(Y_1)$ and $f_2 \in \mathrm{L}^2(Y_2)$. In particular, we obtain

$$\begin{aligned} \mathbb{1}_{T_1 \times T_2(A_1 \times A_2)} &= U_{T_1 \times T_2}(\mathbb{1}_{A_1} \odot \mathbb{1}_{A_2}) = U_{T_1} \mathbb{1}_{A_1} \odot U_{T_2} \mathbb{1}_{A_2} = \mathbb{1}_{T_1(A_1)} \odot \mathbb{1}_{T_2(A_2)} \\ &= \mathbb{1}_{T_1(A_1) \times T_2(A_2)}, \end{aligned}$$

hence $(T_1 \times T_2)(A_1 \times A_2) = T_1(A_1) \times T_2(A_2)$ for all $A_1 \in \Sigma(Y_1)$ and $A_2 \in \Sigma(Y_2)$. Finally, the claimed uniqueness is an easy consequence of Lemma 7.1.6 and uniqueness of U. □

Remark 7.1.8. This construction is compatible with the product of measure-preserving maps from Exercise 2.1: If $\tau_1 \colon X_1 \to Y_1$ and $\tau_2 \colon X_2 \to Y_2$ are measure-preserving maps between probability spaces, then $(\tau_1^* \times \tau_2^*) = (\tau_1 \times \tau_2)^*$.

Corollary 7.1.7 allows us to introduce products of measure-preserving systems.

Definition 7.1.9. The **product system** $(X_1 \times X_2, T_1 \times T_2)$ of measure-preserving systems (X_1, T_1) and (X_2, T_2) is defined by $(T_1 \times T_2)_\gamma \coloneqq (T_1)_\gamma \times (T_2)_\gamma$ for $\gamma \in \Gamma$.

We prove the following characterization of weak mixing via product systems.

Proposition 7.1.10. *For a measure-preserving system (X, T) the following assertions are equivalent.*

(a) *The system (X, T) is weakly mixing.*

(b) *The product system $(X \times X, T \times T)$ is ergodic.*

Proof. Note first that, as a consequence or Propositions 7.1.4 and 7.1.7, there is a unique unitary operator $V\colon \mathrm{L}^2(X) \otimes \mathrm{L}^2(X)' \to \mathrm{L}^2(X \times X)$ with $V(f \otimes \overline{g}) = f \odot \overline{g}$ for all $f, g \in \mathrm{L}^2(X)$. Moreover, we have $V \circ (U_{T_\gamma} \otimes \overline{U_{T_\gamma}}) = U_{T_\gamma \times T_\gamma} \circ V$ for every $\gamma \in \Gamma$ (check this identity on simple tensors).

For "(a) $\Rightarrow$ (b)" assume that (X, T) is weakly mixing, i.e., $\mathrm{L}^2(X)_{\mathrm{ds}} = \mathbb{C} \cdot \mathbb{1}$. By Theorem 5.3.1 we obtain that $\mathrm{fix}(U_T \otimes \overline{U_T}) = \mathbb{C} \cdot (\mathbb{1} \otimes \mathbb{1})$ in $\mathrm{L}^2(X) \otimes \mathrm{L}^2(X)'$. But then $\mathrm{fix}(U_{T \times T}) = V(\mathrm{fix}(U_T \otimes \overline{U_T})) = \mathbb{C} \cdot (\mathbb{1} \odot \mathbb{1})$, hence $(X \times X, T \times T)$ is ergodic.

Now prove "(b) $\Rightarrow$ (a)" by contraposition. If (X, T) is not weakly mixing, we find some eigenvector $e \in \mathrm{L}^2(X)$ with respect to U_T with $(e|\mathbb{1}) = 0$ (see Proposition 5.1.10 (i) and Corollary 5.1.11). But then $e \otimes \overline{e} \in \mathrm{fix}(U_T \otimes \overline{U_T})$ (see Theorem 5.3.1) with $(e \otimes \overline{e}|\mathbb{1} \otimes \mathbb{1}) = |(e|\mathbb{1})|^2 = 0$. Thus, $e \odot \overline{e} = V(e \otimes \overline{e}) \in \mathrm{fix}(U_{T \times T})$ is orthogonal to $V(\mathbb{1} \otimes \mathbb{1}) = \mathbb{1} \in \mathrm{L}^2(X \times X)$. Therefore $(X \times X, T \times T)$ is not ergodic. □

Example 7.1.11. For an infinite abelian group Γ consider the Bernoulli shift (X^Γ, τ^*) defined by a probability space X from Example 2.1.8. One can readily check that the map

$$q\colon X^\Gamma \times X^\Gamma \to (X \times X)^\Gamma, \quad ((x_\gamma)_{\gamma \in \Gamma}, (y_\gamma)_{\gamma \in \Gamma}) \mapsto ((x_\gamma, y_\gamma))_{\gamma \in \Gamma}$$

induces an isomorphism between the product system $(X^\Gamma \times X^\Gamma, \tau^* \times \tau^*)$ and the Bernoulli shift $((X \times X)^\Gamma, \tau^*)$ on the product space $X \times X$. Thus, the system $(X^\Gamma \times X^\Gamma, \tau^* \times \tau^*)$ is ergodic by Proposition 2.1.11, and therefore the Bernoulli shift (X^Γ, τ^*) is weakly mixing.

We conclude this section by proving a dichotomy between systems with discrete spectrum and weakly mixing systems. For this we first establish the following result using the concept of invariant Markov sublattices from Definition 2.2.9.

Proposition 7.1.12. *Let (X, T) be a measure-preserving system. Then $\mathrm{L}^2(X)_{\mathrm{ds}}$ is an invariant Markov sublattice of $\mathrm{L}^2(X)$.*

We use the following lemma (cf. Lemma 6.2.12 for ergodic systems).

Lemma 7.1.13. *Let (X, T) be a measure-preserving system. For every $\chi \in \Gamma^*$ the space $\ker(\chi - U_T) \cap \mathrm{L}^\infty(X)$ is dense in $\ker(\chi - U_T)$.*

Proof. Let $f \in \ker(\chi - U_T)$ for some $\chi \in G^*$. Then $|f| \in \operatorname{fix}(U_T)$ by Lemma 6.2.11. Since $\operatorname{fix}(U_T)$ is an invariant Markov sublattice, we obtain (similarly to the proof of Proposition 2.2.13) that

$$\mathbb{1}_{[|f|\leq n]} = \mathbb{1} - \mathbb{1}_{[|f|-n\mathbb{1}>0]} = \lim_{m\to\infty}(\mathbb{1} - \inf(m\sup(|f| - n\mathbb{1}, 0), \mathbb{1})) \in \operatorname{fix}(U_T)$$

for every $n \in \mathbb{N}$. By Corollary 1.3.8 we have $U_{T_\gamma}(gh) = (U_{T_\gamma}g)\cdot(U_{T_\gamma}h)$ for all $g, h \in \mathrm{L}^\infty(X)$, and by approximation this still holds if $h \in \mathrm{L}^2(X)$. This implies $U_{T_\gamma}(\mathbb{1}_{[|f|\leq n]}f) = \chi(\gamma)\mathbb{1}_{[|f|\leq n]}f$ for every $\gamma \in \Gamma$. Therefore $\mathbb{1}_{[|f|\leq n]}f \in \ker(\chi - U_T) \cap \mathrm{L}^\infty(X)$ for each $n \in \mathbb{N}$ and clearly $f = \lim_{n\to\infty} \mathbb{1}_{[|f|\leq n]}f$ in $\mathrm{L}^2(X)$. □

Proof of Proposition 7.1.12. Consider $F := \operatorname{lin}\bigcup_{\chi\in\Gamma^*}\ker(\chi - U_T) \cap \mathrm{L}^\infty(X)$. We claim that F has the following properties, which, by Proposition 2.2.15, yield that $\mathrm{L}^2(X)_{\mathrm{ds}}$ is indeed an invariant Markov sublattice of $\mathrm{L}^2(X)$.

(i) $f \cdot g \in F$ for all $f, g \in F$,

(ii) $\overline{f} \in F$ for all $f \in F$,

(iii) $\mathbb{1} \in F$,

(iv) $U_{T_\gamma}f \in F$ for all $f \in F$, $\gamma \in \Gamma$, and

(v) F is dense in $\mathrm{L}^2(X)_{\mathrm{ds}}$ (with respect to the L^2-norm).

For (i) first take $f \in \ker(\chi - U_T) \cap \mathrm{L}^\infty(X)$ and $g \in \ker(\varrho - U_T) \cap \mathrm{L}^\infty(X)$ for some $\chi, \varrho \in \Gamma^*$. Then $U_{T_\gamma}(fg) = \chi(\gamma)\varrho(\gamma)fg$ for all $\gamma \in \Gamma$, hence $fg \in F$. Using a linearity argument, we obtain (i). Parts (ii) – (iv) are proved in a similar manner and (v) is a direct consequence of Corollary 5.1.11 and Lemma 7.1.13. □

In view of Proposition 2.2.13 the discrete spectrum part $\mathrm{L}^2(X)_{\mathrm{ds}}$ of a measure-preserving system (X, T) gives rise to a subsystem of (X, T).

Definition 7.1.14. Let (X, T) be a measure-preserving system. We then write $J_{\mathrm{kro}}\colon (X_{\mathrm{kro}}, T_{\mathrm{kro}}) \to (X, T)$ for the extension $J_E\colon (X_E, T_E) \to (X, T)$ defined by the invariant Markov sublattice $E = \mathrm{L}^2(X)_{\mathrm{ds}}$ and call $(X_{\mathrm{kro}}, T_{\mathrm{kro}})$ the **Kronecker subsystem** of (X, T).

Remark 7.1.15. Given a measure-preserving system (X, T) the Kronecker subsystem is maximal among all subsystems of (X, T) with discrete spectrum: If $J\colon (Y, S) \to (X, T)$ is any extension such that (Y, S) has discrete spectrum, then $U_J(\mathrm{L}^2(Y)) \subseteq \mathrm{L}^2(X)_{\mathrm{ds}} = U_{J_{\mathrm{kro}}}(\mathrm{L}^2(X_{\mathrm{kro}}))$. The Markov embedding $U := U_{J_{\mathrm{kro}}} \circ U_J\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X_{\mathrm{kro}})$ then induces an extension $(Y, S) \to (X_{\mathrm{kro}}, T_{\mathrm{kro}})$ (see Exercise 2.7), hence (Y, S) is a subsystem of $(X_{\mathrm{kro}}, T_{\mathrm{kro}})$.

The proof of the following dichotomy between weak mixing and discrete spectrum is now easy. Recall the definition of the trivial system $(\{0\}, \mathrm{Id})$ from Example 2.1.2.

Theorem 7.1.16 (Dichotomy Between Discrete Spectrum and Weak Mixing)**.** *For a measure-preserving system (X, T) exactly one of the following two alternatives is true.*

(1) *(X, T) is weakly mixing.*

(2) *(X, T) has a subsystem (Y, S) with discrete spectrum which is not isomorphic to the trivial system $(\{0\}, \mathrm{Id})$.*

Proof. In view of Remark 7.1.15 assertion (2) is equivalent to $(X_{\mathrm{kro}}, T_{\mathrm{kro}})$ not being isomorphic to the trivial system $(\{0\}, \mathrm{Id})$. This is the case precisely when $\mathrm{L}^2(X_{\mathrm{kro}}) \neq \mathbb{C} \cdot \mathbb{1}$, see Exercise 2.7. Thus (2) is equivalent to $U_{J_{\mathrm{kro}}}(\mathrm{L}^2(X_{\mathrm{kro}})) = \mathrm{L}^2(X)_{\mathrm{ds}} \neq \mathbb{C} \cdot \mathbb{1}$, which means that (X, T) is not weakly mixing. □

7.2 Uniquely Ergodic Systems

In Proposition 6.2.4 we have seen that a rotation system (G, τ_c) defined by a group homomorphism $c\colon \Gamma \to G$ to a compact abelian group G with dense range has precisely one invariant measure $\mu \in \mathrm{P}(G, \tau_c)$ (namely the Haar measure $\mu = \mathrm{m}_G$). We introduce a name for topological dynamical systems with this property.

Definition 7.2.1. A topological dynamical system (K, τ) is **uniquely ergodic** if $\mathrm{P}(K, \tau)$ has exactly one element.

The following is a characterization in terms of mean convergence.

Proposition 7.2.2. *Let $(F_i)_{i \in I}$ be a Følner net for the abelian group Γ. For a topological dynamical system (K, τ) the following assertions are equivalent.*

(a) *(K, τ) is uniquely ergodic.*

(b) *For every $f \in \mathrm{C}(K)$ the net $(\frac{1}{|F_i|} \sum_{\gamma \in F_i} f \circ \tau_\gamma)_{i \in I}$ converges with respect to the supremum norm to a constant function.*

If (a) and (b) hold, then the limit in (b) for $f \in \mathrm{C}(K)$ is given by $(\int_K f \,\mathrm{d}\mu) \cdot \mathbb{1}$ where μ is the unique element of $\mathrm{P}(K, \tau)$.

The proof uses the following basic topological lemma (cf. [Sin19, Section 4.3 and Exercise 5.18]).

Lemma 7.2.3. *If a net $(x_i)_{i \in I}$ in a compact space K has precisely one accumulation point $x \in K$, then $\lim_{i \in I} x_i = x$.*

Proof of Proposition 7.2.2. We first show the implication "(a) ⇒ (b)" and the assertion about the limit. So let (K, τ) be uniquely ergodic with invariant measure $\mu \in \mathrm{P}(K, \tau)$. Assume that there is some $f \in \mathrm{C}(K)$ such that the net $(1/|F_i| \sum_{\gamma \in F_i} f \circ \tau_\gamma)_{i \in I}$ does not converge to $(\int_K f \,\mathrm{d}\mu) \cdot \mathbb{1}$ in $\mathrm{C}(K)$ with respect to the supremum norm. This means that we can find some $\varepsilon > 0$ and for each $i \in I$ some $j(i) \geq i$ as well as a point $x_i \in K$ with

$$\left| \frac{1}{|F_{j(i)}|} \sum_{\gamma \in F_{j(i)}} f(\tau_\gamma(x_i)) - \int_K f \,\mathrm{d}\mu \right| \geq \varepsilon. \tag{7.1}$$

A moment's thought reveals, as in Example 3.2.13, that the net $(\mu_i)_{i\in I}$ in $\mathrm{P}(K)$ given by $\mu_i := \frac{1}{|F_{j(i)}|} \sum_{\gamma \in F_{j(i)}} (\tau_\gamma)_* \delta_{x_i}$ for $i \in I$ is asymptotically invariant, hence all its limit points are elements of $\mathrm{P}(K, \tau)$ by Proposition 3.2.14. Thus, by Lemma 7.2.3, the net $(\mu_i)_{i\in I}$ converges to μ with respect to the weak* topology. But then, in particular,

$$\lim_{i\in I} \frac{1}{|F_{j(i)}|} \sum_{\gamma \in F_{j(i)}} f(\tau_\gamma(x_i)) = \lim_{i\in I} \frac{1}{|F_{j(i)}|} \sum_{\gamma \in F_{j(i)}} ((\tau_\gamma)_* \delta_{x_i})(f) = \int_K f \,\mathrm{d}\mu,$$

contradicting inequality (7.1).

Now assume conversely that (b) holds. For every $f \in \mathrm{C}(K)$ let $c_f \in \mathbb{C}$ with $\lim_{i\in I} {}^1/_{|F_i|} \sum_{\gamma\in F_i} f \circ \tau_\gamma = c_f \cdot \mathbb{1}$. For $\mu \in \mathrm{P}(K, \tau)$ and $f \in \mathrm{C}(K)$ we obtain $\int_K f \,\mathrm{d}\mu = \int_K \frac{1}{|F_i|} \sum_{\gamma\in F_i} f \circ \tau_\gamma \,\mathrm{d}\mu$ for every $i \in I$. Since μ defines a continuous linear functional on $\mathrm{C}(K)$, this implies

$$\mu(f) = \lim_{i\in I} \int_K \frac{1}{|F_i|} \sum_{\gamma\in F_i} f \circ \tau_\gamma \,\mathrm{d}\mu = \int_K c_f \cdot \mathbb{1} \,\mathrm{d}\mu = c_f$$

for every $f \in \mathrm{C}(K)$. Thus, $\mathrm{P}(K, \tau)$ can contain only a single element. □

For further examples of uniquely ergodic systems we first introduce the following concept.

Definition 7.2.4. Let (L, σ) be a topological dynamical system and G a compact group. A continuous map $c\colon \Gamma \times L \to G$ is called a **continuous cocycle** if

$$c(\gamma_1 + \gamma_2, l) = c(\gamma_1, \sigma_{\gamma_2}(l)) \cdot c(\gamma_2, l) \quad \text{for all } \gamma_1, \gamma_2 \in \Gamma \text{ and } l \in L.$$

In this case, we call the topological dynamical system $(L \times G, \sigma \rtimes c)$ defined by $(\sigma \rtimes c)_\gamma(l, x) := (\sigma_\gamma(l), c(\gamma, l)x)$ for $l \in L$, $x \in G$ and $\gamma \in \Gamma$ a **skew-rotation system**.

As an easy consequence of the definition, a continuous cocycle $c\colon \Gamma \times L \to G$ as above satisfies $c(0, l) = 1$ for every $l \in L$ and $c(-\gamma, l) = c(\gamma, \sigma_{-\gamma}(l))^{-1}$ for all $l \in L$ and $\gamma \in \Gamma$. Moreover, it is a simple exercise to check that the skew-rotation system induced by c is indeed a topological dynamical system.

Observe further that the definition of skew-rotation systems reduces to rotation systems from Example 6.2.2 if we take (L, σ) to be a trivial system on $L = \{0\}$.

Remark 7.2.5. If (L,σ) is a topological dynamical system over $\Gamma = \mathbb{Z}$, then a continuous cocycle $c\colon \Gamma \times L \to G$ to a compact group G is uniquely determined by the continuous map $\tilde{c} := c(1,\cdot)\colon L \to G$ since the cocycle rule implies

$$c(m,l) = \begin{cases} \tilde{c}(\sigma_{m-1}(l)) \cdot \tilde{c}(\sigma_{m-2}(l)) \cdots \tilde{c}(\sigma_1(l)) \cdot \tilde{c}(l) & \text{if } m \geq 1, \\ 1 & \text{if } m = 0, \\ \tilde{c}(\sigma_{-m}(l))^{-1} \cdot \tilde{c}(\sigma_{-(m-1)}(l))^{-1} \cdots \tilde{c}(\sigma_{-1}(l))^{-1} & \text{if } m \leq -1, \end{cases}$$

for every $l \in L$. This establishes a one-to-one correspondence between continuous cocycles $c\colon \mathbb{Z} \times L \to G$ and continuous maps $\tilde{c}\colon L \to G$.

The following is a standard example of a skew-rotation system.

Example 7.2.6. Consider the torus rotation $(\mathbb{T}, \tau_a)$ for $a \in \mathbb{T}$ from Example 6.2.2. Then with $G = \mathbb{T}$ and the continuous cocycle defined by the identity map $\tilde{c} := \mathrm{id}_{\mathbb{T}}\colon \mathbb{T} \to \mathbb{T}$ we obtain a skew-rotation $(\mathbb{T}^2, \tau_a \rtimes c)$ on the 2-torus $\mathbb{T}^2$. It is given by the homeomorphism $\mathbb{T}^2 \to \mathbb{T}^2$, $(x,y) \mapsto (ax, xy)$.

We now consider invariant measures on skew-rotation systems. Given a topological dynamical system (L,σ), a continuous cocycle $c\colon \Gamma \times L \to G$ to a compact group G and an invariant measure $\nu \in \mathrm{P}(L,\sigma)$, a natural candidate would be the product measure $\nu \otimes \mathrm{m}_G$ with the Haar measure of G (see Definitions 2.1.7 and 6.1.3). There is a small technical difficulty here since the product σ-algebra $\mathcal{B}(L) \otimes \mathcal{B}(G)$ might be strictly smaller than the Borel σ-algebra $\mathcal{B}(L \times G)$ (and hence the product measure is not a Borel measure). However, one can fix this issue (see [Fol99, Section 7.4]).

Proposition 7.2.7. *Let $\nu \in \mathrm{P}(L)$ and $\mu \in \mathrm{P}(K)$ for compact spaces K and L. Then there is a unique regular Borel probability measure $\nu \overline{\otimes} \mu\colon \mathcal{B}(L \times K) \to [0,1]$ with $(\nu \overline{\otimes} \mu)|_{\mathcal{B}(L) \otimes \mathcal{B}(K)} = \nu \otimes \mu$.*

Remarks 7.2.8.
(a) If the compact spaces L and K are metrizable (or, equivalently, second countable, see, e.g., [AB06, Theorem 3.40]), then $\mathcal{B}(L \times K) = \mathcal{B}(L) \otimes \mathcal{B}(K)$ and thus $\nu \overline{\otimes} \mu = \nu \otimes \mu$, see again [Fol99, Section 7.4].

(b) Write $(f \odot g)(x,y) := f(x)g(y)$ for $(x,y) \in L \times K$ where $f \in \mathrm{C}(L)$ and $g \in \mathrm{C}(K)$. By an application of the Stone-Weierstraß theorem (see Theorem 6.1.18) the functions $f \odot g$ for $f \in \mathrm{C}(L)$ and $g \in \mathrm{C}(K)$ span a dense subspace of $\mathrm{C}(L \times K)$. This implies that $\nu \overline{\otimes} \mu \in \mathrm{P}(L \times K)$ can also be characterized as the unique measure $\varrho \in \mathrm{P}(L \times K)$ satisfying $\int_{L \times K} f \odot g \,\mathrm{d}\varrho = \int_L f \,\mathrm{d}\nu \cdot \int_K g \,\mathrm{d}\mu$ for all $f \in \mathrm{C}(L)$ and $g \in \mathrm{C}(K)$.

The following is easy to check.

Proposition 7.2.9. *Let (L,σ) be a topological dynamical system, $c\colon \Gamma \times L \to G$ a continuous cocycle to a compact group G and $\nu \in \mathrm{P}(L,\sigma)$ an invariant measure.*

Then $\nu\overline{\otimes}\mathrm{m}_G \in \mathrm{P}(L \times G)$ *is an invariant measure for the skew rotation system* $(L \times G, \sigma \rtimes c)$.

Notice that, if we write $\mathrm{pr}_L\colon L \times G \to L$, $(l,x) \mapsto l$ for the projection onto the first component, then $(\mathrm{pr}_L)_*(\nu\overline{\otimes}\mathrm{m}_G) = \nu$. This leads us to the following uniqueness property.

Theorem 7.2.10. *Let* (L,σ) *be a topological dynamical system,* $c\colon \Gamma \times L \to G$ *a continuous cocycle to a compact group* G *and* $\nu \in \mathrm{P}(L,\sigma)$. *If the measure* $\nu\overline{\otimes}\mathrm{m}_G$ *is ergodic, then it is the only invariant measure* $\mu \in \mathrm{P}(L \times G, \sigma \rtimes c)$ *with* $(\mathrm{pr}_L)_*\mu = \nu$.

Proof. Fix functions $f \in \mathrm{C}(L)$ and $g \in \mathrm{C}(G)$ and consider the product function $f \odot g \in \mathrm{C}(L \times G)$ from Remark 7.2.8 (b). By ergodicity of $\nu\overline{\otimes}\mathrm{m}_G$ and the abstract mean ergodic theorem (see Theorem 3.1.5 (ii)) we find $c \in \mathbb{C}$ and $V_n = \sum_{j=1}^{k_n} t_{n,j} U_{(\sigma\rtimes c)_{\gamma_{n,j}}} \in \mathrm{co}\{U_{(\sigma \rtimes c)_\gamma} \mid \gamma \in \Gamma\}$ for every $n \in \mathbb{N}$ such that $\lim_{n\to\infty} V_n(f \odot g) = c\mathbb{1}$ in $\mathrm{L}^2(L \times G, \mathcal{B}(L \times G), \nu\overline{\otimes}\mathrm{m}_G)$. Integrating yields $c = \int_L f\,\mathrm{d}\mu \cdot \int_G g\,\mathrm{dm}_G$ (cf. Exercise 3.4).

We may assume, by passing to a subsequence, that $(V_n(f \odot g))_{n\in\mathbb{N}}$ converges almost everywhere with respect to $\nu\overline{\otimes}\mathrm{m}_G$ to $(\int_L f\,\mathrm{d}\mu)(\int_G g\,\mathrm{dm}_G)\mathbb{1}$. We therefore obtain that the Borel measurable set

$$A := \left\{(l,x) \in L \times G \,\middle|\, \lim_{n\to\infty} \sum_{j=1}^{k_n} t_{n,j} f(\sigma_{\gamma_j}(l)) g(c(l,\sigma_{\gamma_j}(l))x) = \int_L f\,\mathrm{d}\nu \cdot \int_G g\,\mathrm{dm}_G \right\}$$

has full measure with respect to $\nu\overline{\otimes}\mathrm{m}_G$.

For $y \in G$ the map $q_y\colon L \times G \to L \times G$, $(l,x) \mapsto (l, xy^{-1})$ is a homeomorphism and measure-preserving with respect to $\nu\overline{\otimes}\mathrm{m}_G$. It maps A to the set

$$A_y := \left\{(l,x) \in L \times G \,\middle|\, \lim_{n\to\infty} \sum_{j=1}^{k_n} t_{n,j} f(\sigma_{\gamma_j}(l)) g(c(l,\sigma_{\gamma_j}(l))xy) = \int_L f\,\mathrm{d}\nu \cdot \int_G g\,\mathrm{dm}_G \right\}.$$

Thus, A_y is also Borel measurable with full measure for each $y \in G$. We now check that the same holds true for the intersection $B := \bigcap_{y\in G} A_y$.

To do so, note first that the map $G \to \mathrm{C}(G)$, $y \mapsto g \circ r_y$ (where $r_y(x) = xy$ for $x, y \in G$) is continuous with respect to the supremum norm by Lemma 6.1.9. Thus, since G is compact, also the image $\{g \circ r_y \mid y \in G\} \subseteq \mathrm{C}(G)$ is compact. In particular, since compact metric spaces are separable (see, e.g., [AB06, Section 3.7]), we find a countable subset $D \subseteq G$ such that $\{g \circ r_y \mid y \in D\}$ is dense in $\{g \circ r_y \mid y \in G\}$. We claim that $B = \bigcap_{y\in D} A_y$, which then implies that B is indeed measurable with full measure.

So take $z \in G$ and pick a sequence $(z_m)_{m\in\mathbb{N}}$ in D with $\lim_{m\to\infty} \|g \circ r_z - g \circ r_{z_m}\|_\infty = 0$.

Then

$$\lim_{m\to\infty}\sup_{(l,x)}\left|\sum_{j=1}^{k_n} t_{n,j} f(\sigma_{\gamma_j}(l))g(c(l,\sigma_{\gamma_j}(l))xz) - \sum_{j=1}^{k_n} t_{n,j} f(\sigma_{\gamma_j}(l))g(c(l,\sigma_{\gamma_j}(l))xz_m)\right| = 0,$$

and a standard application of the triangle inequality yields $\bigcap_{n\in\mathbb{N}} A_{z_n} \subseteq A_z$. This implies $\bigcap_{y\in D} A_y \subseteq A_z$ for every $z \in G$, hence $B = \bigcap_{y\in D} A_y$ as desired.

Since the map $L \to L \times G$, $l \mapsto (l, 1)$ is continuous, hence Borel measurable, we obtain that the "section" $B_1 := \{l \in L \mid (l, 1) \in B\}$ is Borel measurable as well. Moreover, for $l \in L$ one can readily check that $(l, 1) \in B$ precisely when $(l, y) \in B$ for some $y \in G$. Thus, $\mathrm{pr}_L^{-1}(B_1) = B$ and consequently $\nu(B_1) = (\nu\overline{\otimes}\mathrm{m}_G)(\mathrm{pr}_L^{-1}(B_1)) = (\nu\overline{\otimes}\mathrm{m}_G)(B) = 1$.

Finally, take an invariant measure $\mu \in \mathrm{P}(L \times G, \sigma \rtimes c)$ with $(\mathrm{pr}_L)_*\mu = \nu$. Then also $\mu(B) = \mu(\mathrm{pr}_L^{-1}(B_1)) = (\mathrm{pr}_L)_*\mu(B_1) = \nu(B_1) = 1$. Thus, we have

$$\lim_{n\to\infty}\sum_{j=1}^{k_n} t_{n,j} f(\sigma_{\gamma_j}(l))g(c(l,\sigma_{\gamma_j}(l))x) = \int_L f\,\mathrm{d}\nu \cdot \int_G g\,\mathrm{dm}_G$$

for almost every $(l, x) \in L \times G$ with respect to the measure μ. Using Lebesgue's theorem and invariance of the measure μ, we obtain by integrating on both sides against μ that $\int_{L\times G} f \odot g\,\mathrm{d}\mu = \int_L f\,\mathrm{d}\nu \cdot \int_G g\,\mathrm{dm}_G$. In view of Remark 7.2.8 (b) this implies the desired equality $\mu = \nu\overline{\otimes}\mathrm{m}_G$. □

Corollary 7.2.11 (Furstenberg)**.** *Let (L, σ) be a uniquely ergodic system with invariant measure ν and $c\colon \Gamma \times L \to G$ a continuous cocycle to a compact group G. If $\nu\overline{\otimes}\mathrm{m}_G$ is ergodic, then the skew rotation $(L \times G, \sigma \rtimes c)$ is uniquely ergodic.*

7.3 Weyl's Equidistribution Theorem

As an application of the above we study equidistribution of sequences in the torus.

Definition 7.3.1. We say that a sequence $(a_n)_{n\in\mathbb{N}}$ in $\mathbb{T}$ is **equidistributed** if

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(a_n) = \int_{\mathbb{T}} f\,\mathrm{dm}_{\mathbb{T}} = \int_0^1 f(\mathrm{e}^{2\pi\mathrm{i}t})\,\mathrm{d}t$$

for every $f \in \mathrm{C}(\mathbb{T})$.

Equivalent definitions are discussed in Exercise 7.3.

Theorem 7.3.2 (Weyl)**.** *Let $p = \sum_{k=0}^{d} c_k t^k \in \mathbb{R}[t]$ be a real polynomial with at least one irrational coefficient c_k for $k \in \{1, \ldots, d\}$. Then the sequence $(\mathrm{e}^{2\pi\mathrm{i}p(n)})_{n\in\mathbb{N}}$ is equidistributed.*

We first establish the following lemma.

Lemma 7.3.3. *Assume that* $a \in \mathbb{T}$ *is not a root of unity and let* $k \in \mathbb{N}$*. Then the topological dynamical system over* $\mathbb{Z}$ *defined by the homeomorphism*

$$\tau_k\colon \mathbb{T}^k \to \mathbb{T}^k, \quad (z_1, \dots, z_k) \mapsto (az_1, z_1z_2, z_2z_3, \dots, z_{k-1}z_k)$$

is uniquely ergodic with invariant measure $\mathrm{m}_{\mathbb{T}^k} = \mathrm{m}_{\mathbb{T}} \otimes \cdots \otimes \mathrm{m}_{\mathbb{T}}$.

Proof. Observe that for $k = 1$ the system is just the torus rotation from Example 6.2.2 and hence uniquely ergodic (see Example 6.2.5). We now prove the result via induction on $k \in \mathbb{N}$. Notice that for $k \in \mathbb{N}$ the system $(\mathbb{T}^{k+1}, \tau_{k+1})$ is a skew-rotation defined by the system $(\mathbb{T}^k, \tau_k)$ and the continuous cocycle induced by the map $\tilde{c}\colon \mathbb{T}^k \to \mathbb{T}$, $(z_1, \dots, z_k) \mapsto z_k$ (cf. Remark 7.2.5). By Corollary 7.2.11 it therefore suffices to show that $(\mathbb{T}^k, \mathcal{B}(\mathbb{T}^k), \mathrm{m}_{\mathbb{T}^k}, \tau_k^*)$ is ergodic for each $k \in \mathbb{N}$.

So take $f \in \mathrm{fix}(U_{\tau_k}) \subseteq \mathrm{L}^2(\mathbb{T}^k)$. Using Proposition 6.1.19 it is easy to check (as in Example 6.1.21) that the dual group $(\mathbb{T}^k)'$ consists of the (pairwise distinct) continuous characters $\chi_m\colon \mathbb{T}^k \to \mathbb{T}$ for $m = (m_1, \dots, m_k) \in \mathbb{Z}^k$ given by $\chi_m(z) := z_1^{m_1} \cdots z_k^{m_k}$ for every $z = (z_1, \dots, z_k) \in \mathbb{T}^k$, and these form an orthonormal basis. Hence we can consider the Fourier series expansion $f = \sum_{m\in\mathbb{Z}^k}(f|\chi_m)\chi_m$ of f in $\mathrm{L}^2(\mathbb{T}^k)$ (see Theorem A.2.5). We abbreviate $b_m := (f|\chi_m)$ for $m \in \mathbb{Z}^k$ and show that $b_m = 0$ for each $m \neq 0$, hence $f = b_0\mathbb{1} \in \mathbb{C} \cdot \mathbb{1}$.

Since

$$(\chi_m \circ \tau_k)(z) = (az_1)^{m_1}(z_1z_2)^{m_2} \cdots (z_{k-1}z_k)^{m_k} = a^{m_1}\chi_{m_1+m_2,m_2+m_3,\dots,m_{k-1}+m_k,m_k}(z)$$

for $z = (z_1, \dots, z_k) \in \mathbb{T}^k$ and $m = (m_1, \dots, m_k) \in \mathbb{Z}^k$, we obtain

$$U_{\tau_k} f = \sum_{m_1,\dots,m_k\in\mathbb{Z}} b_{m_1+m_2,\dots,m_{k-1}+m_k,m_k} a^{m_1} \chi_{m_1+m_2,\dots,m_{k-1}+m_k,m_k}.$$

The map $\varphi\colon \mathbb{Z}^k \to \mathbb{Z}^k$, $(m_1, \dots, m_k) \mapsto (m_1+m_2, \dots, m_{k-1}+m_k, m_k)$ is a group isomorphism, so we can also write $f = \sum_{m\in\mathbb{Z}^k} b_{\varphi(m)}\chi_{\varphi(m)}$. Since $U_{\tau_k} f = f$, a comparison of the Fourier coefficients yields $a^{m_1}b_{m_1,\dots,m_k} = b_{\varphi(m_1,\dots,m_k)}$ for all $m_1, \dots, m_k \in \mathbb{Z}$. In particular, $|b_m| = |b_{\varphi(m)}|$ for every $m \in \mathbb{Z}^k$.

Now fix $m = (m_1, \dots, m_k) \in \mathbb{Z}^k \setminus \{0\}$ and consider two cases. Suppose first that the images $\varphi^j(m)$ for $j \in \mathbb{N}_0$ are all pairwise distinct. Then, since the Fourier coefficients are square-summable (see again Theorem A.2.5), the identity $|b_m| = |b_{\varphi^j(m)}|$ for all $j \in \mathbb{N}_0$ implies $b_m = 0$. In the second case we find $i, j \in \mathbb{N}_0$ with $i < j$ such that $\varphi^i(m) = \varphi^j(m)$, hence $\varphi^l(m) = m$ for $l := j - i \geq 1$. By considering the $(k-1)$th component we obtain $m_{k-1} = m_{k-1} + lm_k$, hence $m_k = 0$. But then the same argument allows us to show $m_{k-1} = 0$, and inductively $m_2 = \cdots = m_k = 0$. This implies that $\varphi(m) = m$, and thus $a^{m_1}b_m = b_m$. Since $m \neq 0$ and a is not a root of unity, we have $a^{m_1} \neq 1$, hence $b_m = 0$. □

We also need the following observation. The proof is left as Exercise 7.4.

Lemma 7.3.4. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{T}$ and $m \in \mathbb{N}$. If $(a_{nm+l})_{n\in\mathbb{N}}$ is equidistributed for every $l \in \{0,\dots,m-1\}$, then $(a_n)_{n\in\mathbb{N}}$ is equidistributed.*

We now prove the equidistribution theorem.

Proof of Theorem 7.3.2. We show the claim by induction on the degree of p. So take $d \in \mathbb{N}_0$ assume that $(\mathrm{e}^{2\pi \mathrm{i} q(n)})_{n\in\mathbb{N}}$ is equidistributed for all polynomials $q = \sum_{k=0}^{d-1} b_k t^k \in \mathbb{R}[t]$ of degree at most $d-1$ having at least one irrational coefficient b_j for $j \in \{1,\dots,d-1\}$. Let further $p = \sum_{k=0}^{d} c_k t^k \in \mathbb{R}[t]$ be a polynomial of degree d with some irrational coefficient.

Assume first that c_d is rational. We then choose $m \in \mathbb{N}$ with $mc_d \in \mathbb{Z}$ and consider the polynomials $q_l := \sum_{k=0}^{d} c_k(tm+l)^k - c_d(tm)^d \in \mathbb{R}[t]$ for $l \in \{0,\dots,m-1\}$. Then, since the exponential function is $2\pi\mathrm{i}$-periodic, a moment's thought reveals that $\mathrm{e}^{2\pi\mathrm{i}p(nm+l)} = \mathrm{e}^{2\pi\mathrm{i}q_l(n)}$ for every $n \in \mathbb{N}$. Moreover, the polynomials q_l have degree at most $d-1$ and have at least one irrational coefficient (aside from the constant term) for every $l \in \{0,\dots,m-1\}$. We conclude that the sequence $(\mathrm{e}^{2\pi\mathrm{i}p(nm+l)})_{n\in\mathbb{N}}$ is equidistributed for each $l \in \{0,\dots,m-1\}$. By Lemma 7.3.4 this implies the claim.

Assume now that c_d is irrational. Then $a := \mathrm{e}^{2\pi\mathrm{i}c_d(d!)} \in \mathbb{T}$ is not a root of unity. We consider the corresponding system defined by the homeomorphism $\tau_d\colon \mathbb{T}^d \to \mathbb{T}^d$ from Lemma 7.3.3. By an induction argument, we obtain for all $n \in \mathbb{N}$ and $(z_1,\dots,z_d) \in \mathbb{T}^d$ that

$$\tau_d^n(z_1,\dots,z_d) = (a^n z_1, a^{\binom{n}{2}} z_1^n z_2, \dots, a^{\binom{n}{d}} z_1^{\binom{n}{d-1}} z_2^{\binom{n}{d-2}} \cdots z_{d-1}^n z_d).$$

Since the binomial coefficient polynomials $\binom{t}{0}, \binom{t}{1}, \dots \binom{t}{d-1} \in \mathbb{R}[t]$ all have different degrees, they form a basis of the vector space of real polynomials of degree at most $d-1$. The polynomial $p - c_d(d!)\binom{t}{d}$ is of degree at most $d-1$, so we find $r_1,\dots,r_d \in \mathbb{R}$ with $p = c_d(d!)\binom{t}{d} + r_1 \cdot \binom{t}{d-1} + r_2 \cdot \binom{t}{d-1} + \dots + r_d$. Setting $z_j := \mathrm{e}^{2\pi\mathrm{i}r_j}$ for $j \in \{1,\dots d\}$ we then have

$$\mathrm{e}^{2\pi\mathrm{i}p(n)} = a^{\binom{n}{d}} z_1^{\binom{n}{d-1}} \cdots z_{d-1}^n z_d \text{ for every } n \in \mathbb{N}.$$

Finally, take $f \in \mathrm{C}(\mathbb{T})$ and consider the induced continuous function $g\colon \mathbb{T}^k \to \mathbb{C}$, $(x_1,\dots,x_k) \mapsto f(x_k)$. By Proposition 7.2.2 and Lemma 7.3.3 we obtain, in particular, that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(\mathrm{e}^{2\pi\mathrm{i}p(n)}) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g(\tau_d^n(z_1,\dots,z_d)) = \int_{\mathbb{T}^d} g \,\mathrm{dm}_{\mathbb{T}^d} = \int_{\mathbb{T}} f \,\mathrm{dm}_{\mathbb{T}}.$$

□

7.4 Comments and Further Reading

There is much more to say about weakly (and strongly) mixing measure-preserving systems, see, e.g., [EW11, Sections 2.7 and 2.8] and [EFHN15, Chapter 9]. Here we have covered only the basics.

Our proof of Furstenberg's theorem on unique ergodicity of skew-rotations as well as its application to Weyl's equidistribution theorem is based on [EFHN15, Section 10.15] and [EW11, Subsection 4.4.3] (see also [Fur14, Section III.3]). Skew-rotations will also play a crucial part in a later stage of the course.

Hermann Weyl proved his equidistribution results already in 1916, see [Wey16]. More on equidistribution of sequences can, e.g., be found in Lecture 10 of the previous internet seminar on ergodic theory [EF19], as well as the comprehensive monograph [KN74].

7.5 Exercises

Exercise 7.1. (i) For a unitary operator $V \in \mathscr{U}(H)$ consider the corresponding unitary representation $U_V : \mathbb{Z} \to \mathscr{U}(H)$, $m \mapsto V^m$ (cf. Remark 5.1.9). Show that for each $k \in \mathbb{N}$ the discrete spectrum parts of U_V and U_{V^k} coincide.
Hint: First use Exercise 6.6 to obtain

$$H_{\text{ds}} = \{f \in H \mid \overline{\{V^m f \mid m \in \mathbb{Z}\}} \textit{ compact}\}$$

for the discrete spectrum part of V.

(ii) Show that for a measure-preserving system (X, T) over $\mathbb{Z}$ and $k \in \mathbb{N}$ the following assertions are equivalent.

(a) (X, T) is weakly mixing.

(b) (X, T^k) is weakly mixing.

Exercise 7.2. (i) Let $(F_i)_{i\in I}$ be a Følner net for the abelian group Γ. Show that the following assertions are equivalent for a bounded map $\Gamma \to [0, \infty)$, $\gamma \mapsto r_\gamma$.

(a) $\lim_{i\in I} \frac{1}{|F_i|} \sum_{\gamma\in F_i} r_\gamma^p = 0$ for some $p \in [1, \infty)$.

(b) $\lim_{i\in I} \frac{1}{|F_i|} \sum_{\gamma\in F_i} r_\gamma^p = 0$ for every $p \in [1, \infty)$.

(ii) Prove Proposition 7.1.3.

Exercise 7.3. Call a subset $I \subseteq \mathbb{T}$ a **compact interval** in $\mathbb{T}$ if I is compact and connected. One can check that this is precisely the case when I is the image of a compact interval $[a, b] \subseteq \mathbb{R}$ with respect to the continuous map $\mathbb{R} \to \mathbb{T}$, $t \mapsto \mathrm{e}^{2\pi\mathrm{i}t}$. Now show that the following assertions are equivalent for a sequence $(a_n)_{n\in\mathbb{N}}$ in $\mathbb{T}$.

(a) The sequence $(a_n)_{n\in\mathbb{N}}$ is equidistributed.

(b) $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N a_n^m = 0$ for every $m \in \mathbb{Z} \setminus \{0\}$.

(c) For every compact interval $I \subseteq \mathbb{T}$ we have

$$\lim_{N\to\infty} \frac{|\{n \in \{1, \dots, N\} \mid a_n \in I\}|}{N} = \mathrm{m}_\mathbb{T}(I).$$

Hint: For "(a) ⇒ (c)" construct for $\varepsilon > 0$ continuous functions $f, g \in \mathrm{C}(\mathbb{T})$ with $f \leq \mathbb{1}_I \leq g$ and $\int_\mathbb{T} g - f \,\mathrm{dm}_\mathbb{T} \leq \varepsilon$. For the converse implication "(c) ⇒ (a)" approximate $f \in \mathrm{C}(\mathbb{T})$ uniformly by linear combinations of characteristic functions of compact intervals.

Exercise 7.4. Prove Lemma 7.3.4.

Lecture 8

In the first part of this lecture, we will present an ergodic proof of Roth's theorem, which is the special case of Szemerédi's theorem for three-term arithmetic progressions, using the JdLG-decomposition. In the second part, we will introduce disintegration of measures and relative products of probability spaces that will be used in the following lectures to extend this ergodic proof of Roth's theorem to encompass Szemerédi's theorem for arithmetic progressions of arbitrary length.

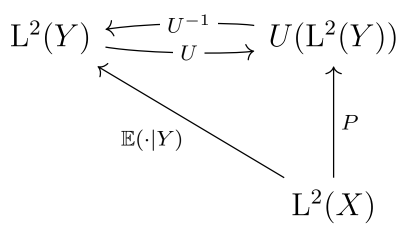
In both parts, the conditional expectation operator plays a crucial role, which we now pause to introduce.

8.1 Conditional expectation

Let X, Y be probability spaces and let $U\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ be a Markov embedding. By Definition and Proposition 1.3.2, $U(\mathrm{L}^2(Y))$ is a closed linear subspace of $\mathrm{L}^2(X)$. Let P denote the orthogonal projection of $\mathrm{L}^2(X)$ onto $U(\mathrm{L}^2(Y))$. Using the injectivity of Markov embeddings (Lemma 1.3.3 (ix)), we define $\mathbb{E}(f \mid Y)$ for $f \in \mathrm{L}^2(X)$ by

$$\mathbb{E}(f \mid Y) \in \mathrm{L}^2(Y), \quad U(\mathbb{E}(f \mid Y)) = P(f). \tag{8.1}$$

The following diagram illustrates this situation:



Lemma 8.1.1. *Let X, Y be probability spaces and let $U\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ be a Markov embedding. The* **conditional expectation operator** *$f \mapsto \mathbb{E}(f \mid Y)$ defined for $f \in \mathrm{L}^2(X)$ by (8.1) has the following properties:*

(i) *$f \mapsto \mathbb{E}(f \mid Y)$ is a linear operator from* $\mathrm{L}^2(X)$ *to* $\mathrm{L}^2(Y)$.

(ii) *If $f \geq 0$, then $\mathbb{E}(f \mid Y) \geq 0$.*

(iii) *If $f \in \mathrm{L}^2(Y)$, then $\mathbb{E}(U(f) \mid Y) = f$. In particular, $\mathbb{E}(\mathbb{1} \mid Y) = \mathbb{1}$.*

(iv) *If $f \in \mathrm{L}^2(X), g \in \mathrm{L}^\infty(Y)$, then $\mathbb{E}(fU(g) \mid Y) = g\mathbb{E}(f \mid Y)$.*

(v) *For $f \in \mathrm{L}^2(X)$, $\int_X f \,\mathrm{d}\mu_X = \int_Y \mathbb{E}(f \mid Y) \,\mathrm{d}\mu_Y$.*

(vi) *For $f \in \mathrm{L}^2(X)$, the conditional expectation $\mathbb{E}(f \mid Y)$ is the unique element of $\mathrm{L}^2(Y)$ satisfying*

$$\int_Y \mathbb{E}(f \mid Y) h \,\mathrm{d}\mu_Y = \int_X f U(h) \,\mathrm{d}\mu_X$$

for all $h \in \mathrm{L}^\infty(Y)$.

(vii) *For $f \in \mathrm{L}^2(X)$, $|\mathbb{E}(f \mid Y)|^2 \leq \mathbb{E}(|f|^2 \mid Y)$.*

(viii) *The conditional expectation operator extends to a linear operator from $\mathrm{L}^1(X)$ to $\mathrm{L}^1(Y)$ satisfying the properties (a)–(e). Moreover, it maps each $\mathrm{L}^p(X)$ to $\mathrm{L}^p(Y)$, for $1 \leq p \leq \infty$, with $\|\mathbb{E}(f \mid Y)\|_{\mathrm{L}^p(Y)} \leq \|f\|_{\mathrm{L}^p(X)}$ for every $f \in \mathrm{L}^p(X)$.*

(ix) *If $0 \leq f_1 \leq f_2 \leq \ldots$ is a monotone sequence and f is an element in $\mathrm{L}^2(X)$ such that $(f_n)_{n\in\mathbb{N}}$ converges to f almost surely, then $\mathbb{E}(f_n \mid Y)$ converges to $\mathbb{E}(f \mid Y)$ almost surely.*

(x) *If $(f_n)_{n\in\mathbb{N}}$ is a sequence and f is an element in $\mathrm{L}^2(X)$ such that f_n converges to f almost surely and there is $g \in \mathrm{L}^1(X)$ with $|f_n| \leq g$ almost surely for all n, then $\mathbb{E}(f_n \mid Y)$ converges to $\mathbb{E}(f \mid Y)$ almost surely.*

Proof. Properties (i)–(viii) follow from the properties of Markov embeddings and orthogonal projections, while property (ix) follows from property (ii), the characterization in (vi) and the monotone convergence theorem, and property (x) can be proved from (ix), we leave the details to the interested reader (cf. [EFHN15, Section 13.3]). □

Recall that if $J\colon \Sigma(Y) \to \Sigma(X)$ is a measure algebra homomorphism of probability spaces X and Y, then by Proposition 1.3.4, there is a unique Markov embedding $U_J\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ such that $U_J(\mathbb{1}_A) = \mathbb{1}_{J(A)}$ for all $A \in \Sigma(Y)$. In this situation we consider the conditional expectation operator $\mathbb{E}(\,\cdot \mid Y)$ with respect to U_J.

The conditional expectation intertwines with the dynamics of extensions:

Proposition 8.1.2. *Let $J\colon (Y,S) \to (X,T)$ be an extension of measure-preserving systems. Then for all $f \in \mathrm{L}^2(X)$ and $\gamma \in \Gamma$,*

$$\mathbb{E}(U_{T_\gamma}(f) \mid Y) = U_{S_\gamma}(\mathbb{E}(f \mid Y)).$$

Proof. Let $f \in \mathrm{L}^2(X)$ and $\gamma \in \Gamma$. By Lemma 8.1.1(vi), the definition of extensions in Definition 2.2.1, and the unitarity of U_{T_γ}, for every $h \in \mathrm{L}^\infty(Y)$,

$$\begin{aligned}\int_Y \mathbb{E}(U_{T_\gamma}(f) \mid Y) h \,\mathrm{d}\mu_Y &= \int_X U_{T_\gamma}(f) U_J(h) \,\mathrm{d}\mu_X \\ &= \int_X f U_J(U_{S_{-\gamma}}(h)) \,\mathrm{d}\mu_X \\ &= \int_Y \mathbb{E}(f \mid Y) U_{S_{-\gamma}}(h) \,\mathrm{d}\mu_Y \\ &= \int_Y U_{S_\gamma}(\mathbb{E}(f \mid Y)) h \,\mathrm{d}\mu_Y.\end{aligned}$$

Since $\mathrm{L}^\infty(Y)$ is dense in $\mathrm{L}^2(Y)$, we conclude the claim

$$\mathbb{E}(U_{T_\gamma}(f) \mid Y) = U_{S_\gamma}(\mathbb{E}(f \mid Y)).$$

□

8.2 Roth's Theorem

We are now ready to prove a special case of Szemerédi's theorem.

Theorem 8.2.1 (Roth)**.** *Let $A \subseteq \mathbb{N}$ with $\overline{\mathrm{d}}(A) > 0$. Then A contains arithmetic progressions of length 3.*

In view of Furstenberg's correspondence principle (see Theorem 4.2.8) this is a consequence of the following result.

Theorem 8.2.2. *Let (X, T) be an ergodic measure-preserving system over $\Gamma = \mathbb{Z}$. For every $f \in \mathrm{L}^\infty(\mathrm{X})$ with $f \geq 0, \int_X f \mathrm{d}\mu_X > 0$ the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f \cdot U_T^n f \cdot U_T^{2n} f$$

exists and is strictly positive.

The proof uses the following estimate for bounded sequences in a Hilbert space. An elementary, but rather technical proof will be discussed as a supplement at the end of this lecture.

Lemma 8.2.3 (van der Corput)**.** *For every bounded sequence $(a_n)_{n\in\mathbb{N}_0}$ in a Hilbert space H the inequality*

$$\limsup_{N\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} a_n \right\|^2 \leq \limsup_{M\to\infty} \frac{1}{M} \sum_{m=0}^{M-1} \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathrm{Re}\,(a_n | a_{n+m})$$

holds.

This leads to the following observation about the weakly mixing part of a measure-preserving system.

Lemma 8.2.4. *Let (X,T) be an ergodic measure-preserving system over $\Gamma = \mathbb{Z}$. Assume that $f, g \in \mathrm{L}^\infty(X)$ with $f \in \mathrm{L}^2(X)_{\mathrm{wm}}$ or $g \in \mathrm{L}^2(X)_{\mathrm{wm}}$, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f \cdot U_T^{2n} g = 0$$

in $\mathrm{L}^2(X)$.

Proof. To apply the van der Corput lemma consider $a_n := (U_T^n f)\cdot(U_T^{2n} g)$ for $n \in \mathbb{N}_0$. For $n, m \in \mathbb{N}_0$ we then have

$$\begin{aligned}(a_n|a_{n+m}) &= \int_X (U_T^n f)\cdot(U_T^{2n} g)\cdot(U_T^{n+m}\overline{f})\cdot(U_T^{2n+2m}\overline{g})\\ &= \int_X U_T^n(f\cdot(U_T^n g)\cdot(U_T^m\overline{f})\cdot(U_T^{n+2m}\overline{g})) = \int_X f\cdot(U_T^m\overline{f})\cdot U_T^n(g\cdot U_T^{2m}\overline{g}).\end{aligned}$$

Since the system is ergodic, we can use Exercise 3.4 to obtain

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (a_n|a_{n+m}) = \left(\int_X f\cdot U_T^m\overline{f}\right)\cdot\left(\int_X g\cdot U_T^{2m}\overline{g}\right).$$

for all $m \in \mathbb{N}_0$. By the Cauchy–Schwarz inequality applied to the second integral, we obtain

$$\frac{1}{M}\sum_{m=0}^{M-1}\left|\left(\int_X f\cdot U_T^m\overline{f}\right)\cdot\left(\int_X g\cdot U_T^{2m}\overline{g}\right)\right| \le \|g\|^2\cdot\frac{1}{M}\sum_{m=0}^{M-1}|(U_T^m f|f)|$$

for every $M \in \mathbb{N}$. Since $f \in \mathrm{L}^2(X)_{\mathrm{wm}}$, by Proposition 5.3.3 and Exercise 7.2 (i)

$$\lim_{M\to\infty}\frac{1}{M}\sum_{m=0}^{M-1}\left|\left(\int_X f\cdot U_T^m\overline{f}\right)\cdot\left(\int_X g\cdot U_T^{2m}\overline{g}\right)\right| = 0.$$

Similarly, if $g \in \mathrm{L}^2(X)_{\mathrm{wm}}$, we have

$$\frac{1}{M}\sum_{m=0}^{M-1}\left|\left(\int_X f\cdot U_T^m\overline{f}\right)\cdot\left(\int_X g\cdot U_T^{2m}\overline{g}\right)\right| \le 2\|f\|^2\cdot\frac{1}{2M}\sum_{m=0}^{2M-1}|(U_T^m g|g)|$$

for every $M \in \mathbb{N}$ and we again obtain[1]

$$\lim_{M\to\infty} \frac{1}{M} \sum_{m=0}^{M-1} \left| \left(\int_X f \cdot U_T^m \overline{f} \right) \cdot \left(\int_X g \cdot U_T^{2m} \overline{g} \right) \right| = 0.$$

In both cases Lemma 8.2.3 yields the claim. □

Proof of Theorem 8.2.2. Take $f \in \mathrm{L}^\infty(X)$ with $f \geq 0, \int_X f \mathrm{d}\mu_X > 0$. Suppose first that (X,T) is a measure-preserving system over $\Gamma = \mathbb{Z}$ with discrete spectrum. By the Halmos–von Neumann representation result (Theorem 6.2.6) we may assume that (X,T) is induced by a rotation map $l_a \colon G \to G$, $x \mapsto ax$ on a compact abelian group G for some $a \in G$. As an easy consequence of Proposition 6.1.8 the function

$$g\colon G \to \mathbb{C}, \quad x \mapsto \int_G f \cdot (f \circ l_x) \cdot (f \circ l_{x^2})\, \mathrm{dm}_G$$

is continuous. Notice further that $g(a^n) = \int_G f \cdot U_T^n f \cdot U_T^{2n} f \,\mathrm{d}m_G$ for every $n \in \mathbb{Z}$. In particular, $g(1) = \int_G f^3 \,\mathrm{d}m_G > 0$ since $f \geq 0, \int_X f \mathrm{d}\mu_X > 0$, and this implies $\int_G g \,\mathrm{d}m_G > 0$, see Remark 6.1.4. Using unique ergodicity of the rotation system (see Proposition 6.2.4) we obtain by Proposition 7.2.2 that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_G f \cdot U_T^n f \cdot U_T^{2n} f \,\mathrm{d}m_G = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_{T_a}^n g(1) = \int_G g \,\mathrm{m}_G > 0.$$

For a general ergodic system (X,T) consider the Kronecker subsystem $(X_{\mathrm{kro}}, T_{\mathrm{kro}})$ from Definition 7.1.14. Since $U_{J_{\mathrm{kro}}}(\mathrm{fix}(U_{T_{\mathrm{kro}}})) \subseteq \mathrm{fix}(U_T)$ we obtain that $(X_{\mathrm{kro}}, T_{\mathrm{kro}})$ is also ergodic. We use the conditional expectation $g := \mathbb{E}(f \mid X_{\mathrm{kro}}) \in \mathrm{L}^\infty(X_{\mathrm{kro}})$ of f with respect to this subsystem. By Lemma 8.1.1 (ii) and (v) we have $g > 0$, and $h := f - U_{J_{\mathrm{kro}}} g \in \mathrm{L}^2(X)_{\mathrm{wm}} \cap \mathrm{L}^\infty(X)$ by Lemma 8.1.1 (iii) and (iv). We obtain

$$\begin{aligned} &\frac{1}{N} \sum_{n=0}^{N-1} U_T^n f \cdot U_T^{2n} f - \frac{1}{N} \sum_{n=0}^{N-1} U_T^n U_{J_{\mathrm{kro}}} g \cdot U_T^{2n} U_{J_{\mathrm{kro}}} g \\ &= \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h \cdot U_T^{2n} h - \frac{1}{N} \sum_{n=0}^{N-1} U_T^n U_{J_{\mathrm{kro}}} g \cdot U_T^{2n} h - \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h \cdot U_T^{2n} U_{J_{\mathrm{kro}}} g \end{aligned}$$

and the right hand side of this equation converges to zero in $\mathrm{L}^2(X)$ by Lemma 8.2.4. This implies that the sequence $(\frac{1}{N} \sum_{n=0}^{N-1} \int_X f \cdot U_T^n f \cdot U_T^{2n} f)_{N \in \mathbb{N}}$ converges if and only if the sequence $(\frac{1}{N} \sum_{n=0}^{N-1} \int_X f \cdot U_T^n U_{J_{\mathrm{kro}}} g \cdot U_T^{2n} U_{J_{\mathrm{kro}}} g)_{N \in \mathbb{N}}$ converges, and in that case

[1] One can also use Exercise 7.1 to treat the second case exactly like the first one.

both limits agree. By Lemma 8.1.1 (iv) and (v) we have

$$\begin{aligned}\int_X f \cdot U_T^n U_{J_{\mathrm{kro}}} g \cdot U_T^{2n} U_{J_{\mathrm{kro}}} g &= \int_{X_{\mathrm{kro}}} \mathbb{E}(f \cdot U_T^n U_{J_{\mathrm{kro}}} g \cdot U_T^{2n} U_{J_{\mathrm{kro}}} g \mid X_{\mathrm{kro}}) \\ &= \int_{X_{\mathrm{kro}}} \mathbb{E}(f \cdot U_{J_{\mathrm{kro}}}(U_{T_{\mathrm{kro}}}^n g \cdot U_{T_{\mathrm{kro}}}^{2n} g) \mid X_{\mathrm{kro}}) \\ &= \int_{X_{\mathrm{kro}}} g \cdot U_{T_{\mathrm{kro}}}^n g \cdot U_{T_{\mathrm{kro}}}^{2n} g\end{aligned}$$

for every $n \in \mathbb{N}$. Thus,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f \cdot U_T^n f \cdot U_T^{2n} f = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_{X_{\mathrm{kro}}} g \cdot U_{T_{\mathrm{kro}}}^n g \cdot U_{T_{\mathrm{kro}}}^{2n} g > 0$$

by the first part. □

8.3 The relative perspective

In the previous section, we proved Furstenberg's multiple recurrence theorem in the double recurrence situation $f \cdot U_T^n f \cdot U_T^{2n} f$ (see Theorem 8.2.2). The case of single recurrence corresponds to von Neumann's mean ergodic theorem (see Theorem 3.1.1). In both situations - single and double recurrence - we observed that the proofs relied on a decomposition result. In the single recurrence case, we decomposed $\mathrm{L}^2(X)$ into the fixed space $\mathrm{fix}(U_T)$ and its orthogonal complement (see Theorem 3.2.7). In the double recurrence situation, we decomposed $\mathrm{L}^2(X)$ into the structured part $\mathrm{L}^2(X)_{\mathrm{ds}}$ and its orthogonal complement, which we identified with the weakly mixing part $\mathrm{L}^2(X)_{\mathrm{wm}}$ (the JdLG-decomposition; see Theorem 5.3.4).

To prove Furstenberg's multiple recurrence theorem in full generality (Theorem 4.2.10), we will establish a *relative* version of the JdLG-decomposition. We motivate this relative version with the following example of a skew-product extension (cf. Example 7.2.6).

We consider an ergodic skew-rotation on the 2-torus $\mathbb{T}^2$ equipped with the Haar measure, given by the homeomorphism

$$\mathbb{T}^2 \to \mathbb{T}^2, \quad (x, y) \mapsto (ax, xy),$$

where a is irrational. The JdLG-decomposition of this measure-preserving system identifies the rotational system $(\mathbb{T}, \tau_a)$ as its discrete spectrum part via the projection onto the first coordinate (this will be one part of an exercise in the next lecture). The remaining "structure" of this system can only be understood relative to its discrete spectrum part $(\mathbb{T}, \tau_a)$ (this will be the other part of an exercise in the next lecture). Indeed, relatively, the remaining part resembles a bundle of rotational systems $x \to (y \mapsto xy)$ where the rotation depends on the point x in the base system. Thus, relative to this base, we can hope to decompose the unitary representation of the system into an invariant bundle of 1-dimensional subspaces. The theory we will develop in the following lectures, known as *Furstenberg–Zimmer structure theory*, makes the previous heuristic rigorous. This theory will be crucial for establishing Furstenberg's recurrence theorem and, via Furstenberg's correspondence principle, Szemerédi's theorem in full generality.

To study the relative structure in a system, such as a relative version of the JdLG-decomposition theorem, it is necessary to develop tools for relative analysis. In our context, "relative" always means relativizing with respect to a subsystem (Y, S) of a system (X, T). Concretely, this relativization is achieved by conditioning on the subsystem (Y, S) using the conditional expectation operator. Therefore, we will start by developing the tools necessary for relative analysis before delving into the Furstenberg–Zimmer structure theory in the subsequent lectures.

8.4 Disintegration of measures and relative products

The disintegration of measures is the idea that, given a measure-preserving transformation $X \to Y$ between probability spaces X and Y, the measure μ_X can be expressed as a bundle of measures μ_y parameterized over the base space Y. Disintegration of measures will be a key tool for introducing relative products of measure-preserving systems, which are, in turn, essential for establishing a relative version of the JdLG-decomposition.

An example of disintegration of measures occurs in the construction of the 2-dimensional Lebesgue measure $\lambda_{\mathbb{R}^2}$: For a Borel set $A \subseteq \mathbb{R}^2$,

$$\lambda_{\mathbb{R}^2}(A) = \int_{\mathbb{R}} \lambda_{\mathbb{R}}(A_y)\, d\lambda_{\mathbb{R}}(y),$$

where

$$A_y = \{x \in \mathbb{R} : (x, y) \in A\}$$

is the vertical slice of A at y and $\lambda_{\mathbb{R}}$ is the 1-dimensional Lebesgue measure on the real line.

The classical theory of disintegration, which we will introduce shortly, applies to Lebesgue probability spaces. For alternative approaches that work for general probability spaces, see the Comments and Further Reading section of this lecture.

To prove the disintegration of measure theorem for Lebesgue spaces, we will first establish that such spaces are isomorphic to probability spaces formed on compact metric spaces. We begin with the following lemma.

Lemma 8.4.1. *The measure algebra $\Sigma(X)$ of a Lebesgue space X is countably generated; that is, there exists a countable algebra $\mathcal{D}$ of $\Sigma(X)$ such that any element of $\Sigma(X)$ can be written as a countable combination of unions and intersections of elements in $\mathcal{D}$.*

Proof. Recall that a probability space (X, Σ_X, μ_X) is a Lebesgue space if there exists a complete separable metric space Y equipped with the Borel σ-algebra $\Sigma_Y = \mathcal{B}(Y)$ and a Borel probability measure μ_Y, such that there is an invertible map $\varphi \in \mathrm{M}(X, Y)$ in the sense of Definition 1.1.5. In particular, the pullback map $\varphi^*\colon \Sigma(Y) \to \Sigma(X)$ is an isomorphism of measure algebras. Therefore, it suffices to show that $\Sigma(Y)$ is countably generated.

Since the space Y is separable, its topology has a countable basis $\{U_n\}_{n=1}^{\infty}$. Consider the algebra $\mathcal{D}$ generated by finite Boolean combinations of the U_n; that is, $\mathcal{D}$ consists of all sets that can be formed using a finite number of unions, intersections, and complements of the U_n. Since the U_n are countable, the collection $\mathcal{D}$ is also countable. The Borel σ-algebra $\mathcal{B}(Y)$ is generated by the open sets, so we have $\mathcal{B}(Y) = \sigma(\mathcal{D})$.

We equip $\Sigma(Y)$ with the metric $d(A, B) = \mu_Y(A \triangle B)$ (see Exercise 1.4). By Lemma 2.1.12, for any measurable set $A \in \mathcal{B}(Y)$ and any $\varepsilon > 0$, there exists a set $D \in \mathcal{D}$ such that $\mu_Y(A \triangle D) < \varepsilon$. Thus, the equivalence classes of elements of $\mathcal{D}$ are dense in $\Sigma(Y)$ with respect to the metric d.

Given any $A \in \Sigma(Y)$, there are $D_n \in \mathcal{D}$ such that $d(D_n, A) \leq 2^{-n}$. Let

$$B = \bigcap_{N \in \mathbb{N}} \bigcup_{n \geq N} D_n.$$

We get

$$\mu_Y(B \cap A^c) = \mu_Y\left(\bigcap_{N \in \mathbb{N}} \bigcup_{n \geq N} (D_n \cap A^c)\right) \leq \lim_{N \to \infty} \sum_{n \geq N} \mu_Y(D_n \cap A^c) \leq \lim_{N \to \infty} 2^{1-N} = 0,$$

and

$$\mu_Y(B^c \cap A) = \mu_Y\left(\bigcup_{N \in \mathbb{N}} \bigcap_{n \geq N} (D_n^c \cap A)\right) \leq \liminf_{n \to \infty} \mu_Y(D_n^c \cap A) = 0.$$

Then $d(A, B) = 0$, which completes the proof. □

Proposition 8.4.2. *Let X be a Lebesgue space. Then there exists a compact metric space K equipped with the Borel σ-algebra $\mathcal{B}(K)$ and a regular Borel probability measure μ_K such that (X, Σ_X, μ_X) and $(K, \mathcal{B}(K), \mu_K)$ are isomorphic in the sense of Definition 1.1.5.*

Proof. By Lemma 8.4.1, $\Sigma(X)$ is countably generated, and let $\mathcal{D}$ be a countable dense algebra. Let $\mathcal{D}_0 = \{A_1, A_2, \ldots\}$ be a subset of Σ_X consisting of a representative for each element in $\mathcal{D}$. We consider the compact metric space $K = \{0,1\}^{\mathbb{N}}$, and for every $m \in \mathbb{N}$, consider the (open and closed) cylinder set

$$A'_m = \{(x_n)_{n\in\mathbb{N}} \in K : x_m = 1\}.$$

For every $N \in \mathbb{N}$, let $\mathcal{A}_N$ be the algebra of sets on $\{0,1\}^N$ generated by the A'_m with $m \leq N$. This corresponds to the power set and the product σ-algebra on $\{0,1\}^N$.

Define a measure μ_N on $\{0,1\}^N$ by

$$\mu_N(A'_{i_1} \cap \ldots \cap A'_{i_k}) \coloneqq \mu_X(A_{i_1} \cap \ldots \cap A_{i_k}), \tag{8.2}$$

for any choice of $i_1 < \ldots < i_k \leq N$.

Extend μ_N arbitrarily to $\{0,1\}^{\mathbb{N}}$ equipped with the Borel σ-algebra $\mathcal{B}(K)$ after embedding $\{0,1\}^N$ into $\{0,1\}^{\mathbb{N}}$ by sending $(x_n)_{1\leq n\leq N}$ to $(y_n)_{n\in\mathbb{N}}$ where $x_n = y_n$ for $1 \leq n \leq N$ and $y_n = 0$ for $n > N$. By a theorem of Ulam (see, e.g., [Dud02, Theorem 7.1.3]), every Borel probability measure on a compact metric space is inner regular, and therefore also outer regular. Let μ_K be a weak* accumulation point of the sequence $(\mu_N)_{N\in\mathbb{N}}$ of regular Borel probability measures. The measure μ_K satisfies (8.2) for any choice of $i_1 < \ldots < i_k$.

We can extend the map $T(A_n) = A'_n$ to a measure algebra isomorphism from $\Sigma(X)$ to $\Sigma(K)$ using Lemma 2.1.12, as the algebras generated by the A_n and A'_n in $\Sigma(X)$ and $\Sigma(K)$, respectively, are dense with respect to the metrics associated with μ_X and μ_K, and T is an isometry. □

Standing assumptions: For the remainder of this lecture series, unless mentioned otherwise, we assume that all probability spaces are formed on compact metric spaces equipped with regular probability measures, and we consider invertible measure-preserving transformations between them in the sense of Definition 1.1.5. We further assume that Γ is a countable discrete abelian group (cf. Remark 2.1.4). Then, without loss of generality, we can assume that all measure-preserving systems are concrete systems in the sense of Definition 2.1.3. Following Exercise 2.6 and Proposition 8.4.2, we can always reduce to such a setup if we work with measure-preserving systems in the sense of Definition 2.1.1 formed on Lebesgue probability spaces. We point out that these assumptions are not necessary for developing ergodic structure theory (see the Comments and Further Reading section of this lecture for a discussion). However, we state these assumptions to reduce the amount of background and technology needed and to keep the lectures as self-contained as possible. These assumptions are satisfied in all of the applications.

Recall that $\mathrm{P}(K)$ denotes the space of regular Borel probability measures on a compact space K. By Proposition 4.1.2, $\mathrm{P}(K)$ is a compact space when identified with a subset of the dual space $\mathrm{C}(K)'$, equipped with the weak* topology. Moreover, if K is metrizable, the Banach space $\mathrm{C}(K)$ is separable (see, e.g., [AB06, Lemma 3.99]), and therefore the topology on $\mathrm{P}(K)$ is metrizable (see, e.g., [AB06, Theorem 6.30]).

With these preliminaries in place, we can state the following useful result on the disintegration of measures, where we consider the Borel σ-algebra on $\mathrm{P}(X)$.

Theorem 8.4.3. *Let $\tau\colon X \to Y$ be a measure-preserving transformation between probability spaces X, Y. Then there exists a measurable map $y \mapsto \mu_y$ from Y to $\mathrm{P}(X)$, called the* **disintegration of μ_X over** *Y, with the following properties:*

(i) *For every $f \in \mathrm{L}^2(X, \Sigma_X, \mu_X)$, we have $f \in \mathrm{L}^2(X, \Sigma_X, \mu_y)$, and*

$$\mathbb{E}(f \mid Y)(y) = \int_X f \,\mathrm{d}\mu_y \quad \text{for almost every } y \in Y.$$

(ii) *For every $f \in \mathrm{L}^2(X, \Sigma_X, \mu_X)$,*

$$\int_Y \left(\int_X f \,\mathrm{d}\mu_y \right) \mathrm{d}\mu_Y = \int_X f \,\mathrm{d}\mu_X.$$

(iii) *If there is another measurable map $y \mapsto \mu'_y$ satisfying properties (i) and (ii), then $\mu_y = \mu'_y$ for almost every $y \in Y$.*

Proof. Write $Q := \mathbb{Q} + \mathrm{i}\mathbb{Q} \subseteq \mathbb{C}$. Using separability, we choose a countable dense subset $E \subseteq \mathrm{C}(X)$ and denote by E_Q the Q-linear hull of E. For every $f \in E_Q$, pick a representative of $\mathbb{E}(f \mid Y)$ in $\mathrm{L}^2(Y)$. For almost every y, by the countability of E_Q, the map

$$\varphi_y(f) := \mathbb{E}(f \mid Y)(y)$$

is a well-defined positive, unital, continuous, and Q-linear functional on E_Q. Using continuity with respect to the sup-norm, we can extend φ_y to a positive, unital, continuous, and $\mathbb{C}$-linear functional on $\mathrm{C}(X)$ for almost every y.

By Theorem 3.2.6, we can represent φ_y by a measure μ_y for almost every y. Note that a subbasis of open sets for the weak* topology on $\mathrm{P}(X)$ is formed by sets of the form

$$\{\varphi \in \mathrm{P}(X) \colon \varphi(f) > 0\},$$

as f ranges over $\mathrm{C}(X)$. Since $\{y \in Y \colon \varphi_y(f) > 0\}$ is measurable in Y for $f \in E_Q$, it follows that $\{y \in Y \colon \varphi_y(f) > 0\}$ is measurable in Y for every $f \in C(X)$. Thus, by Proposition 1.1.4(i), the map $y \mapsto \mu_y$ is measurable.

The extension to L^2 in (i) follows from Lemma 4.1.7 and Lemma 8.1.1(ix) via a standard approximation procedure. Property (ii) is immediate from Lemma 8.1.1(v). The uniqueness property (iii) follows from uniqueness in Theorem 3.2.6. □

As a corollary, we obtain the following variant of Proposition 8.1.2.

Corollary 8.4.4. *Let $q\colon (X,\tau) \to (Y,\sigma)$ be a factor map of concrete measure-preserving systems $(X,\tau),(Y,\sigma)$, and let $y \mapsto \mu_y$ be the disintegration of μ_X over Y. Then for all $\gamma \in \Gamma$ and almost every $y \in Y$,*

$$\mu_{\sigma_\gamma y} = (\tau_\gamma)_*\mu_y.$$

Proof. The claim follows from Proposition 8.1.2. □

We can now introduce an important construction:

Definition 8.4.5. Let $\tau_i\colon X_i \to Y$, for $i = 1,2$, be measure-preserving transformations of probability spaces. Denote by $y \mapsto \mu_{1,y}$, $y \mapsto \mu_{2,y}$ the disintegration of the measures μ_{X_1}, μ_{X_2} over Y. Using the monotone convergence theorem, we can define a measure $\mu_{X_1} \otimes_Y \mu_{X_2}$ on the product measurable space $(X_1 \times X_2, \Sigma_{X_1} \otimes \Sigma_{X_2})$ by

$$\mu_{X_1} \otimes_Y \mu_{X_2}(A) = \int_Y \mu_{1,y} \otimes \mu_{2,y}(A)\,\mathrm{d}\mu_Y, \tag{8.3}$$

for $A \in \Sigma_{X_1} \otimes \Sigma_{X_2}$. We call the resulting probability space $(X_1 \times X_2, \Sigma_{X_1} \otimes \Sigma_{X_2}, \mu_{X_1} \otimes_Y \mu_{X_2})$ the **product of X_1 and X_2 relative to Y**, and denote it by $X_1 \otimes_Y X_2$.

Proposition 8.4.6. *The measure $\mu_{X_1} \otimes_Y \mu_{X_2}$ is uniquely characterized by the equality*

$$\int_{X_1\times X_2} f_1 \odot f_2\,\mathrm{d}(\mu_{X_1} \otimes_Y \mu_{X_2}) = \int_Y \mathbb{E}(f_1 \mid Y)\mathbb{E}(f_2 \mid Y)\,\mathrm{d}\mu_Y$$

for all $f_1 \in \mathrm{L}^\infty(X_1)$ and $f_2 \in \mathrm{L}^\infty(X_2)$.

Proof. This is immediate for f_1 and f_2 being characteristic functions of measurable sets. The general case follows from a standard approximation procedure using the properties of the conditional expectation in Lemma 8.1.1. □

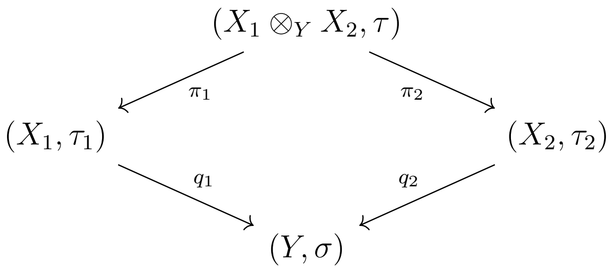
We obtain the following construction of relative products of measure-preserving systems:

Corollary 8.4.7. *Let $q_i\colon (X_i,\tau_i) \to (Y,\sigma)$, for $i = 1,2$, be factor maps of concrete measure-preserving systems. Define for every $\gamma \in \Gamma$, the map*

$$\tau_\gamma = \tau_{1,\gamma} \times \tau_{2,\gamma}. \tag{8.4}$$

Then $(X_1 \otimes_Y X_2, \tau)$ is a concrete measure-preserving system.

Moreover, the coordinate projections $\pi_i\colon X_1 \times X_2 \to X_i$ *for* $i = 1, 2$ *define factor maps of concrete measure-preserving systems, making the following diagram commute:*



Proof. Let $A_i \in \Sigma_{X_i}$ and denote by $f_i = \mathbb{1}_{A_i}$ for $i = 1, 2$. By Propositions 8.4.6 and 8.1.2,

$$\begin{aligned}\int f_1 \odot f_2 \circ (\tau_{1,\gamma} \times \tau_{2,\gamma}) \,\mathrm{d}(\mu_{X_1} \otimes_Y \mu_{X_2}) &= \int (f_1 \circ \tau_{1,\gamma}) \odot (f_2 \circ \tau_{2,\gamma}) \,\mathrm{d}(\mu_{X_1} \otimes_Y \mu_{X_2}) \\ &= \int_Y \mathbb{E}(f_1 \circ \tau_{1,\gamma} \mid Y) \odot \mathbb{E}(f_2 \circ \tau_{2,\gamma} \mid Y) \,\mathrm{d}\mu_Y \\ &= \int_Y \mathbb{E}(f_1 \mid Y) \circ \sigma_\gamma \odot \mathbb{E}(f_2 \mid Y) \circ \sigma_\gamma \,\mathrm{d}\mu_Y \\ &= \int_Y \mathbb{E}(f_1 \mid Y) \odot \mathbb{E}(f_2 \mid Y) \,\mathrm{d}\mu_Y \\ &= \int f_1 \odot f_2 \,\mathrm{d}(\mu_{X_1} \otimes_Y \mu_{X_2}).\end{aligned}$$

By linearity, this computation extends to the algebra of finite disjoint unions of product sets $A_1 \times A_2$ with $A_i \in \Sigma_{X_i}$. Finally, using Lemma 2.1.12, we obtain the first claim. The remaining claim follows by a similar line of reasoning, which we leave to the interested reader. ☐

8.5 Supplement: The van der Corput Inequality

We now give an elementary proof of Lemma 8.2.3 based Joel Moreira's blog[2]. A more structural proof based on the mean ergodic theorem (even providing a slightly sharper estimate) can be found in [EKN21][3] We start with the following observation.

Lemma 8.5.1. *Let $(b_m)_{m\in\mathbb{N}_0}$ be a bounded sequence in $\mathbb{R}$. Then*

$$\limsup_{M\to\infty} \frac{2}{M^2}\sum_{m_1=0}^{M-1}\sum_{m=0}^{M-m_1-1} b_m \le \limsup_{M\to\infty}\frac{1}{M}\sum_{m=0}^{M-1} b_m.$$

Proof. Denote the right side by c. For $\varepsilon > 0$ choose $M_0 \in \mathbb{N}$ such that $\frac{1}{M}\sum_{m=0}^{M-1} b_m \le c+\varepsilon$ for every $M \ge M_0$. Since $\limsup_{M\to\infty}\frac{2}{M^2}\sum_{m_1=M-M_0}^{M-1}\sum_{m=0}^{M-m_1-1} b_m = 0$, it suffices to check that

$$\limsup_{M\to\infty} \frac{2}{M^2}\sum_{m_1=0}^{M-M_0-1}\sum_{m=0}^{M-m_1-1} b_m \le \limsup_{M\to\infty}\frac{1}{M}\sum_{m=0}^{M-1} b_m.$$

But for $M \in \mathbb{N}$ with $M \ge M_0$ we have $\frac{1}{M}\sum_{m=0}^{M-m_1-1} b_m \le \frac{M-m_1}{M}(c+\varepsilon)$ for each $m_1 \in \{0,\dots,M-M_0-1\}$, hence

$$\begin{aligned}\frac{2}{M^2}\sum_{m_1=0}^{M-M_0-1}\sum_{m=0}^{M-m_1-1} b_m &\le (c+\varepsilon)\frac{2}{M^2}\sum_{m_1=0}^{M-M_0-1}(M-m_1)\\ &= (c+\varepsilon)\cdot\frac{(M-M_0-1)(M+M_0)}{M^2}.\end{aligned}$$

Taking the limit superior yields the claim. □

Let us now recall the statement of Lemma 8.2.3 and then prove it.

Lemma (van der Corput)**.** *For every bounded sequence $(a_n)_{n\in\mathbb{N}_0}$ in a Hilbert space H the inequality*

$$\limsup_{N\to\infty}\left\|\frac{1}{N}\sum_{n=0}^{N-1} a_n\right\|^2 \le \limsup_{M\to\infty}\frac{1}{M}\sum_{m=0}^{M-1}\limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\mathrm{Re}\,(a_n|a_{n+m})$$

holds.

[2] https://joelmoreira.wordpress.com/2012/04/24/proof-of-roths-theorem-using-ergodic-theory/

[3] See also https://joelmoreira.wordpress.com/2015/04/12/alternative-proofs-of-two-classical-lemmas/

Proof. Note first that by a telescopic sum argument we have $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1}(a_n - a_{n+m}) = 0$ for every $m \in M$, hence also

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} a_n - \frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{M}\sum_{m=0}^{M-1} a_{n+m} = 0 \tag{8.5}$$

for every $M \in M$. We obtain by the Cauchy–Schwarz inequality that

$$\left\|\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{M}\sum_{m=0}^{M-1} a_{n+m}\right\|^2 \le \frac{1}{N}\sum_{n=0}^{N-1}\left\|\frac{1}{M}\sum_{m=0}^{M-1} a_{n+m}\right\|^2 = \frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{M^2}\sum_{m_1,m_2=0}^{M-1} c_{n+m_1,n+m_2}$$

for all $N, M \in \mathbb{N}$, where $c_{k,l} := \operatorname{Re}(a_k|a_l)$ for $k, l \in \mathbb{N}_0$. For $n \in \mathbb{N}_0$ and $M \in \mathbb{N}$ we write the sum $\sum_{m_1,m_2=0}^{M-1} c_{n+m_1,n+m_2}$ as

$$\sum_{m=0}^{M-1} c_{n+m,n+m} + \sum_{m_1=0}^{M-1}\sum_{m_2=m_1+1}^{M-1} c_{n+m_1,n+m_2} + \sum_{m_2=0}^{M-1}\sum_{m_1=m_2+1}^{M-1} c_{n+m_1,n+m_2}.$$

Since $c_{k,l} = c_{l,k}$ and $c_{l,l} \ge 0$ for $l, k \in \mathbb{N}_0$ taking the real part on both sides yields

$$\begin{aligned}\sum_{m_1,m_2=0}^{M-1} c_{n+m_1,n+m_2} &= \sum_{m=0}^{M-1} c_{n+m,n+m} + 2\operatorname{Re}\left(\sum_{m_1=0}^{M-1}\sum_{m_2=m_1+1}^{M-1} c_{n+m_1,n+m_2}\right)\\ &\le 2\operatorname{Re}\sum_{m_1=0}^{M-1}\sum_{m_2=m_1+1}^{M-1} c_{n+m_1,n+m_2}.\end{aligned}$$

We conclude that

$$\begin{aligned}\left\|\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{M}\sum_{m=0}^{M-1} a_{n+m}\right\|^2 &\le \frac{2}{M^2}\sum_{m_1=0}^{M-1}\sum_{m_2=m_1}^{M-1}\frac{1}{N}\sum_{n=m_2}^{N-1+m_2} c_{n,n+m_2-m_1}\\ &= \frac{2}{M^2}\sum_{m_1=0}^{M-1}\sum_{m=0}^{M-m_1-1}\frac{1}{N}\sum_{n=m+m_1}^{N-1+m+m_1} c_{n,n+m}\end{aligned}$$

for all $N, M \in \mathbb{N}$. Again using telescopic summing we have

$$\lim_{N\to\infty}\left(\frac{1}{N}\sum_{n=m+m_1}^{N-1+m+m_1} c_{n,n+m_2-m_1} - \frac{1}{N}\sum_{n=0}^{N-1} c_{n,n+m_2-m_1}\right) = 0$$

for all $m, m_1 \in \mathbb{N}_0$. Combined with (8.5) we obtain for all $M \in \mathbb{N}$ that

$$\limsup_{N\to\infty}\left\|\frac{1}{N}\sum_{n=0}^{N-1} a_n\right\|^2 \le \frac{2}{M^2}\sum_{m_1=0}^{M-1}\sum_{m=0}^{M-m_1-1}\limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} c_{n,n+m}$$

Finally, we apply Lemma 8.5.1 with $b_m := \limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} c_{n,n+m}$ for $m \in \mathbb{N}_0$ to obtain the claim. $\square$

8.6 Comments and Further Reading

Theorem 8.2.1 was proved using Fourier-analytic methods in 1953 by Klaus Roth [Rot53], confirming a conjecture of Paul Erdős and Pál Turán from 1936 (see [ET36]). Our proof combines the reduction to the Kronecker subsystem via the van der Corput lemma (see, e.g., [EW11, Section 7.6] or [EFHN15, Section 20.4]) with an argument of Furstenberg (see [Fur77, Paragraph 3]).

We point out that today there are many different proofs of Roth's theorem, including different ergodic theoretic ones. For example, Furstenberg uses a more involved structural argument instead of the van der Corput lemma to pass to the Kronecker subsystem. On the other hand, one can use Exercise 6.6 to treat the discrete spectrum case without using the Halmos–von Neumann representation theorem (see again [EW11, Section 7.6] and [EFHN15, Section 20.4]).

The compact metric space with a regular Borel probability measure that we associated with the measure algebra of a Lebesgue probability space in Proposition 8.4.2 is an example of a *topological model* of a probability space, called the *Cantor model*. There are many topological models of probability spaces designed for different purposes (see, e.g., the Koopman model in [JST24, Appendix A.4], and for a thorough discussion [JT23a] and [EFHN15, Chapter 12]). The canonical model, as termed in [JT23a] (see therein for alternative names), represents the measure algebra of an arbitrary probability space on a, in general, inseparable compact Hausdorff space.

Using such strong topological models, it is possible to prove a version of the disintegration theorem (Theorem 8.4.3) for arbitrary probability spaces, and consequently, also to define relative products (see [JT23a, Section 8]). Surprisingly, such canonical disintegrations, as termed in [JT23a], do not yield a canonical ergodic decomposition (cf. Exercise 8.5); a counterexample was constructed in [JT23a, Appendix B].

However, one can still use the canonical disintegration, among other tools, to establish a Furstenberg–Zimmer structure theory for arbitrary (not necessarily countable) group actions on the measure algebra of arbitrary (not necessarily countably generated) probability spaces. We will comment more on this in the following lectures.

8.7 Exercises

Exercise 8.1. Use Lemma 8.2.3 to show *van der Corput's difference theorem*: If $(a_n)_{n\in\mathbb{N}}$ is a sequence in $\mathbb{T}$ such that $(a_{n+m}\overline{a_n})_{n\in\mathbb{N}}$ is equidistributed for every $m \in \mathbb{N}$, then $(a_n)_{n\in\mathbb{N}}$ is equidistributed.

Exercise 8.2. Let (X, T) be a weakly mixing measure-preserving system over $\Gamma = \mathbb{Z}$. Show that for all $f_1, \ldots, f_k \in \mathrm{L}^\infty(X)$ we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f_1 \cdots U_T^{kn} f_k = \left(\int_X f_1 \cdots \int_X f_k \right) \cdot \mathbb{1}$$

in $\mathrm{L}^2(X)$.
Hint: Use an induction on k and apply the van der Corput Lemma 8.2.3.

Exercise 8.3. Let (X, T) be a measure-preserving system with discrete spectrum over $\Gamma = \mathbb{Z}$. Show that for all $f_1, \ldots, f_k \in \mathrm{L}^\infty(X)$ the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f_1 \cdots U_T^{kn} f_k$$

exists in $\mathrm{L}^2(X)$.
Hint: First consider eigenfunctions $f_1, \ldots, f_k \in \mathrm{L}^\infty(X)$.

Exercise 8.4. The *Furstenberg–Sárközy theorem* states that for every subset $A \subseteq \mathbb{N}$ with upper density $\overline{\mathrm{d}}(A) > 0$ there are $a, d \in \mathbb{N}$ with $a, a+d^2 \in A$. In this exercise we prove a known generalization of this result. In the following $U \in \mathscr{U}(H)$ is a unitary operator on a Hilbert space H and $p \in \mathbb{Z}[t]$ a polynomial with integer coefficients.

(i) Show that

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

for every $f \in \mathrm{fix}(U^m)$ where $m \in \mathbb{N}$.

(ii) Define the **rational spectrum part** by

$$H_{\mathrm{rs}} := \overline{\mathrm{lin}} \bigcup_{k\in\mathbb{N}} \bigcup_{\substack{a\in\mathbb{T} \\ a^k=1}} \ker(a - U) \subseteq H_{\mathrm{ds}}.$$

Show that the limit $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U^{p(n)} f$ exists for every $f \in H_{\mathrm{rs}}$.
Hint: Observe that the inclusion $H_{\mathrm{rs}} \subseteq \overline{\bigcup_{m\in\mathbb{N}} \mathrm{fix}(U^m)}$ holds[4]

[4] In fact, this is even an equality, see, e.g., [EF19, Section 6.2].

(iii) Show that if $p \in \mathbb{Z}[t]$ is non-constant, then $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} U^{p(n)}f = 0$ for every $f \in H_{\mathrm{ds}} \cap H_{\mathrm{rs}}^{\perp}$.
Hint: Use Theorem 7.3.2.

(iv) Show that if $p \in \mathbb{Z}[t]$ is non-constant, then $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} U^{p(n)}f = 0$ for every $f \in H_{\mathrm{wm}}$.
Hint: Use an induction on the degree of p and Lemma 8.2.3.

(v) Show that the limit $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} U^{p(n)}f$ exists for every $f \in H$.

(vi) Show that for every measure-preserving system (X,T) over $\mathbb{Z}$ the rational spectrum part $\mathrm{L}^2(X)_{\mathrm{rs}}$ is an invariant Markov sublattice of $\mathrm{L}^2(X)$.

(vii) Show that for every polynomial $p \in \mathbb{Z}[t]$ with $p(0) = 0$, every measure-preserving system (X,T) over $\mathbb{Z}$ and every $f \in \mathrm{L}^\infty(X)$ with $f \geq 0$, $\int_X f\,\mathrm{d}\mu_X > 0$ the limit

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \int_X f \cdot U_T^{p(n)} f$$

exists and is strictly positive.
Hint: Argue as in the proof of Theorem 8.2.2 to reduce to the case $\mathrm{L}^2(X) = \mathrm{L}^2(X)_{\mathrm{rs}}$. *We may further assume that* $\|f\|_2 = 1$. *Writing* P_m *for the orthogonal projection onto* $\mathrm{fix}(U_T^m)$, *we find some* $m \in \mathbb{N}$ *with* $\|f - P_m f\|_2 \leq \frac{1}{4}(\int_X f)^2$. *Use the Cauchy–Schwarz inequality as well as the facts* $P_m f \geq 0$ *and* $\int_X f = \int_X P_m f$ *to check that* $\int_X f \cdot U_T^{p(mn)} f \geq \frac{1}{2}(\int_X f)^2$ *for all* $n \in \mathbb{N}$.

(viii) For every polynomial $p \in \mathbb{Z}[t]$ with $p(0) = 0$ and every subset $A \subseteq \mathbb{N}$ with $\overline{\mathrm{d}}(A) > 0$ there are $a, d \in \mathbb{Z}$ with $a, a + p(d) \in A$.

Exercise 8.5. Let (X,τ) be a concrete measure-preserving system. Let $Y = X$, let Σ_Y be the σ-algebra of almost invariant subsets of Σ_X, let $\mu_Y = \mu_X$, and let $\sigma\colon \Gamma \to \mathrm{Aut}(Y)$ be the trivial homomorphism. Consider the disintegration $y \mapsto \mu_y$ of μ_X over Y.

Prove that for almost every $y \in Y$, the measure-preserving system on X associated with the action τ and the measure μ_y is ergodic. While doing so, verify all the unproved claims in the statement of this exercise.

Lecture 9

In this lecture, we introduce *compact extensions* and *weakly mixing extensions*, and establish their dichotomy. From this, we infer the foundational Furstenberg–Zimmer structure theorem. We begin by developing a conditional version of Hilbert space theory to perform the *relative analysis* required for the proofs of these results.

9.1 The Furstenberg–Zimmer Structure Theorem

We emphasize that equalities and inequalities between measurable functions will always be understood in the almost sure sense, unless explicitly mentioned otherwise. Additionally, we will freely pass from the underlying probability spaces to their measure algebras and work with representatives if needed, without explicitly mentioning it. In our countable setting, these passages are all well-defined and safe; however, we encourage the interested reader to verify for themselves that the arguments are sound and the notations are well-defined.

Throughout this section, let $q\colon (X,\tau) \to (Y,\sigma)$ be a factor map of concrete measure-preserving systems, and let $U_q\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X)$ be the associated Markov embedding. Let $y \mapsto \mu_y$ denote the disintegration of μ_X relative to Y.

Recall from Section 8.4 that we denote by $(X \otimes_Y X, \tau \times \tau)$ the product system of (X,τ) with itself relative to its subsystem (Y,σ). We can now use the first coordinate projection $\pi_1\colon X \times X \to X$ (alternatively, we could use the second coordinate projection π_2) to define the factor map

$$p\colon (X \otimes_Y X, \tau \times \tau) \to (Y,\sigma)$$

(cf. the commutative diagram 8.4.7), and let $U_p\colon \mathrm{L}^2(Y) \to \mathrm{L}^2(X \otimes_Y X)$ be the associated Markov embedding.

The key idea for studying extensions $q : (X,\tau) \to (Y,\sigma)$ is to replace the classical analysis with "relative analysis" over Y. For example, instead of vector spaces, we are now interested in "vector spaces" over the algebra $\mathrm{L}^\infty(Y)$. Since $\mathrm{L}^\infty(Y)$ is generally not a field, the right concept is that of a module. Recall that an **R-module** M over a commutative unital ring R consists of an abelian group $(M,+)$ and an operation $\cdot : R \times M \to M$ such that for all $r,s \in R$ and $x,y \in M$, the following properties hold:

$$\begin{aligned} &r \cdot (x+y) = r \cdot x + r \cdot y, \\ &(r+s) \cdot x = r \cdot x + s \cdot x, \\ &(rs) \cdot x = r \cdot (s \cdot x), \\ &1 \cdot x = x. \end{aligned}$$

For $h \in \mathrm{L}^\infty(Y)$ and $f \in \mathrm{L}^2(X)$, we have $U_q(h)f \in \mathrm{L}^2(X)$. This equips $\mathrm{L}^2(X)$ with the structure of an $\mathrm{L}^\infty(Y)$-module. Similarly, for $f \in \mathrm{L}^2(X \otimes_Y X)$, we have $U_p(h)f \in \mathrm{L}^2(X \otimes_Y X)$, and this equips $\mathrm{L}^2(X \otimes_Y X)$ with the structure of an $\mathrm{L}^\infty(Y)$-module as well.

Definition 9.1.1. Let $f,g \in \mathrm{L}^2(X)$. We define the **conditional inner product** of f and g by

$$(f \mid g)_{X|Y} \coloneqq \mathbb{E}(f\bar{g} \mid Y),$$

and the **conditional norm** of f by

$$\|f\|_{X|Y} \coloneqq \sqrt{(f \mid f)_{X|Y}}.$$

Moreover, we say that f,g are **conditionally orthogonal** if $(f \mid g)_{X|Y} = 0$.

For $f,g \in \mathrm{L}^2(X)$, the conditional inner product is only an element of $\mathrm{L}^1(Y)$. To obtain an $\mathrm{L}^\infty(Y)$-valued inner product, we introduce the following definition.

Definition 9.1.2. We define the **conditional L^2-space**

$$\mathrm{L}^2(X \mid Y) \coloneqq \{f \in \mathrm{L}^2(X) \colon \| \|f\|_{X|Y} \|_{\mathrm{L}^\infty(Y)} < \infty\},$$

and equip it with the norm

$$\|f\|_{\mathrm{L}^2(X|Y)} \coloneqq \| \|f\|_{X|Y} \|_{\mathrm{L}^\infty(Y)}.$$

Note that we have the inclusions

$$\mathrm{L}^\infty(X) \subseteq \mathrm{L}^2(X \mid Y) \subseteq \mathrm{L}^2(X).$$

Remark 9.1.3. Similarly, we define the conditional inner product $(\cdot \mid \cdot)_{X \otimes_Y X|Y}$, the conditional norm $\|\cdot\|_{X \otimes_Y X|Y}$, and the **conditional L^2-tensor space** $\mathrm{L}^2(X \otimes_Y X \mid Y)$ with norm $\|\cdot\|_{\mathrm{L}^2(X \otimes_Y X|Y)}$.

The next proposition explains why we call $\mathrm{L}^2(X \mid Y)$ a conditional Hilbert space (similar properties also hold for $\mathrm{L}^2(X \otimes_Y X \mid Y)$).

Proposition 9.1.4. *The following properties are satisfied.*

(i) *If* $f, g \in \mathrm{L}^2(X \mid Y)$, *then* $(f \mid g)_{X|Y} \in \mathrm{L}^\infty(Y)$ *and consequently also* $\|f\|_{X|Y} \in \mathrm{L}^\infty(Y)$.

(ii) $\mathrm{L}^2(X \mid Y)$ *is an* $\mathrm{L}^\infty(Y)$*-module and a Banach space with respect to the norm* $\|\cdot\|_{\mathrm{L}^2(X|Y)}$.

(iii) *If* $f, g \in \mathrm{L}^2(X \mid Y)$ *and* $h \in \mathrm{L}^\infty(Y)$, *then*

$$(U_q(h)f \mid g)_{X|Y} = h(f \mid g)_{X|Y} = (f \mid U_q(\bar{h})g)_{X|Y}.$$

(iv) *If* $f, g \in \mathrm{L}^2(X \mid Y)$, *the* **conditional Cauchy–Schwarz inequality** *holds:*

$$|(f \mid g)_{X|Y}| \leq \|f\|_{X|Y}\|g\|_{X|Y}.$$

(v) *If* $f, g \in \mathrm{L}^2(X \mid Y)$, *the* **conditional triangle inequality** *holds:*

$$\|f + g\|_{X|Y} \leq \|f\|_{X|Y} + \|g\|_{X|Y}.$$

(vi) *If* $f, g \in \mathrm{L}^2(X \mid Y)$, *the* **conditional Pythagorean identity** *holds:*

$$\|f - g\|_{X|Y}^2 = \|f\|_{X|Y}^2 - 2\mathrm{Re}(f \mid g)_{X|Y} + \|g\|_{X|Y}^2.$$

Proof. Exercise. □

In Exercise 6.6, we described systems with discrete spectrum in terms of precompact orbits. The following notion of conditional precompactness in the conditional Hilbert space $\mathrm{L}^2(X \mid Y)$, and for extensions of measure-preserving systems, is a relative version of this concept.

Definition 9.1.5. (i) A subset $Z \subseteq \mathrm{L}^2(X \mid Y)$ is said to be **conditionally precompact** if for every $\varepsilon > 0$, there exist finitely many $g_1, \ldots, g_n \in \mathrm{L}^2(X \mid Y)$ such that

$$\inf_{1 \leq i \leq n} \|f - g_i\|_{X|Y} \leq \varepsilon \mathbb{1}.$$

for each $f \in Z$.

(ii) An element $f \in \mathrm{L}^2(X \mid Y)$ is said to be **conditionally almost periodic** if its orbit

$$\{U_{T_\gamma}(f) : \gamma \in \Gamma\} \subseteq \mathrm{L}^2(X \mid Y)$$

is conditionally precompact.

(iii) The extension $q\colon (X,\tau)\to(Y,\sigma)$ is said to be **compact** if the set

$$\{f\in \mathrm{L}^2(X\mid Y)\colon f \text{ conditionally almost periodic}\}$$

is dense in $\mathrm{L}^2(X)$.

Some important examples related to the notion of conditionally almost periodicity are discussed in the exercises to this lecture.

We show that the collection of functions that are conditionally almost periodic forms a subsystem lying between (X,τ) and (Y,σ):

Proposition 9.1.6. *There exists a concrete measure-preserving system (Z,ρ) with the following properties.*

(i) (Z,ρ) is a subsystem of (X,τ) and a compact extension of (Y,σ).

(ii) If (Z',ρ') is another concrete measure-preserving system satisfying both properties in (i), then (Z',ρ') is a subsystem of (Z,ρ).

We call (Z,ρ) the **maximal compact extension of** (Y,σ) **below** (X,τ).

Proof. Exercise. □

The following definition of weakly mixing extensions is motivated by the characterization of weak mixing in Proposition 7.1.3. Observe that countable discrete abelian groups admit Følner sequences since the set of finite subsets of a countable set is countable, as used in the proof of Proposition 3.1.10 to construct Følner nets.

Definition 9.1.7. Let (F_n) be a Følner sequence for Γ. A function $f\in\mathrm{L}^2(X\mid Y)$ is said to be **conditionally weakly mixing** if for every $g\in\mathrm{L}^2(X\mid Y)$, we have

$$\lim_{n\to\infty}\frac{1}{|F_n|}\sum_{\gamma\in F_n}\|(U_{\tau_\gamma}(f)\mid g)_{X|Y}\|^2_{\mathrm{L}^2(Y)}=\lim_{n\to\infty}\frac{1}{|F_n|}\sum_{\gamma\in F_n}\|(f\mid U_{\tau_\gamma}(g))_{X|Y}\|^2_{\mathrm{L}^2(Y)}=0. \tag{9.1}$$

The extension $q\colon(X,\tau)\to(Y,\sigma)$ is called **weakly mixing** if every $f\in\mathrm{L}^2(X\mid Y)$ with $\mathbb{E}(f\mid Y)=0$ is conditionally weakly mixing.

Remark 9.1.8. From the identity

$$\begin{aligned}\|(U_{\tau_\gamma}(f)\mid g)_{X|Y}\|^2_{\mathrm{L}^2(Y)}&=\int_Y (U_{\tau_\gamma}(f)\mid g)_{X|Y}(g\mid U_{\tau_\gamma}(f))_{X|Y}\,d\mu_Y\\&=(U_{\tau_\gamma}(f\odot\bar f)\mid g\odot\bar g)_{\mathrm{L}^2(X\otimes_Y X)},\end{aligned}$$

we can rewrite the expression in (9.1) in terms of ergodic averages:

$$\frac{1}{|F_n|}\sum_{\gamma\in F_n}(U_{\tau_\gamma}(f\odot\bar f)\mid g\odot\bar g)_{\mathrm{L}^2(X\otimes_Y X)}.$$

From the abstract ergodic theorem (cf. Theorem 3.1.5), these averages converge to

$$(P_{\mathrm{fix}(U_{\tau\times\tau})}(f \odot \bar{f})|g \odot \overline{g}).$$

This readily implies that (9.1) is equivalent to $P_{\mathrm{fix}(U_{\tau\times\tau})}(f \odot \bar{f}) = 0$. In particular, Definition 9.1.7 is independent of the choice of the Følner sequence for Γ.

The aim of this section is to establish the following relative version of the dichotomy between compactness and weak mixing (cf. Theorem 7.1.16).

Theorem 9.1.9. *Let $q\colon (X, \tau) \to (Y, \sigma)$ be an extension of concrete measure-preserving systems. Then exactly one of the following statements is true:*

(i) There exists a concrete measure-preserving system (Z, ρ) such that the following diagram commutes:

$$(X,\tau) \xrightarrow{q_1} (Z,\rho) \xrightarrow{q_2} (Y,\sigma), \qquad (X,\tau) \xrightarrow{q} (Y,\sigma)$$

where q_1 and q_2 are extensions, with q_2 being a compact extension, which is not an isomorphism.

(ii) q is a weakly mixing extension.

In order to prove Theorem 9.1.9, we will establish first:

Proposition 9.1.10. *Let $q\colon (X, \tau) \to (Y, \sigma)$ be an extension of concrete measure-preserving systems. A function $f \in \mathrm{L}^2(X \mid Y)$ is conditionally weakly mixing if and only if $(f \mid g)_{X|Y} = 0$ for all conditionally almost periodic functions $g \in \mathrm{L}^2(X \mid Y)$.*

We show how Proposition 9.1.10 implies Theorem 9.1.9.

Proof of Theorem 9.1.9. First, observe that Proposition 9.1.10 implies, in particular, that the properties (i) and (ii) in Theorem 9.1.9 cannot occur simultaneously.

Now, if q is not a weakly mixing extension, there exists $f \in \mathrm{L}^2(X \mid Y)$ with $\mathbb{E}(f \mid Y) = 0$ that is not conditionally weakly mixing. By Proposition 9.1.10, there exists a conditionally almost periodic function $g \in \mathrm{L}^2(X \mid Y)$ such that $(f \mid g)_{X|Y} \neq 0$.

By assumption, f is orthogonal to $U_q(\mathrm{L}^\infty(Y))$, and thus g is not an element of $U_q(\mathrm{L}^\infty(Y))$. By Proposition 9.1.6, the maximal compact extension of (Y, σ) is not trivial. □

We split the proof of Proposition 9.1.10. We first prove the easier "only if" part:

Proof. Let $f \in \mathrm{L}^2(X \mid Y)$ be conditionally weakly mixing and $g \in \mathrm{L}^2(X \mid Y)$ be conditionally almost periodic. Let (F_n) be a Følner sequence for Γ and let $\varepsilon > 0$ be arbitrary. Since g is conditionally almost periodic, there are $f_1, \dots, f_n \in \mathrm{L}^2(X \mid Y)$ such that

$$\sup_{\gamma \in \Gamma} \inf_{i=1,\dots,n} \|U_{\tau_\gamma} g - f_i\|_{X|Y} \le \varepsilon \mathbb{1}.$$

We claim that

$$\|(f|g)_{X|Y}\|_{\mathrm{L}^2(Y)} \le \varepsilon \|f\|_{\mathrm{L}^2(X)} + \sum_{i=1}^{n} \|(U_{\tau_\gamma}(f) \mid f_i)_{X|Y}\|_{\mathrm{L}^2(Y)} \tag{9.2}$$

for all $\gamma \in \Gamma$. By definition of conditional weakly mixing, this then yields

$$\|(f|g)_{X|Y}\|_{\mathrm{L}^2(Y)} = \lim_{n \to \infty} \frac{1}{|F_n|} \sum_{\gamma \in F_n} \|(f|g)_{X|Y}\|_{\mathrm{L}^2(Y)} \le \varepsilon \|f\|_{\mathrm{L}^2(X)},$$

hence $(f|g)_{X|Y} = 0$ as desired.

So take $\gamma \in \Gamma$. We find a measurable partition $\{A_1, \dots, A_n\}$ of Y such that

$$\inf_{i=1,\dots,n} \|U_{\tau_\gamma} g - f_i\|_{X|Y}(y) = \|U_{\tau_\gamma} g - f_j\|_{X|Y}(y)$$

for almost every $y \in A_j$ where $j \in \{1, \dots, n\}$. Then $w \coloneqq \sum_{i=1}^n \mathbb{1}_{A_i} f_i \in \mathrm{L}^2(X \mid Y)$ satisfies

$$\|U_{\tau_\gamma} g - w\|_{X|Y} = \sum_{i=1}^{n} \mathbb{1}_{A_i} \|U_{\tau_\gamma} g - w\|_{X|Y} = \sum_{i=1}^{n} \mathbb{1}_{A_i} \|U_{\tau_\gamma} g - f_i\|_{X|Y} \le \varepsilon \mathbb{1}.$$

By the conditional Cauchy–Schwarz inequality in Proposition 9.1.4 we therefore obtain

$$\begin{aligned} |(U_{\tau_\gamma} f \mid U_{\tau_\gamma} g)_{X|Y}| &\le |(U_{\tau_\gamma} f \mid U_{\tau_\gamma} g - w)_{X|Y}| + |(U_{\tau_\gamma} f \mid w)_{X|Y}| \\ &\le \|U_{\tau_\gamma} f\|_{X|Y} \cdot \|U_{\tau_\gamma} g - w\|_{X|Y} + |(U_{\tau_\gamma} f \mid w)_{X|Y}| \\ &\le \varepsilon \|U_{\tau_\gamma} f\|_{X|Y} + |(U_{\tau_\gamma} f \mid w)_{X|Y}|. \end{aligned}$$

Since $\|(U_{\tau_\gamma} f \mid U_{\tau_\gamma} g)_{X|Y})\|_{\mathrm{L}^2(Y)} = \|U_{\sigma_\gamma}(f \mid g)_{X|Y}\|_{\mathrm{L}^2(Y)} = \|(f|g)_{X|Y}\|_{\mathrm{L}^2(Y)}$, this implies

$$\|(f \mid g)_{X|Y}\|_{\mathrm{L}^2(Y)} = \|(U_{\tau_\gamma} f \mid U_{\tau_\gamma} g)_{X|Y}\|_{\mathrm{L}^2(Y)} \le \varepsilon \|f\|_{\mathrm{L}^2(X)} + \|(U_{\tau_\gamma} f \mid w)_{X|Y}\|_{\mathrm{L}^2(Y)}.$$

Using Proposition 9.1.4 (iii), we finally have

$$\varepsilon \|f\|_{\mathrm{L}^2(X)} + \|(U_{\tau_\gamma} f \mid w)_{X|Y}\|_{\mathrm{L}^2(Y)} \le \varepsilon \|f\|_{\mathrm{L}^2(X)} + \sum_{i=1}^{n} \|(U_{\tau_\gamma}(f) \mid f_i)_{X|Y}\|_{\mathrm{L}^2(Y)},$$

and therefore obtain the desired estimate (9.2). □

For the "if" part of Proposition 9.1.10, we prove that if $f \in \mathrm{L}^2(X \mid Y)$ is not conditionally weakly mixing, then it must non-trivially conditionally correlate with a conditionally almost periodic function.

From the computation in Remark 9.1.8, it follows that $f \in \mathrm{L}^2(X \mid Y)$ is conditionally weakly mixing if and only if

$$P_{\mathrm{fix}(U_{\tau\times\tau})}(f \odot \bar{f}) = 0.$$

Suppose $f \in \mathrm{L}^2(X \mid Y)$ is not conditionally weakly mixing; thus,

$$K := P_{\mathrm{fix}(U_{\tau\times\tau})}(f \odot \bar{f}) \neq 0.$$

Notice that $K \in \mathrm{L}^2(X \otimes_Y X \mid Y)$.

Now, we define the $\mathrm{L}^\infty(Y)$-linear operator $K *_Y \colon \mathrm{L}^2(X \mid Y) \to \mathrm{L}^2(X \mid Y)$ by

$$(K *_Y f)(x) := \int_X K(x, x') f(x')\, d\mu_{q(x)}(x'). \tag{9.3}$$

By Exercise 9.6, for all $f \in \mathrm{L}^2(X \mid Y)$,

$$\|K *_Y f\|_{X|Y} \leq \|K\|_{X\otimes_Y X|Y} \|f\|_{X|Y}. \tag{9.4}$$

Since $P_{\mathrm{fix}(U_{\tau\times\tau})}$ is an orthogonal projection, we have

$$\begin{aligned}
\int_Y (K *_Y f \mid f)_{X|Y}\, d\mu_Y &= (K *_Y f \mid f)_{\mathrm{L}^2(X)} \\
&= (K \mid f \odot \bar{f})_{\mathrm{L}^2(X\otimes_Y X)} \\
&= (P_{\mathrm{fix}(U_{\tau\times\tau})}(f \odot \bar{f}) \mid f \odot \bar{f})_{\mathrm{L}^2(X\otimes_Y X)} \\
&= \|P_{\mathrm{fix}(U_{\tau\times\tau})}(f \odot \bar{f})\|^2_{\mathrm{L}^2(X\otimes_Y X)} \\
&= \|K\|^2_{\mathrm{L}^2(X\otimes_Y X)}.
\end{aligned}$$

Thus, if $K \neq 0$, then $(K *_Y f \mid f)_{X|Y} \neq 0$. To complete the proof of the "if" part of Proposition 9.1.10, it suffices to show that for every $\varepsilon > 0$ there is some measurable subset $E \subseteq Y$ with $\mu_Y(E) \geq 1 - \varepsilon$ such that $\mathbb{1}_E K *_Y f$ is conditionally almost periodic.

This is because, by choosing a measurable subset $E \subseteq Y$ with sufficiently large measure, we can ensure that the conditionally almost periodic function $\mathbb{1}_E K *_Y f$ satisfies

$$(\mathbb{1}_E K *_Y f \mid f)_{X|Y} = \mathbb{1}_E (K *_Y f \mid f)_{X|Y} \neq 0.$$

The following lemma is needed to complete the argument.

Lemma 9.1.11. *Let $K \in \mathrm{L}^2(X \otimes_Y X \mid Y)$, let $C > 0$ be some constant, and let*

$$M = \{f \in \mathrm{L}^2(X \mid Y) \colon \|f\|_{\mathrm{L}^2(X|Y)} \leq C\}.$$

Then for every $\varepsilon > 0$ there is a a measurable set $E \subseteq Y$ of measure $\mu_Y(E) \geq 1 - \varepsilon$ such that for every $\delta > 0$ there exist finitely many $g_1, \dots, g_n \in \mathrm{L}^2(X \mid Y)$ with

$$\sup_{f \in M} \min_{1 \leq i \leq n} \|\mathbb{1}_E((K * f) - g_i)\|_{\mathrm{L}^2(X|Y)} \leq \delta.$$

Proof. Fix $\varepsilon > 0$. Let (K_n) be a sequence of linear combinations of elementary tensors $g \odot h \in \mathrm{L}^\infty(X \otimes_Y X)$ such that $\|K_n - K\|_{\mathrm{L}^2(X \otimes_Y X)} \to 0$. Since

$$\|K_n - K\|_{\mathrm{L}^2(X \otimes_Y X)} = \|\|K_n - K\|_{X \otimes_Y X|Y}\|_{\mathrm{L}^2(Y)},$$

by passing to a subsequence, we have that $\|K_n - K\|_{X \otimes_Y X|Y} \to 0$ almost surely in $y \in Y$. By Egorov's theorem, there is a measurable set E in Y with $\mu_Y(E) \geq 1 - \varepsilon$ such that $\mathbb{1}_E \|K_n - K\|_{X \otimes_Y X|Y} \to 0$ in the norm of $\mathrm{L}^\infty(Y)$.

Now for $\delta > 0$ choose n large enough to guarantee that

$$\|K_n - K\|_{X \otimes_Y X|Y} \mathbb{1}_E \leq \frac{\delta}{2C} \mathbb{1}.$$

By (9.4), we obtain for all $f \in M$,

$$\|K_n *_Y f - K *_Y f\|_{X|Y} \mathbb{1}_E \leq \|K_n - K\|_{X \otimes_Y X|Y} \|f\|_{X|Y} \mathbb{1}_E \leq \frac{\delta}{2C} \cdot C\mathbb{1} = \frac{\delta}{2}\mathbb{1}. \quad (9.5)$$

Write $K_n = \sum_{i=1}^N g_i \odot h_i$ with $g_i, h_i \in \mathrm{L}^\infty(X)$. For all $f \in M$, we have

$$K_n *_Y f = \sum_{i=1}^N U_q(c_i) g_i,$$

where $c_i \coloneqq (f|\overline{h_i})_{X|Y} \in \mathrm{L}^\infty(Y)$ satisfies $\|c_i\|_{\mathrm{L}^\infty(Y)} \leq C \cdot \max_{j=1,\dots,N} \|\|h_j\|_{X|Y}\|_{\mathrm{L}^\infty(Y)}$ for all i.

Denote by $D \coloneqq \max_{j=1,\dots,N} \|g_j\|_{\mathrm{L}^2(X|Y)}$. Let F be the finite set of all linear combinations of $g_1, \dots, g_N$ with coefficients from a finite set of complex numbers which is $\frac{\delta}{2ND}$-dense in the disc of radius C. By the conditional triangle inequality, there exists for each $f \in M$ some $g \in F$ such that

$$\|K_n *_Y f - g\|_{\mathrm{L}^2(X|Y)} < \delta/2.$$

The claim follows from (9.5) by another application of the conditional triangle inequality. □

We continue with the proof of the "if" part of Proposition 9.1.10. Since $f \in \mathrm{L}^2(X \mid Y)$, there is $C > 0$ such that $\|f\|_{\mathrm{L}^2(X|Y)} \leq C$, and therefore also $\|U_{\tau_\gamma}(f)\|_{\mathrm{L}^2(X|Y)} \leq C$ for all $\gamma \in \Gamma$.

Let $\varepsilon > 0$. Let E be the measurable subset of Y with $\mu_Y(E) \geq 1 - \varepsilon$ from Lemma 9.1.11. Let $A = \bigcup_{\gamma\in\Gamma} \sigma_\gamma^{-1}(E)$, and set

$$\varphi := K *_Y f \mathbb{1}_{q^{-1}(A)}.$$

Since $E \subseteq A$, $\mu_Y(A) \geq 1 - \varepsilon$, so it suffices to show that φ is conditionally almost periodic.

So take $\delta > 0$ and let F be the finite set from Lemma 9.1.11. Write A as a disjoint union $\bigcup_n A_n$ of measurable sets $A_n \subseteq \sigma_{\gamma_n}^{-1}(E)$ for some $\gamma_n \in \Gamma$, and define $g' := \sum_{n\geq 1} U_{\tau_{\gamma_n}}(g)\mathbb{1}_{A_n}$ for all $g \in F$. Denote their collection by $F' = \{g' \colon g \in F\}$. By construction, $U_{\tau_\gamma}(\varphi)\mathbb{1}_{A^c} = 0$ for all $\gamma \in \Gamma$. Since $U_{\tau_\gamma}(K *_Y h) = K *_Y U_{\tau_\gamma}(h)$ by the invariance of K, we have

$$\sup_{\gamma\in\Gamma} \inf_{g'\in F'} \|U_{\tau_\gamma}(\varphi) - g'\|_{X|Y} \leq \delta\mathbb{1},$$

concluding the proof of Proposition 9.1.10.

We now state and prove the beautiful Furstenberg–Zimmer structure theorem in the setting of countable abelian group actions on standard Lebesgue spaces. This theorem describes any such system as a tower, called the *Furstenberg–Zimmer tower*, of compact extensions, followed by a weakly mixing extension of the inverse limit of the compact extensions.

Theorem 9.1.12. *Let (X, τ) be a concrete measure-preserving system. Then there exists an ordinal number α and a subsystem (Y_β, σ_β) for every $\beta \leq \alpha$ with the following properties:*

- *(i) $(Y_\emptyset, \sigma_\emptyset)$ is the trivial system on a singleton.*
- *(ii) For every successor ordinal $\beta + 1 \leq \alpha$, $(Y_{\beta+1}, \sigma_{\beta+1})$ is a compact extension of (Y_β, σ_β).*
- *(iii) For every limit ordinal $\beta \leq \alpha$, (Y_β, σ_β) is the inverse limit of the (Y_η, σ_η) in the sense that $\mathrm{L}^2(Y_\beta)$ is the closure of $\bigcup_{\eta<\beta} \mathrm{L}^2(Y_\eta)$.*
- *(iv) (X, τ) is the weakly mixing extension of $(Y_\alpha, \sigma_\alpha)$.*

Proof. By Theorem 9.1.9, the extension $(X, \tau) \to (Y_\emptyset, \sigma_\emptyset)$ is either weakly mixing, in which case we set $\alpha = 0$, or the maximal compact extension $(Y_1, \sigma_1) \to (Y_\emptyset, \sigma_\emptyset)$ below (X, τ) is not isomorphic to $(Y_\emptyset, \sigma_\emptyset)$.

If the latter holds, then by Theorem 9.1.9, the extension $(X, \tau) \to (Y_1, \sigma_1)$ is either weakly mixing, in which case we set $\alpha = 1$, or the maximal compact extension $(Y_2, \sigma_2) \to (Y_1, \sigma_1)$ below (X, τ) is not isomorphic to (Y_1, σ_1).

We may need to continue this process and repeat the previous step.

Repeating these steps in a transfinite recursion, while passing to inverse limits at limit ordinals, this process will eventually terminate at an ordinal number α.

Since $\mathrm{L}^2(X)$ is separable, α is a countable ordinal. $\square$

9.2 Comments and Further Reading

Furstenberg–Zimmer structure theory originates from the foundational works of Furstenberg [Fur77] and Zimmer [Zim76a, Zim76b]. The proof we present is inspired by and combines arguments from Furstenberg's presentation in [Fur14] and Tao's presentation in [Tao09], while extending the relative dichotomy and the Furstenberg–Zimmer structure theorem to the setting of arbitrary countable discrete abelian groups.

One can also formalize and prove a relative dichotomy and a Furstenberg–Zimmer structure theorem for the measure-preserving action of arbitrary discrete groups [Jam23, EHK24].

In Lectures 5, 6, and 7, we proved characterizations of compact and weakly mixing systems. Analogous characterizations exist in the setting of compact extensions and weakly mixing extensions, and we will address these after proving Furstenberg's multiple recurrence theorem (Theorem 1.1.7) by an induction on the Furstenberg–Zimmer tower in Theorem 9.1.12 for $\mathbb{Z}$-actions, and via the Furstenberg correspondence principle, also Szemerédi's theorem in the next lectures.

9.3 Exercises

Exercise 9.1. Prove Proposition 9.1.4.

Exercise 9.2. Prove Proposition 9.1.6.

In the next two exercises, we consider the skew-rotation $(\mathbb{T}^2, \tau_a \rtimes c)$ from Example 7.2.6, which is the cocycle extension of the torus rotation $(\mathbb{T}, \tau_a)$, $a \in \mathbb{T}$, by the identity map $c := \mathrm{id}_{\mathbb{T}} \colon \mathbb{T} \to \mathbb{T}$. This is a measure-preserving $\mathbb{Z}$-system if we equip $\mathbb{T}^2$ with the Haar measure (check this!).

Exercise 9.3. Let a not be a root of unity. Let $f \in \mathrm{L}^2(\mathbb{T}^2 \mid \mathbb{T})$ be the function defined by

$$f(x, y) := y^n \quad \text{whenever } n \geq 1 \text{ and } x \in \left\{ e^{2\pi i z} \colon z \in \left(\frac{1}{n+1}, \frac{1}{n} \right] \right\}.$$

Show that f is not conditionally almost periodic but is a limit of a sequence $(f_n)_{n\in\mathbb{N}}$ of conditionally almost periodic elements $f_n \in \mathrm{L}^2(\mathbb{T}^2 \mid \mathbb{T})$ for $n \in \mathbb{N}$.

Exercise 9.4.
(i) Identify the Kronecker subsystem (see Definition 7.1.14) of the skew-rotation $(\mathbb{T}^2, \tau_a \rtimes c)$.
(ii) Show that $(\mathbb{T}^2, \tau_a \rtimes c)$ is a compact extension of its Kronecker subsystem.

Exercise 9.5. Let (Y, σ) be a measure-preserving $\mathbb{Z}$-system, let G be a compact metrizable group with a closed subgroup H, let $\varphi : Y \to G$ be measurable (where G is equipped with the Borel σ-algebra), and let $Y \times_\varphi G/H$ be the extension of Y with underlying space $Y \times G/H$, with measure equal to the product of μ_Y and Haar measure, and shift map $T : (y, g) \mapsto (\sigma(y), \varphi(y)g)$. Show that $Y \times_\varphi G/H$ is a compact extension of Y.

Exercise 9.6. Establish the bound (9.4).

Lecture 10

In this lecture and the next, we prove van der Waerden's theorem (cf. Theorem 4.2.2) and Szemerédi's theorem (cf. Theorem 1.1.8).

In both cases, we will use dynamics. For van der Waerden's theorem, we employ topological dynamics, translating the problem into a multiple recurrence statement within a shift system.

For Szemerédi's theorem, we use ergodic theory. Specifically, we establish Szemerédi's theorem by proving Furstenberg's multiple recurrence theorem (cf. Theorem 1.1.7) and applying Furstenberg's correspondence principle (cf. Theorem 4), as demonstrated in the special case of Roth's theorem in Lecture 8. This will be achieved through induction on the Furstenberg–Zimmer tower (cf. Theorem 9.1.12). More precisely, we show that the multiple recurrence property is preserved under compact extensions, weakly mixing extensions, and inverse limits.

To achieve this, we will use van der Waerden's theorem to prove the preservation of the multiple recurrence property under compact extensions. This will be addressed in this lecture, alongside the dynamical proof of van der Waerden's theorem.

10.1 Van der Waerden's Theorem

We restate van der Waerden's theorem:

Theorem 10.1.1 (Infinitary version)**.** *Partition the natural numbers $\mathbb{N}$ into finitely many cells $\mathbb{N} = A_1 \cup \cdots \cup A_r$. Then there exists one cell A_j that contains arithmetic progressions of arbitrary (finite) length.*

Before translating the infinitary version of van der Waerden's theorem into a statement about multiple recurrence in topological dynamics, we state the following finitary version, which will be useful in the proof of Szemerédi's theorem. The equivalence of the infinitary version and the finitary version is left as Exercise 10.1.

Theorem 10.1.2 (Finitary Version)**.** *Let $r \in \mathbb{N}$ and $k \in \mathbb{N}$. Then there exists* *a smallest number $W(r,k) \in \mathbb{N}$ such that any coloring of $\{1,\dots,N\}$, where $N \geq W(r,k)$, into r colors contains a monochromatic arithmetic progression of length k.*

Remark 10.1.3. The numbers $W(r,k)$ are known as *van der Waerden numbers*, and it remains an open and important problem in Ramsey theory to establish "reasonable bounds" for the values of $W(r,k)$ for most values of r and k. The best currently known upper bound is due to Gowers [Gow01], who proved that

$$W(r,k) \leq 2^{2^{r^{2^{2^{k+9}}}}}.$$

It is a fun combinatorial exercise to show that $W(2,3) = 9$.

Let $\mathbb{N} = C_1 \cup \dots \cup C_r$ be a partition of the natural numbers. Define $D_0 = \{0\}$, $D_i := C_i$ for $1 \leq i \leq r$, and $D_i := -C_{i-r}$ for $r+1 \leq i \leq 2r$. This construction yields a partition of the integers $\mathbb{Z} = D_0 \cup D_1 \cup \dots \cup D_{2r}$. Now, suppose one of the cells D_i contains an arithmetic progression $a, a+n, \dots, a+(k-1)n$ of length $k \geq 2$. Then either $a, a+n, \dots, a+(k-1)n \in C_j$ for some $1 \leq j \leq r$, or $a, a+n, \dots, a+(k-1)n \in -C_j$ for some $1 \leq j \leq r$. In the latter case, the negative sequence $-(a+(k-1)n), \dots, -(a+n), -a$ lies in C_j. Thus, we conclude that in order to prove Theorem 10.1.1, it suffices to prove a version of the theorem where $\mathbb{N}$ is replaced by $\mathbb{Z}$.

We start with a more general definition.

Definition 10.1.4. Let $k \geq 1$. A set $P \subseteq \mathbb{Z}^k$ is said to be a **van der Waerden collection** if for every finite partition $\mathbb{Z} = C_1 \cup C_2 \cup \ldots \cup C_r$, there exist $(\gamma_1, \gamma_2, \dots, \gamma_k) \in P$ and $j \in \{1, 2, \dots, r\}$ such that $\{\gamma_1, \gamma_2, \dots, \gamma_k\} \subseteq C_j$.

By the pigeonhole principle, to prove Theorem 10.1.1, it is enough to prove:

Theorem 10.1.5. *For every $k \geq 1$, the arithmetic progressions of length k in $\mathbb{Z}$ form a van der Waerden collection.*

We will translate the statement in Theorem 10.1.5 into a statement about the existence of multiply recurrent points in a topological dynamical system (K, τ) over $\Gamma = \mathbb{Z}$. Throughout this lecture, we will consider topological dynamical systems (K, τ) over $\mathbb{Z}$, where K is a metrizable compact space.

We provide the dynamical counterpart of a van der Waerden collection:

Definition 10.1.6. Let $k \geq 1$. A set $P \subseteq \mathbb{Z}^k$ is said to be a **Birkhoff collection** if for every topological dynamical system (K, τ), $x \in K$, and any $\varepsilon > 0$, there exists $(\gamma_1, \gamma_2, \dots, \gamma_k) \in P$ such that $\tau^{\gamma_1} x, \tau^{\gamma_2} x, \dots, \tau^{\gamma_k} x$ are pairwise ε-close.

In the next section, we will prove the following multiple recurrence theorem:

Theorem 10.1.7. *Let (K, τ) be a topological dynamical system and let d be a metric* *for K. For any $k \in \mathbb{N}$ and any $\varepsilon > 0$, there exist $x \in K$ and $\gamma \in \mathbb{N}$ such that $d(\tau^{j\gamma}x, x) < \varepsilon$ for all $j = 1, 2, \ldots, k$.*

Assuming Theorem 10.1.7 for now, in order to prove Theorem 10.1.5, it remains to establish the following two results.

Proposition 10.1.8. *Let $k \in \mathbb{N}$. Then the set of k-term arithmetic progressions is a Birkhoff collection.*

Proof. Let $x \in K$ and let L be the orbit closure of x. By Theorem 10.1.7, there are $y \in L$ and $n \in \mathbb{N}$ such that

$$y, \tau^n(y), \tau^{2n}(y), \ldots, \tau^{(k-1)n}(y)$$

are pairwise ε-close to each other. Since y is in the orbit closure of x, there is $m \in \mathbb{Z}$ such that

$$y, \tau^n(y), \tau^{2n}(y), \ldots, \tau^{(k-1)n}(y)$$

are ε-close to

$$\tau^m(x), \tau^{n+m}(x), \tau^{2n+m}(x), \ldots, \tau^{(k-1)n+m}(x)$$

respectively. By the triangle inequality,

$$\tau^m(x), \tau^{n+m}(x), \tau^{2n+m}(x), \ldots, \tau^{(k-1)n+m}(x)$$

are pairwise 3ε-close to each other. This finishes the proof. □

Lemma 10.1.9. *Every Birkhoff collection is a van der Waerden collection.*

Proof. Assume that P is a Birkhoff collection. Fix a partition $\mathbb{Z} = C_1 \cup C_2 \cup \ldots \cup C_r$ and consider the topological dynamical system $K = \{1, 2, \ldots, r\}^{\mathbb{Z}}$ with the shift map given by $\tau((x_n)_{n\in\mathbb{Z}}) := (x_{n-1})_{n\in\mathbb{Z}}$ (cf. Example 3.2.5). We consider the metric d on K given by

$$d(x, y) = \begin{cases} 2^{-k} & \text{if } x \neq y \text{ and } k = \min\{|n| \mid x_n \neq y_n\}, \\ 0 & \text{if } x = y. \end{cases}$$

Note that $d(x, y) < 1$ if and only if $x_0 = y_0$.

Define the point $x \in K$ by setting $x_n = i$ if $n \in C_i$. Let $1 > \varepsilon > 0$. Since P is a Birkhoff collection, there exists $(\gamma_1, \gamma_2, \ldots, \gamma_k) \in P$ such that $\tau^{\gamma_1}(x), \tau^{\gamma_2}(x), \ldots, \tau^{\gamma_k}(x)$ are pairwise ε-close. By the choice of d, this implies $x_{\gamma_1} = x_{\gamma_2} = \ldots = x_{\gamma_k}$, and thus $\{\gamma_1, \gamma_2, \ldots, \gamma_k\} \subseteq C_j$ for some $j \in \{1, 2, \ldots, r\}$. □

10.2 Minimality and Recurrence

In this section, we prove Theorem 10.1.7. We need some preparation.

Definition 10.2.1. Let (K, τ) be a topological dynamical system. A subset $L \subseteq K$ is said to be **τ-invariant** if $\tau(L) = L$. We say that $L \subseteq K$ is **minimal** if L is closed, τ-invariant, and contains no proper, closed, non-empty τ-invariant subsets. The system (K, τ) is called **minimal** if K is a minimal subset.

Remark 10.2.2. Any minimal subset $L \subseteq K$ of a topological dynamical system (K, τ) induces a subsystem (L, σ) of (K, τ), where σ is the restriction of τ to L.

Proposition 10.2.3. *Every topological dynamical system (K, τ) admits a minimal subsystem.*

Proof. Let $\mathcal{M}$ be the family of all non-empty closed τ-invariant subsets L of K. Since $K \in \mathcal{M}$, $\mathcal{M}$ is not empty. Define a partial order on $\mathcal{M}$ by inclusion.

For any totally ordered collection $\mathcal{C} \subseteq \mathcal{M}$, the intersection of any finite number of elements in $\mathcal{C}$ is non-empty. As K is compact, the intersection of all elements in $\mathcal{C}$ is non-empty. Denote this intersection by L_0. Since each element in $\mathcal{C}$ is closed and τ-invariant, L_0 is also closed and τ-invariant. Therefore, $L_0 \in \mathcal{M}$, and L_0 is a lower bound for $\mathcal{C}$.

By Zorn's Lemma, the family $\mathcal{M}$ has a minimal element. □

We also need the following notions (cf. Exercise 4.7).

Definition 10.2.4. A subset $A \subseteq \mathbb{Z}$ is said to be **syndetic** if there exists a finite set $F \subseteq \mathbb{Z}$ such that

$$\bigcup_{\gamma \in F} (A + \gamma) = \mathbb{Z},$$

that is, if finitely many translates of A cover $\mathbb{Z}$.

Let (K, τ) be a topological dynamical system. A point $x \in K$ is said to be **uniformly recurrent** if for every neighborhood U of x, the set

$$\{\gamma \in \mathbb{Z} : \tau^\gamma(x) \in U\}$$

is a syndetic subset of $\mathbb{Z}$.

Theorem 10.2.5. *In a minimal topological dynamical system (K, τ), every point of K is uniformly recurrent.*

Proof. Let $x \in K$ and let U be an open neighborhood of x. By Exercise 10.2 and compactness, there exist $\gamma_1, \dots, \gamma_n \in \mathbb{Z}$ such that

$$\bigcup_{i=1}^{n} \tau^{-\gamma_i}(U) = K.$$

For any $\gamma \in \mathbb{Z}$, $\tau^{\gamma}(x) \in \tau^{-\gamma_i}(U)$ for some $1 \leq i \leq n$. Thus, $\tau^{\gamma+\gamma_i}(x) \in U$.

Hence, the set $\{\gamma \in \mathbb{Z} : \tau^{\gamma}(x) \in U\}$ is syndetic, and therefore x is uniformly recurrent. □

If $x \in K$ is uniformly recurrent, then x is also **recurrent** in the sense that for every neighborhood U of x, there exists $\gamma \in \mathbb{N}$ such that $\tau^{\gamma}(x) \in U$.

In this sense, Theorem 10.1.7 extends Theorem 10.2.5 to the setting of multiple recurrence.

We need the following two preliminary lemmas.

Lemma 10.2.6 (Lebesgue's Covering Lemma)**.** *Let K be a compact metric space, and let $\mathcal{O}$ be an open covering of K. Then there exists $\varepsilon > 0$ such that for all $x \in K$ there exists $O \in \mathcal{O}$ such that $B(x, \varepsilon) \subseteq O$.*

Proof. For a proof, see [Sin19, Lemma 5.3.9]. □

Lemma 10.2.7. *Let (K, τ) be a minimal topological dynamical system and let $k \in \mathbb{N}$. Suppose that for every $\varepsilon > 0$ there exist $x \in K$ and $\gamma \in \mathbb{N}$ such that $d(\tau^{j\gamma}(x), x) < \varepsilon$ for all $j = 1, 2, \dots, k$. Then for any $\varepsilon > 0$, there exists a dense subset $D \subseteq K$ such that for each $y \in D$, there exists $\gamma \in \mathbb{N}$ with $d(\tau^{j\gamma}(y), y) < \varepsilon$ for all $j = 1, 2, \dots, k$.*

Proof. Let $\varepsilon > 0$ and let U be an open ball of radius ε. By minimality and compactness, there exists a finite set $H \subseteq \mathbb{Z}$ such that K is covered by the H-translates of U. By Lemma 10.2.6, we find $\delta > 0$ such that for each $z \in K$ there is $\gamma \in H$ with $B(x, \delta) \subseteq \tau^{\gamma}(U)$.

By assumption, there exist $z_0 \in K$ and $\gamma' \in \mathbb{N}$ such that $d(\tau^{j\gamma'}(z_0), z_0) < \delta$ for all $j = 1, 2, \dots, k$. By the choice of δ, there exists some $\tilde{\gamma} \in H$ such that $B(z_0, \delta) \subseteq \tau^{\tilde{\gamma}}(U)$. Then $\tau^{-\tilde{\gamma}}(B(z_0, \delta)) \subseteq U$, and setting $w := \tau^{-\tilde{\gamma}}(z_0) \in U$, we have $d(w, \tau^{j\gamma'}(w)) < \varepsilon$ for all $j = 1, \dots, k$. Since ε, U were arbitrary, this proves the claim. □

Proof of Theorem 10.1.7. By Proposition 10.2.3, we may assume without loss of generality that (K, τ) is minimal. The proof is by induction on k. The base case $k = 1$ is Theorem 10.2.5.

Assume that the statement holds for some $k \geq 1$, meaning that for any $\varepsilon > 0$, there exist $x \in K$ and $\gamma \in \mathbb{N}$ such that $d(\tau^{j\gamma}x, x) < \varepsilon$ for $j = 1, 2, \dots, k$. By Lemma 10.2.7, for each $\varepsilon > 0$, there is a dense set $D \subseteq K$ such that for all $y \in D$, there exists $\gamma \in \mathbb{N}$ with $d(\tau^{j\gamma}y, y) < \varepsilon$ for $j = 1, 2, \ldots, k$. We show that the same conclusion holds for $k+1$.

Fixing $\varepsilon > 0$, we can choose $x_0 \in K$ and an integer $\gamma_0 \in \mathbb{N}$ such that $d(\tau^{j\gamma_0}x_0, x_0) < \varepsilon/2$ for $j = 0, 1, 2, \ldots, k$. Since τ is a bijection, we can choose $x_1 \in K$ such that $\tau^{\gamma_0}x_1 = x_0$. Then for $j = 1, 2, \ldots, k$,

$$d(\tau^{(j+1)\gamma_0}x_1, x_0) = d(\tau^{j\gamma_0}\tau^{\gamma_0}x_1, x_0) = d(\tau^{j\gamma_0}x_0, x_0) < \varepsilon/2.$$

This means that for $j = 1, 2, \ldots, k+1$,

$$d(\tau^{j\gamma_0}x_1, x_0) < \varepsilon/2.$$

Since τ is continuous, the same conclusion holds in some neighborhood of x_1. Thus we can choose ε_1 with $0 < \varepsilon_1 < \frac{\varepsilon}{2}$ such that $d(\tau^{j\gamma_0}y, x_0) < \varepsilon/2$ for $j = 1, 2, \ldots, k+1$ and for all $y \in B(x_1, \varepsilon_1)$. By the inductive assumption, there exists a point $y_1 \in B(x_1, \varepsilon_1/2)$ and $\gamma_1 \in \mathbb{Z}$ such that

$$d(\tau^{j\gamma_1}y_1, y_1) < \varepsilon_1/2 \text{ for } j = 1, 2, \ldots, k.$$

This means that y_1 and $\tau^{j\gamma_1}y_1$, for $j = 1, 2, \ldots, k$, lie in $B(x_1, \varepsilon_1)$. Thus for $j = 1, 2, \ldots, k+1$,

$$d(\tau^{j\gamma_0}(\tau^{(j-1)\gamma_1}y_1), x_0) < \varepsilon/2.$$

Taking any point $x_2 \in K$ such that $\tau^{\gamma_1}x_2 = y_1$, we have

$$d(\tau^{j\gamma_1}x_2, x_1) < \varepsilon_1 < \varepsilon/2$$

for $j = 1, 2, \ldots, k+1$, as well as:

$$d(\tau^{j(\gamma_1+\gamma_0)}x_2, x_0) < \varepsilon/2.$$

Inductively, we find $x_0, x_1, x_2, \ldots \in K$ and $\gamma_1, \gamma_2, \gamma_3, \ldots \in \mathbb{N}$ such that for any $i \in \mathbb{N}$ and for $j = 1, 2, \ldots, k+1$,

$$d(\tau^{j\gamma_i}x_i, x_{i-1}) < \varepsilon/2,$$
$$d(\tau^{j(\gamma_{i-1}+\gamma_{i-2})}x_i, x_{i-2}) < \varepsilon/2,$$

and

$$d(\tau^{j(\gamma_{i-1}+\ldots+\gamma_0)}x_i, x_0) < \varepsilon/2.$$

By compactness of K, there exist integers $0 < m < l$ such that $d(x_l, x_m) < \varepsilon/2$. Thus

$$d(\tau^{j(\gamma_{l-1}+\ldots+\gamma_m)}x_l, x_l) \leq d(x_l, x_m) + d(x_m, \tau^{j(\gamma_{l-1}+\ldots+\gamma_m)}x_l) < \varepsilon$$

for $j = 1, 2, \ldots, k+1$. Taking $x = x_l$ and $\gamma = \gamma_{l-1} + \ldots + \gamma_m$, we have produced a point $x \in K$ such that $d(\tau^{j\gamma}x, x) < \varepsilon$ for $j = 1, 2, \ldots, k+1$. $\square$

10.3 Multiple Recurrence and Compact Extensions

We will establish the multiple recurrence statement for measure-preserving systems in Theorem 4.2.10.

Definition 10.3.1. We say that a measure-preserving system (X,T) over $\Gamma = \mathbb{Z}$ has the **multiple recurrence property** if for every $k \in \mathbb{N}$ and each $f \in \mathrm{L}^\infty(X)$ with $f \geq 0, \int_X f \mathrm{d}\mu_X > 0$ we have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f \cdot U_T^n f \cdots U_T^{kn} f > 0.$$

Our goal is to show that every measure-preserving system (X,T) over $\Gamma = \mathbb{Z}$ has the multiple recurrence property. Clearly, trivial systems have this property. Using the Furstenberg–Zimmer structure theorem (see Theorem 9.1.12) and an induction argument, it now suffices to show that the multiple recurrence property is preserved by compact extensions, weakly mixing extensions and inverse limits of measure-preserving systems. We start with the first of these three problems.

Theorem 10.3.2. *Let $q\colon (X,\tau) \to (Y,\sigma)$ be a compact extension of concrete measure-preserving systems over $\Gamma = \mathbb{Z}$. If (Y,σ) has the multiple recurrence property, then so does (X,τ).*

We first prove two lemmas.

Lemma 10.3.3. *Let $q\colon (X,\tau) \to (Y,\sigma)$ be a compact extension of concrete measure-preserving systems over $\Gamma = \mathbb{Z}$. If $A \subseteq X$ is measurable with $\mu_X(A) > 0$, we find a measurable subset $B \subseteq A$ with $\mu_X(B) > 0$ and the following two properties.*

(i) *$\mathbb{1}_B$ is conditionally almost periodic.*

(ii) *For almost every $y \in Y$ we have $\mathbb{E}(\mathbb{1}_B \mid Y)(y) > \frac{1}{2}\mu_X(B)$ or $\mathbb{E}(\mathbb{1}_B \mid Y)(y) = 0$.*

Proof. We first build a subset with property (ii). Consider the element

$$C \coloneqq \{y \in Y \mid \mathbb{E}(\mathbb{1}_A \mid Y)(y) > {}^{\mu_X(A)}/_2\} \in \Sigma(Y)$$

and $A' \coloneqq A \cap q^*(C) \in \Sigma(X)$. Then $\mathbb{E}(\mathbb{1}_{A'} \mid Y) = \mathbb{E}(U_q(\mathbb{1}_C)\mathbb{1}_A \mid Y) = \mathbb{1}_C\mathbb{E}(\mathbb{1}_A \mid Y)$, and therefore clearly

$$\mathbb{E}(\mathbb{1}_{A'} \mid Y)(y) > {}^{\mu_X(A)}/_2 \geq {}^{\mu_X(A')}/_2$$

for almost every $y \in C$, and $\mathbb{E}(\mathbb{1}_{A'} \mid Y)(y) = 0$ for almost every $y \in Y \setminus C$.

Assuming that A' is a nullset, we have

$$\mu_X(A) = \mu_X(A \cap q^*(Y \setminus C)) = \int_{Y\setminus C} \mathbb{E}(\mathbb{1}_A \mid Y) \leq {}^{\mu_X(A)}/_2,$$

a contradiction. We conclude that $\mu_X(A') > 0$.

Replacing A by A' we may therefore assume that property (ii) holds (where $B = A$). We now construct $B \subseteq A$ with $\mu_X(B) > 0$, which still satisfies (ii), but in addition also property (i). For $n \in \mathbb{N}$ let $\varepsilon_n := \frac{\mu_X(A)}{2^{n+1}}$ and choose a conditionally almost periodic $f_n \in \mathrm{L}^2(X \mid Y)$ with $\|\mathbb{1}_A - f_n\|_{\mathrm{L}^2(X)} < \varepsilon_n$. Let further

$$C_n := \{y \in Y \mid \|\mathbb{1}_A - f_n\|_{X|Y}^2(y) \geq \varepsilon_n\} \in \Sigma(Y) \text{ for } n \in \mathbb{N}$$

and consider $B := A \setminus \bigcup_{n\in\mathbb{N}} q^*(C_n) \in \Sigma(X)$. Since

$$\mu_Y(C_n) \leq \frac{1}{\varepsilon_n}\int_{C_n} \|\mathbb{1}_A - f_n\|_{X|Y}^2 \leq \frac{1}{\varepsilon_n}\int_Y \|\mathbb{1}_A - f_n\|_{X|Y}^2 = \frac{1}{\varepsilon_n}\|\mathbb{1}_A - f_n\|_{\mathrm{L}^2(X)}^2 \leq \varepsilon_n$$

for each $n \in \mathbb{N}$, we obtain that

$$\mu_X(B) \geq \mu_X(A) - \sum_{n=1}^{\infty} \mu_Y(C_n) \geq \mu_X(A) - \sum_{n=1}^{\infty} \frac{\mu_X(A)}{2^{n+1}} = \frac{1}{2}\mu_X(A) > 0.$$

We show that B still has property (ii). Write

$$E := \left\{y \in Y : \mathbb{E}(\mathbb{1}_A \mid Y)(y) \geq \frac{1}{2}\mu_X(A)\right\} \in \Sigma(Y).$$

Since $\mathbb{E}(\mathbb{1}_B \mid Y) = \mathbb{1}_{Y\setminus\bigcup_{n\in\mathbb{N}} C_n}\mathbb{E}(\mathbb{1}_A \mid Y)$ we have $\mathbb{E}(\mathbb{1}_B \mid Y)(y) \geq \frac{1}{2}\mu_X(A) \geq \frac{1}{2}\mu_X(B)$ for almost every $y \in E \setminus \bigcup_{n\in\mathbb{N}} C_n$, and $\mathbb{E}(\mathbb{1}_B \mid Y)(y) = 0$ for almost every $y \in (Y \setminus E) \cup \bigcup_{n\in\mathbb{N}} C_n$.

We finally prove that it also satisfies (i), i.e., that $\mathbb{1}_B$ is conditionally almost periodic. For $\varepsilon > 0$ let $k \in \mathbb{N}$ with $\sqrt{\varepsilon_k} \leq \frac{\varepsilon}{2}$. For $m \in \mathbb{Z}$ we then have

$$\begin{aligned}
\mathbb{1}_{Y\setminus\sigma^{-m}(\bigcup_{n\in\mathbb{N}} C_n)}\|U_\tau^m \mathbb{1}_B - U_\tau^m f_k\|_{X|Y} &= U_\sigma^m(\mathbb{1}_{Y\setminus\bigcup_{n\in\mathbb{N}} C_n}\|\mathbb{1}_B - f_k\|_{X|Y}) \\
&= U_\sigma^m(\mathbb{1}_{Y\setminus\bigcup_{n\in\mathbb{N}} C_n}\|\mathbb{1}_A - f_k\|_{X|Y}) \\
&\leq \sqrt{\varepsilon_k}\mathbb{1}_{Y\setminus\sigma^{-m}(\bigcup_{n\in\mathbb{N}} C_n)} \leq \frac{\varepsilon}{2}\mathbb{1}_{Y\setminus\sigma^{-m}(\bigcup_{n\in\mathbb{N}} C_n)}.
\end{aligned}$$

On the other hand,

$$\mathbb{1}_{\sigma^{-m}(\bigcup_{n\in\mathbb{N}} C_n)}\|U_\tau^m \mathbb{1}_B\|_{X|Y} = \|U_\tau^m(\mathbb{1}_{q^*(\bigcup_{n\in\mathbb{N}} C_n)\cap B})\|_{X|Y} = 0.$$

Since f_k is conditionally almost periodic, we find $g_1, \dots, g_l \in \mathrm{L}^2(X \mid Y)$ with

$$\inf_{1\leq j\leq l} \|U_\tau^m f_k - g_j\|_{X|Y} \leq \frac{\varepsilon}{2}$$

for every $m \in \mathbb{Z}$. Setting $g_0 := 0$ we then have $\inf_{0\leq j\leq l}\|U_\tau^m \mathbb{1}_B - g_j\|_{X|Y} \leq \varepsilon\mathbb{1}$ for each $m \in \mathbb{Z}$ by the triangle inequality. □

Lemma 10.3.4. *For a sequence* $(a_n)_{\in\mathbb{N}}$ *in* $[0,1]$ *the following assertions are equivalent.*

(a) $\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} a_n > 0$.

(b) *There is some* $c > 0$ *such that* $\underline{\mathrm{d}}(\{n \in \mathbb{N} \mid a_n > c\}) > 0$.

Proof. Let $c > 0$. Writing $E_N := \{n \in \{1,\dots,N\} \mid a_n > c\}$ for $N \in \mathbb{N}$ we obtain

$$\frac{1}{N}\sum_{n=1}^{N} a_n \geq \frac{1}{N}\sum_{n\in E_N} a_n \geq c\cdot\frac{|E_N|}{N}$$

and

$$\frac{1}{N}\sum_{n=1}^{N} a_n \leq \frac{1}{N}\sum_{n\in E_N} 1 + \frac{1}{N}\sum_{n\notin E_N} a_n \leq \frac{|E_N|}{N} + \frac{1}{N}\sum_{n\notin E_N} c \leq \frac{|E_N|}{N} + c.$$

This yields

$$c\cdot\underline{\mathrm{d}}(\{n\in\mathbb{N} \mid a_n > c\}) \leq \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} a_n \leq \underline{\mathrm{d}}(\{n\in\mathbb{N} \mid a_n > c\}) + c$$

for every $c > 0$ which readily implies the desired equivalence. □

We now prove the lifting property for compact extensions.

Proof of Theorem 10.3.2. We take $f \in \mathrm{L}^\infty(X)$ with $f \geq 0$, $\int_X f\,\mathrm{d}\mu_X > 0$ and $k \in \mathbb{N}$. Since there is $A \in \Sigma(X)$ with $\mu_X(A) > 0$ and $a\mathbb{1}_A \leq f$ for some $a > 0$, we may assume that $f = \mathbb{1}_A$ for $A \in \Sigma(X)$ from the get go. Applying Lemma 10.3.3 we may further assume that there is a measurable non-null subset $B \subseteq Y$ such that $\mathbb{E}(\mathbb{1}_A \mid Y)(y) > \frac{1}{2}\mu_X(A)$ for every $y \in B$ and that $\mathbb{1}_A$ is conditionally almost periodic. For $\varepsilon := \frac{\mu_X(A)}{6k} > 0$ choose $g_1,\dots,g_m \in \mathrm{L}^2(X \mid Y)$ such that

$$\inf_{1\leq i\leq m}\|U_\tau^n\mathbb{1}_A - g_i\|_{X|Y} \leq \varepsilon\mathbb{1}$$

for every $n \in \mathbb{Z}$.

Now let $N := W(m, k+1)$ be the van der Waerden number (see Theorem 10.1.2) for arithmetic progressions of length $k+1$ and m colors, and consider the finite index set

$$I := \{(i,d,j) \in \mathbb{N}^3 \mid i \leq N, i + kd \leq N, j \leq m\}.$$

We further abbreviate

$$B_n := B \cap \sigma^{-n}(B) \cap \cdots \cap \sigma^{-Nn}(B) \in \Sigma(Y)$$

for $n \in \mathbb{N}$. Since (Y, σ) has the multiple recurrence property we can apply Lemma 10.3.4 to find some $c > 0$ such that the set

$$D \coloneqq \{n \in \mathbb{N} \mid \mu_Y(B_n) \geq c\}$$

satisfies $\underline{\mathrm{d}}(D) > 0$. For $n \in D$ we claim that

$$\exists d \in \{1, \ldots, N\} \colon \mu_X(A \cap \tau^{-dn}(A) \cap \cdots \cap \tau^{-kdn}(A)) \geq \frac{c}{6|I|}\mu_X(A). \tag{10.1}$$

We deduce this statement below. Let us however first show how this claim implies the desired assertion

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f \cdot U_\tau^n f \cdots U_\tau^{kn} f > 0.$$

Using that $f = \mathbb{1}_A$ and applying Lemma 10.3.4, it suffices to check that the set

$$E \coloneqq \left\{ n \in \mathbb{N} \middle| \mu_X(A \cap \tau^{-n}(A) \cap \cdots \cap \tau^{-kn}(A)) \geq \frac{c}{6|I|}\mu_X(A) \right\}$$

satisfies $\underline{\mathrm{d}}(E) > 0$. Let $M_0 \in \mathbb{N}$ be such that

$$|D \cap \{1, \ldots, M\}]| \geq \frac{M\underline{\mathrm{d}}(D)}{2}$$

for all $M \geq M_0$. Let $L \geq M_0 N$ and write $L = MN + l$ for some $M \geq M_0$ and $l \in \{0, \ldots, N-1\}$. For each $n \in D \cap \{1, \ldots, M\}$ we use (10.1) to choose some $d(n) \in \{1, \ldots, N\}$ with $n \cdot d(n) \in E$. For every $d \in \{1, \ldots, N\}$ the map

$$\{n \in D \cap \{1, \ldots, M\} \mid d(n) = d\} \to E \cap \{1, \ldots, NM\}, \quad n \mapsto d(n)n$$

is clearly injective. Therefore, for any given $n' \in E \cap \{1, \ldots, NM\}$ there are at most N many distinct elements in $D \cap \{1, \ldots, M\}$ with $n' = d(n)n$. Thus,

$$|E \cap \{1, \ldots, L\}| \geq |E \cap \{1, \ldots, MN\}| \geq \frac{1}{N}|D \cap \{1, \ldots, M\}]| \geq \frac{M\underline{\mathrm{d}}(D)}{2N},$$

hence

$$\frac{|E \cap \{1, \ldots, L\}|}{L} \geq \frac{M\underline{\mathrm{d}}(D)}{2NL} = \frac{MN\underline{\mathrm{d}}(D)}{2N^2L} = \frac{(L-l)\underline{\mathrm{d}}(D)}{2N^2L} \geq \frac{L-(N-1)}{L} \cdot \frac{\underline{\mathrm{d}}(D)}{2N^2}.$$

It follows that $\underline{\mathrm{d}}(E) \geq \frac{\underline{\mathrm{d}}(D)}{2N^2} > 0$. Thus, if we can show (10.1) for each $n \in D$, we have proven the theorem.

Now fix $n \in D$ and show (10.1). We choose a representative $h_{j,l}$ of $\|U_\tau^{ln}\mathbb{1}_A - g_j\|_{X|Y} \in \mathrm{L}^\infty(Y)$ for each $l \in \{0,\dots,N\}$ and $j \in \{1,\dots,m\}$. For every $l \in \{0,\dots,N\}$ we then find a nullset Z_l in Y such that $\min_{j=1,\dots,m} h_{j,l}(y) \le \varepsilon$ for all $y \in Y \setminus Z_l$, and write $Z := \bigcup_{l=0}^N Z_l$. By setting all representatives $h_{j,l}$ to 0 on Z for $l \in \{0,\dots,N\}$ and $j \in \{1,\dots,m\}$, we may even assume that $\min_{j=1,\dots,m} h_{j,l}(y) \le \varepsilon$ holds for all $y \in Y$. Consider the measurable sets

$$C_{i,d,j} := \bigcap_{l=0}^{k} \{y \in Y \mid h_{j,(i+ld)}(y) \le \varepsilon\} \subseteq Y$$

for all $(i,d,j) \in I$. We claim that $\bigcup_{(i,d,j)\in I} C_{i,d,j} = Y$. In fact, if $y \in Y$, then

$$\{1,\dots,N\} = \bigcup_{j=1}^{m} \{l \in \{1,\dots,N\} \mid h_{j,l}(y) \le \varepsilon\},$$

and by choice of N as the van der Waerden number $W(m,k+1)$ (and disjointifying the sets) we find some $(i,d,j) \in I$ such that $h_{j,i+ld}(y) \le \varepsilon$ holds for all $l \in \{0,\dots,k\}$, i.e., $y \in C_{i,d,j}$.

By the pigeonhole principle we now find some element $(i,d,j) \in I$ such that $\mu_Y(C_{i,d,j} \cap B_n) \ge \frac{\mu_Y(B_n)}{|I|}$. For almost every $y \in \sigma^{in}(C_{i,d,j} \cap B_n)$ we have

$$\|U_\tau^{ldn}\mathbb{1}_A - U_\tau^{-in} g_j\|_{X|Y}(y) = \|U_\tau^{(i+ld)n}\mathbb{1}_A - g_j\|_{X|Y}(\sigma^{-in}y) = h_{j,i+ld}(\sigma^{-in}y) \le \varepsilon,$$

and therefore

$$\begin{aligned}\|U_\tau^{ldn}\mathbb{1}_A - \mathbb{1}_A\|_{X|Y}(y) &\le \|U_\tau^{ldn}\mathbb{1}_A - U_\tau^{-in} g_j\|_{X|Y}(y) + \|U_\tau^{-in} g_j - U_\tau^{0dn}\mathbb{1}_A\|_{X|Y}(y)\\ &\le 2\varepsilon = \frac{\mu_X(A)}{3k}\end{aligned}$$

for all $l \in \{0,\dots,k\}$. By an application of the conditional Cauchy–Schwarz inequality and a telescopic sum this yields for almost every $y \in \sigma^{in}(C_{i,d,j} \cap B_n)$ that

$$|\mathbb{E}(\mathbb{1}_A \cdot U_\tau^{dn}\mathbb{1}_A \cdots U_\tau^{kdn}\mathbb{1}_A \mid Y) - \mathbb{E}(\mathbb{1}_A^{k+1} \mid Y)|(y) \le \frac{\mu_X(A)}{3},$$

and thus

$$\mathbb{E}(\mathbb{1}_A \cdot U_\tau^{dn}\mathbb{1}_A \cdots U_\tau^{kdn}\mathbb{1}_A \mid Y)(y) \ge \mathbb{E}(\mathbb{1}_A \mid Y)(y) - \frac{\mu_X(A)}{3}$$

since $\mathbb{1}_A^{k+1} = \mathbb{1}_A$. Using that $\sigma^{in}(C_{i,d,j} \cap B_n) \subseteq \sigma^{in}(B_n) \subseteq B$, we further obtain that $\mathbb{E}(\mathbb{1}_A \mid Y)(y) > \frac{1}{2}\mu_X(A)$ for almost every $y \in \sigma^{in}(C_{i,d,j} \cap B_n)$ (by choice of B). But then

$$\mathbb{E}(\mathbb{1}_A \cdot U_\tau^{dn}\mathbb{1}_A \cdots U_\tau^{kdn}\mathbb{1}_A \mid Y)(y) \ge \frac{\mu_X(A)}{6}$$

for almost every $y \in \sigma^{in}(C_{i,d,j} \cap B_{n,N})$. We integrate over Y to finally obtain

$$\begin{aligned}
\mu_X(A \cap \tau^{-dn}(A) \cap \cdots \cap \tau^{-kdn}(A)) &= \int_Y \mathbb{E}(\mathbb{1}_A \cdot U_\tau^{dn}\mathbb{1}_A \cdots U_\tau^{kdn}\mathbb{1}_A \mid Y) \\
&\geq \int_{\sigma^{in}(C_{i,d,j}\cap B_n)} \mathbb{E}(\mathbb{1}_A \cdot U_\tau^{dn}\mathbb{1}_A \cdots U_\tau^{kdn}\mathbb{1}_A \mid Y) \\
&\geq \frac{\mu_X(A)}{6}\mu_Y(\sigma^{in}(C_{i,d,j} \cap B_n)) \geq \frac{\mu_Y(B_n) \cdot \mu_X(A)}{6|I|} \\
&\geq \frac{c\mu_X(A)}{6|I|}
\end{aligned}$$

since $n \in D_N$. This shows (10.1). □

10.4 Comments and Further Reading

Solving a conjecture of Schur, van der Waerden proved Theorem 10.1.1 in [vdW27] using combinatorial methods. The translation of this result into the dynamical formulation of Theorem 10.1.7, along with its proof, was first provided by Furstenberg and Weiss in [FW78]. Our proof is based on a proof from unpublished lecture notes by Bryna Kra.

There are several ways to demonstrate that compact extensions preserve the multiple recurrence property. Here, we adopt an approach by Bergelson (see [Ber06, Theorem 4.2.17] and also [EW11, Subsection 7.9.1]), which utilizes van der Waerden's theorem (see also [Tao09] and [Fur14, Section 7.3]).

In his original ergodic-theoretic approach to Szemerédi's theorem [Fur77], Furstenberg established the lifting property for compact extensions between ergodic systems by leveraging a representation result (similar to the use of the Halmos–von Neumann theorem in the proof of Theorem 8.2.1). We will discuss such a representation result later in this course.

For an alternative proof of the lifting property for compact extensions, see [FKO82] and [EW11, Subsection 7.9.2].

10.5 Exercises

Exercise 10.1. Show that Theorem 10.1.1 is equivalent to Theorem 10.1.2.

Exercise 10.2. Show that a topological dynamical system (K, τ) is minimal if and only if the orbit closure of any element $x \in K$ is K itself.

Exercise 10.3. Show that a syndetic set in $\mathbb{Z}$ contains arbitrarily long arithmetic progressions.

Exercise 10.4. Show that the conclusion of van der Waerden's Theorem does not hold for infinite length progressions, meaning show that there exists a finite partition of $\mathbb{N}$ such that no piece contains an infinite length arithmetic progression.

Exercise 10.5. Use van der Waerden's Theorem to show that for all $\alpha \in \mathbb{R}$ and $\varepsilon > 0$ there exist $m, n \in \mathbb{N}$ such that $|n^2\alpha - m| < \varepsilon$. Generalize this for any polynomial $p(n)$ with integer coeffiecents such that $p(0) = 0$.

Exercise 10.6. Prove *Grünwald's theorem*: For any finite partition $\mathbb{N}^m = C_1 \cup C_2 \cup \cdots \cup C_r$, and any $k \geq 1$, there exist some C_j, some $d \in \mathbb{N}$, and some $b \in \mathbb{N}^m$ such that

$$b + d(x_1, x_2, \ldots, x_m) \subseteq C_j, \quad 1 \leq x_i \leq k,\ 1 \leq i \leq m.$$

Lecture 11

This lecture is divided into two parts. In the first part, we establish that the multiple recurrence property is preserved under weakly mixing extensions and under taking inverse limits. Since it was demonstrated in the previous lecture that the multiple recurrence property is also preserved under compact extensions, applying the Furstenberg–Zimmer structure theorem completes the proof of Furstenberg's multiple recurrence theorem. Through Furstenberg's correspondence principle, this result also establishes Szemerédi's theorem.

In the second part of this lecture (and continuing into the next lecture), we delve deeper into the understanding and classification of compact extensions. This classification of compact extensions will be useful in the final segment of our ISem lectures, where we introduce a significant enhancement of Furstenberg–Zimmer structure theory, known as the Host–Kra or Host–Kra–Ziegler structure theory.

11.1 Multiple Recurrence and Weakly Mixing Extensions

In this section, we establish that the multiple recurrence property as defined in Definition 10.3.1 is preserved under weakly mixing extensions:

Theorem 11.1.1. *Let $q\colon (X,\tau) \to (Y,\sigma)$ be a weakly mixing extension of concrete measure-preserving system over $\Gamma = \mathbb{Z}$. If (Y,σ) has the multiple recurrence property, then so does (X,τ).*

Theorem 11.1.1 is an easy consequence of the following statement, see Exercise 11.1.

Proposition 11.1.2. *Let $q\colon (X,\tau) \to (Y,\sigma)$ be a weakly mixing extension of concrete measure-preserving system over $\Gamma = \mathbb{Z}$. For all $k \in \mathbb{N}$ and $f_1, \dots, f_k \in \mathrm{L}^\infty(X)$*

we have

$$\lim_{N\to\infty}\left[\frac{1}{N}\sum_{n=0}^{N-1}U_\tau^n f_1\cdots U_\tau^{kn}f_k - U_q\biggl(\frac{1}{N}\sum_{n=0}^{N-1}U_\sigma^n\mathbb{E}(f_1\mid Y)\cdots U_\sigma^{kn}\mathbb{E}(f_k\mid Y)\biggr)\right]=0.$$

Proof. We show the statement by induction on $k\in\mathbb{N}$. For $k=1$ and $f_1\in \mathrm{L}^\infty(X)$ we obtain that the limit

$$\begin{aligned}h &\coloneqq \lim_{N\to\infty}\left[\frac{1}{N}\sum_{n=0}^{N-1}U_\tau^n f_1 - U_q\biggl(\frac{1}{N}\sum_{n=0}^{N-1}U_\sigma^n\mathbb{E}(f_1\mid Y)\biggr)\right]\\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}U_\tau^n(f_1-U_q\mathbb{E}(f_1\mid Y))\end{aligned}$$

exists by the mean ergodic theorem and is an element of $\mathrm{L}^\infty(X)$. However, since $f\coloneqq f_1-U_q\mathbb{E}(f_1\mid Y)\in\mathrm{L}^\infty(X)$ satisfies $\mathbb{E}(f\mid Y)=0$, we obtain

$$\begin{aligned}(h|h)_{\mathrm{L}^2(X)} &= \left|\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}(U_\tau^n f|h)_{\mathrm{L}^2(X)}\right| = \left|\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int_Y(U_\tau^n f|h)_{X|Y}\right|\\ &\leq \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\|(U_\tau^n f|h)_{X|Y}\|_{\mathrm{L}^2(Y)}=0\end{aligned}$$

by the definition of weakly mixing extensions and Exercise 7.2 (i). This shows $h=0$.

Now let $k\in\mathbb{N}$ and assume that the claim holds for $k-1$. Let $f_1,\dots,f_k\in\mathrm{L}^\infty(X)$ and abbreviate $g_i\coloneqq\mathbb{E}(f_i\mid Y)$ for $i\in\{1,\dots,k\}$. By telescopic summing we can write

$$\begin{aligned}&U_\tau^n f_1\cdots U_\tau^{kn}f_k - U_q(U_\sigma^n g_1\cdots U_\sigma^{kn}g_k)\\ &=\sum_{j=1}^{k}U_\tau^n f_1\cdots U_\tau^{(j-1)n}f_{j-1}\cdot U_\tau^{jn}(f_j-U_qg_j)\cdot U_\tau^{(j+1)n}U_qg_{j+1}\cdots U_\tau^{kn}U_qg_k\end{aligned}$$

for each $n\in\mathbb{N}$. It therefore suffices to show that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}U_\tau^n h_1\cdots U_\tau^{kn}h_k=0$$

whenever $\mathbb{E}(h_i\mid Y)=0$ for some $i\in\{1,\dots,k\}$. We may assume that $\|h_i\|_\infty\leq 1$ for all $i\in\{1,\dots,k\}$. We want to apply the van der Corput Lemma (see Lemma 8.2.3) and set $a_n\coloneqq U_\tau^n h_1\cdots U_\tau^{kn}h_k$ for $n\in\mathbb{N}$. Then

$$\begin{aligned}(a_n|a_{n+m}) &= \int_X U_\tau^n h_1\cdots U_\tau^{kn}h_k\cdot\overline{U_\tau^{n+m}h_1\cdots U_\tau^{k(n+m)}h_k}\\ &=\int_X g_{1,m}\cdot U_\tau^n g_{2,m}\cdots U_\tau^{(k-1)n}g_{k,m}\end{aligned}$$

for $n, m \in \mathbb{N}$ where $g_{j,m} := h_j U_\tau^{jm} \overline{h_j}$ for $j \in \{1, \ldots, k\}$. Now consider the numbers $c_m := \limsup_{N\to\infty} |\frac{1}{N} \sum_{n=0}^{N-1} (a_n|a_{n+m})|$ for $m \in \mathbb{N}$. As a consequence of the induction hypothesis and the Cauchy–Schwarz inequality we obtain

$$\begin{aligned} c_m &= \limsup_{N\to\infty} \left| \frac{1}{N} \sum_{n=0}^{N-1} \int_X g_{1,m} \cdot U_q(U_\sigma^n \mathbb{E}(g_{2,m} \mid Y) \cdots U_\sigma^{(k-1)n} \mathbb{E}(g_{k,m} \mid Y)) \right| \\ &= \limsup_{N\to\infty} \left| \frac{1}{N} \sum_{n=0}^{N-1} \int_Y \mathbb{E}(g_{1,m} \mid Y) \cdot U_\sigma^n \mathbb{E}(g_{2,m} \mid Y) \cdots U_\sigma^{(k-1)n} \mathbb{E}(g_{k,m} \mid Y) \right| \end{aligned}$$

for every $m \in \mathbb{N}$. Now choose $i \in \{1, \ldots, k\}$ with $\mathbb{E}(h_i \mid Y) = 0$. Since $\|h_j\|_\infty \leq 1$ for all $j \in \{1, \ldots, k\}$, we obtain that

$$\begin{aligned} c_m &\leq \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_Y U_\sigma^{(i-1)n} |\mathbb{E}(g_{i,m} \mid Y)| = \int_Y |\mathbb{E}(h_i U_\tau^{im} \overline{h_i} \mid Y)| \\ &\leq \|(h_i | U_\tau^{im} h_i)_{X|Y}\|_{\mathrm{L}^2(Y)} \end{aligned}$$

for every $m \in \mathbb{N}$. Since $h_i \in \mathrm{L}^\infty(X)$ is weakly mixing, a similar reasoning as in the proof of Lemma 8.2.4 (using Exercise 7.2 (i)) shows that

$$\begin{aligned} \limsup_{M\to\infty} \frac{1}{M} \sum_{m=0}^{M-1} c_m &\leq \limsup_{M\to\infty} \frac{1}{M} \sum_{m=0}^{M-1} \|(h_i | U_\tau^{im} h_i)_{X|Y}\|_{\mathrm{L}^2(Y)} \\ &\leq i \limsup_{M\to\infty} \frac{1}{iM} \sum_{m=0}^{iM-1} \|(h_i | U_\tau^{m} h_i)_{X|Y}\|_{\mathrm{L}^2(Y)} = 0. \end{aligned}$$

The van der Corput Lemma (see Lemma 8.2.3) therefore yields the claim. □

11.2 Multiple recurrence and Inverse Limits

In this section, we establish that the multiple recurrence property, as defined in Definition 10.3.1, is preserved under taking inverse limits. The following lemma reduces this task to showing that a function in the inverse limit has relatively dense support within the subsystems forming the inverse limit.

Lemma 11.2.1. *Let $q\colon (X, \tau) \to (Y, \sigma)$ be an extension of concrete measure-preserving systems over $\Gamma = \mathbb{Z}$. Suppose that (Y, σ) satisfies the multiple recurrence property. Let $k \geq 1$ be an integer and let $f \in \mathrm{L}^\infty(X)$ with $f \geq 0$, $\int_X f \,\mathrm{d}\mu_X > 0$ such that*

$$\mu_Y \left(\left\{ \mathbb{E}(\mathbb{1}_{\{f>0\}} \mid Y) > 1 - \frac{1}{k} \right\} \right) > 0.$$

Then f satisfies

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f \cdot U_\tau^n f \cdots U_\tau^{kn} f > 0.$$

Proof. Let $E_n = \{f > \frac{1}{n}\}$. Applying the monotone convergence theorem for conditional expectations (see Lemma 8.1.1) to the sequence $\mathbb{1}_{E_1} \leq \mathbb{1}_{E_2} \leq \ldots$, we find some $\varepsilon > 0$ such that

$$\mu_Y\left(\left\{\mathbb{E}(\mathbb{1}_{\{f>\varepsilon\}} \mid Y) > 1 - \left(\frac{1}{k} - \varepsilon\right)\right\}\right) > 0.$$

Denote $F = \left\{\mathbb{E}(\mathbb{1}_{\{f>\varepsilon\}} \mid Y) > 1 - \left(\frac{1}{k} - \varepsilon\right)\right\}$.

Since (Y, σ) satisfies the multiple recurrence property, by Lemma 10.3.4 there exists some $c > 0$ such that the set of $n \in \mathbb{Z}$ satisfying

$$\mu_Y\big(F \cap \sigma^{-n}(F) \cap \ldots \cap \sigma^{-(k-1)n}(F)\big) > c \tag{11.1}$$

has positive lower density.

From $f \geq \varepsilon \cdot \mathbb{1}_E$, where $E = \{f > \varepsilon\}$, we obtain

$$f U_{\tau^n}(f) \cdots U_{\tau^{(k-1)n}}(f) \geq \varepsilon^k \mathbb{1}_{E \cap \tau^{-n}(E) \cap \ldots \cap \tau^{-(k-1)n}(E)}.$$

Using the equality $\mathbb{1}_A = \mathbb{1} - \mathbb{1}_{A^c}$, we further get

$$f U_{\tau^n}(f) \cdots U_{\tau^{(k-1)n}}(f) \geq \varepsilon^k \left(\mathbb{1} - \sum_{j=0}^{k-1} \mathbb{1}_{\tau^{-jn}(E^c)}\right).$$

Taking conditional expectations, we obtain

$$\mathbb{E}(f U_{\tau^n}(f) \cdots U_{\tau^{(k-1)n}}(f) \mid Y) \geq \varepsilon^k \left(\mathbb{1} - \sum_{j=0}^{k-1} U_{\sigma^{jn}}(\mathbb{E}(\mathbb{1}_{E^c} \mid Y))\right).$$

Since $\mathbb{1}_{E^c} = \mathbb{1} - \mathbb{1}_E$, we can write

$$\mathbb{E}(f U_{\tau^n}(f) \cdots U_{\tau^{(k-1)n}}(f) \mid Y) \geq \varepsilon^k \left(\sum_{j=0}^{k-1} U_{\sigma^{jn}}(\mathbb{E}(\mathbb{1}_E \mid Y)) - (k-1)\mathbb{1}\right).$$

By the definition of F, we know

$$\mathbb{E}(\mathbb{1}_E \mid Y) > 1 - \left(\frac{1}{k} - \varepsilon\right) \text{ on } F.$$

Thus,

$$\mathbb{E}(fU_{\tau^n}(f)\cdots U_{\tau^{(k-1)n}}(f)\mid Y)\geq \varepsilon^{k+1}\cdot k \text{ on } F\cap\sigma^{-n}(F)\cap\ldots\cap\sigma^{-(k-1)n}(F).$$

Integrating this inequality over Y and using (11.1), we conclude that the set of $n\in\mathbb{Z}$ such that

$$\int_X fU_{\tau^n}(f)\cdots U_{\tau^{(k-1)n}}(f)\,\mathrm{d}\mu_X\geq c\cdot\varepsilon^{k+1}\cdot k$$

has positive lower density. The claim follows by applying Lemma 10.3.4. □

We recall the definition of an inverse limit of a totally ordered family of subsystems.

Let β be a countable limit ordinal, and let $(Y_\alpha,\sigma_\alpha)_{\alpha<\beta}$ be a totally ordered family of factors of a fixed system (X,τ) over $\Gamma=\mathbb{Z}$. By this, we mean that $\mathrm{L}^2(Y_\alpha),\alpha<\beta$, when identified with subspaces of $\mathrm{L}^2(X)$, form an increasing family of closed, invariant Markov sublattices of $\mathrm{L}^2(X)$. Let (Y_β,σ_β) be the concrete measure-preserving system corresponding to the invariant Markov sublattice $\overline{\bigcup_{\alpha<\beta}\mathrm{L}^2(Y_\alpha)}$. Without further mention, for every $\alpha<\beta$, we identify $\mathrm{L}^2(Y_\alpha)$ with the Markov sublattice $U_{q_\alpha}(\mathrm{L}^2(Y_\alpha))$ of $\mathrm{L}^2(Y_\beta)$ where $q_\alpha\colon(Y_\beta,\sigma_\beta)\to(Y_\alpha,\sigma_\alpha)$ denotes the corresponding factor map.

We now establish that the multiple recurrence property is preserved under taking inverse limits:

Theorem 11.2.2. *Let β be a countable limit ordinal, let $(Y_\alpha,\sigma_\alpha)_{\alpha<\beta}$ be a totally ordered family of concrete measure-preserving subsystems of a fixed system (X,τ) over $\Gamma=\mathbb{Z}$, and let (Y_β,σ_β) be the inverse limit of the family $(Y_\alpha,\sigma_\alpha)_{\alpha<\beta}$ in the above sense. If every $(Y_\alpha,\sigma_\alpha),\alpha<\beta$, satisfies the multiple recurrence property, then so does (Y_β,σ_β).*

Proof. Let $f\in\mathrm{L}^\infty(Y_\beta)$ be such that $f\geq 0,\int_{Y_\beta}f=c>0$, and without loss of generality assume that $\|f\|_{\mathrm{L}^\infty(Y_\beta)}\leq 1$. Since (Y_β,σ_β) is the inverse limit of the (Y_α,σ_α), the orthogonal projections $\mathbb{E}(f\mid Y_\alpha)$ converge in $\mathrm{L}^2(X)$ norm to $\mathbb{E}(f\mid Y_\beta)=f$. Thus, for any $\varepsilon>0$, there is $\alpha<\beta$ such that

$$\|f-\mathbb{E}(f\mid Y_\alpha)\|_{\mathrm{L}^2(X)}\leq\varepsilon. \tag{11.2}$$

By Lemma 8.1.1, $\int_{Y_\beta}\mathbb{E}(f\mid Y_\alpha)\,\mathrm{d}\mu_{Y_\beta}=c$ and $\|\mathbb{E}(f\mid Y_\alpha)\|_{\mathrm{L}^\infty(Y_\alpha)}\leq 1$. Therefore $F:=\{\mathbb{E}(f\mid Y_\alpha)\geq c/2\}$ must have measure at least $c/2$. Setting $E:=\{f>0\}$, we have

$$|f-\mathbb{E}(f\mid Y_\alpha)|\geq\frac{c}{2}\mathbb{1}_F\mathbb{1}_{E^c}.$$

Squaring this and taking conditional expectations, we obtain

$$\mathbb{E}(|f-\mathbb{E}(f\mid Y_\alpha)|^2\mid Y_\alpha)\geq\frac{c^2}{4}(\mathbb{1}-\mathbb{E}(\mathbb{1}_E\mid Y_\alpha))\mathbb{1}_F. \tag{11.3}$$

Let $k \geq 1$ be an integer. By Markov's inequality[1], and then applying first (11.3) and second (11.2), we obtain

$$\begin{aligned}\mu_X(\{(\mathbb{1} - \mathbb{E}(\mathbb{1}_E \mid Y_\alpha))\mathbb{1}_F \geq 1/k\}) &\leq k \int_X (\mathbb{1} - \mathbb{E}(\mathbb{1}_E \mid Y_\alpha))\mathbb{1}_F \, \mathrm{d}\mu_X \\ &\leq \frac{4k}{c^2} \|f - \mathbb{E}(f \mid Y_\alpha)\|^2_{\mathrm{L}^2(X)} \\ &\leq \frac{4k\varepsilon^2}{c^2}.\end{aligned}$$

Thus

$$\mu_X(\{(\mathbb{1} - \mathbb{E}(\mathbb{1}_E \mid Y_\alpha))\mathbb{1}_F < 1/k\}) \geq 1 - \frac{4k\varepsilon^2}{c^2}.$$

Choosing ε sufficiently small depending on c, we conclude (from the lower bound $\mu_X(F) \geq c/2$) that

$$\mathbb{E}(\mathbb{1}_E \mid Y_\alpha) > 1 - 1/k$$

on a set of positive measure. The claim now follows from Lemma 11.2.1. □

We have now completed the proof of Furstenberg's multiple recurrence theorem (Theorem 4.2.10) and, as a consequence, Szemerédi's theorem (Theorem 4.2.6).

11.3 Classification of Compact Extensions, Part I

In Lectures 5 and 6, we defined a measure-preserving system (X, τ) to have discrete spectrum[2] if $\mathrm{L}^2(X)$ is spanned by finite-dimensional invariant subspaces. In Exercise 6.6, we suggested that this definition is equivalent to the property that the orbit of every element of $\mathrm{L}^2(X)$ is precompact. This latter characterization inspired us to formalize a relative version for extensions of measure-preserving systems within the conditional Hilbert space $\mathrm{L}^2(X \mid Y)$ in Lecture 9. We termed such functions conditionally almost periodic and defined an extension $q\colon (X, \tau) \to (Y, \sigma)$ of concrete measure-preserving systems to be compact if the set of conditionally almost periodic elements of $\mathrm{L}^2(X \mid Y)$ is dense in $\mathrm{L}^2(X)$.

In this section, we establish that there is an analogous notion of discrete spectrum for extensions, formulated in terms of finite rank invariant $\mathrm{L}^\infty(Y)$-submodules of

[1] Markov's inequality states that for any extended real-valued measurable function f on an arbitrary measure space (X, μ_X) and any $\varepsilon > 0$ it holds that

$$\mu_X(\{|f| \geq \varepsilon\}) \leq \frac{1}{\varepsilon} \int_X |f| \mathrm{d}\mu_X.$$

[2] It is also customary to call systems with discrete spectrum compact systems.

the conditional Hilbert space $\mathrm{L}^2(X \mid Y)$. Furthermore, we show that this notion is equivalent to the compactness of the extension.

Before delving into the details, let us recall the construction of the conditional Hilbert space:

$$\mathrm{L}^2(X \mid Y) = \{f \in \mathrm{L}^2(X)\colon \|\mathbb{E}(|f|^2 \mid Y)\|_{\mathrm{L}^\infty(Y)} < \infty\},$$

which is equipped with the conditional norm

$$\|f\|_{X|Y} = \mathbb{E}(|f|^2 \mid Y)^{1/2} \in \mathrm{L}^\infty(Y).$$

The following provides a replacement for finite-dimensional subspaces in the definition of discrete spectrum within the relative setting of extensions:

Definition 11.3.1. A set $M \subseteq \mathrm{L}^2(X \mid Y)$ is said to be **conditionally orthonormal** if:

(i) $(f \mid g)_{X|Y} = 0$ for all $f, g \in M$ with $f \neq g$.

(ii) For every $f \in M$, there exists a measurable set $E \subseteq Y$ such that $(f \mid f)_{X|Y} = \mathbb{1}_E$.

The $\mathrm{L}^\infty(Y)$-linear hull

$$\mathcal{H} = \left\{ \sum_{i=1}^n h_i f_i \mid h_1, \dots, h_n \in \mathrm{L}^\infty(Y) \right\}$$

of a finite conditionally orthonormal subset $M = \{f_1, \dots, f_n\} \subseteq \mathrm{L}^2(X \mid Y)$ is called a **finite rank** $\mathrm{L}^\infty(Y)$**-submodule** (generated by $\{f_1, \dots, f_n\}$).

The following orthogonal expansion for elements of finite rank submodules will be useful at several points. The proof is left as Exercise 11.2.

Lemma 11.3.2. *Let $\mathcal{H} \subseteq \mathrm{L}^2(X \mid Y)$ be a finite rank $\mathrm{L}^\infty(Y)$-submodule generated by a conditionally orthonormal subset $M = \{f_1, \dots, f_n\}$. Then every $f \in \mathcal{H}$ can be expressed as*

$$f = \sum_{i=1}^n (f \mid f_i)_{X|Y} f_i.$$

Our next major goal is to establish the following implication (the reverse implication is left as Exercise 11.3):

Theorem 11.3.3. *Let $q\colon (X, \tau) \to (Y, \sigma)$ be a compact extension of concrete measure-preserving systems. Then the union of all invariant finite rank $\mathrm{L}^\infty(Y)$-submodules of $\mathrm{L}^2(X \mid Y)$ is dense in $\mathrm{L}^2(X)$.*

The strategy to prove this theorem is to relate the assumptions to conditional Hilbert–Schmidt operators and derive the conclusion through a spectral analysis of these conditional operators. We start by relating the assumptions in Theorem 11.3.3 to conditional Hilbert–Schmidt operators.

Recall that $X \otimes_Y X$ denotes the product of the probability space X with itself relative to the probability space Y (see Definition 8.4.5). For $K \in \mathrm{L}^2(X \otimes_Y X \mid Y)$, recall that we define the conditional kernel operator $K *_Y \colon \mathrm{L}^2(X) \to \mathrm{L}^2(X)$ by

$$(K *_Y f)(x) \coloneqq \int_Y K(x, x') f(x') d\mu_{q(x)}(x').$$

Proposition 11.3.4. *Let $K \in \mathrm{L}^2(X \otimes_Y X \mid Y)$. The conditional kernel operator $K *_Y \colon \mathrm{L}^2(X) \to \mathrm{L}^2(X)$ is a well-defined bounded linear operator.*

Proof. Suppose $\|K\|_{\mathrm{L}^2(X \otimes_Y X|Y)} = C$ for some constant $C > 0$. Recall the inequality[3] (9.4)

$$\|K *_Y f\|_{X|Y} \leq \|K\|_{X \otimes_Y X|Y} \|f\|_{X|Y},$$

which yields

$$\begin{aligned} \|K *_Y f\|^2_{\mathrm{L}^2(X)} &= \int_Y \|K *_Y f\|^2_{X|Y} \mathrm{d}\mu_Y \\ &\leq \int_Y \|K\|^2_{X \otimes_Y X|Y} \|f\|^2_{X|Y} \mathrm{d}\mu_Y \\ &\leq C^2 \int_Y \|f\|^2_{X|Y} \mathrm{d}\mu_Y \\ &= C^2 \|f\|^2_{\mathrm{L}^2(X)}. \end{aligned}$$

This proves the claim. □

We will now argue that $K *_Y \colon \mathrm{L}^2(X \mid Y) \to \mathrm{L}^2(X \mid Y)$ is a conditional Hilbert–Schmidt operator in the following sense:

Definition 11.3.5. A $\mathbb{C}$-linear map $V \colon \mathrm{L}^2(X \mid Y) \to \mathrm{L}^2(X \mid Y)$ is said to be **conditionally Hilbert–Schmidt** if it satisfies the following two properties:

(i) $\mathrm{L}^\infty(Y)$**-linearity:** For all $h \in \mathrm{L}^\infty(Y)$ and $f \in \mathrm{L}^2(X \mid Y)$,

$$V(U_q(h) f) = U_q(h) V(f).$$

(ii) **Conditional Hilbert–Schmidt property:** There exists $C > 0$ such that for all conditionally orthonormal sets $M \subseteq \mathrm{L}^2(X \mid Y)$,

$$\sum_{f \in M} \|V(f)\|^2_{X|Y} \leq C \mathbb{1}.$$

[3]This inequality is established for $f \in \mathrm{L}^2(X \mid Y)$, but one can easily check that it also holds for f in the larger space $\mathrm{L}^2(X)$.

Remark 11.3.6. Remember that we assume that X is a Lebesgue space, and consequently, $\mathrm{L}^2(X)$ is separable. This separability implies that any conditional orthonormal set is at most countable since such a set defines an ordinary suborthonormal set. To see this, note that if $(f \mid g)_{X|Y} = 0$, then by integrating both sides, it follows that $(f \mid g)_{\mathrm{L}^2(X)} = 0$. Similarly, $(f \mid f)_{X|Y} = \mathbb{1}_E$ implies $(f \mid f)_{\mathrm{L}^2(X)} \leq 1$.

Proposition 11.3.7. *Let $K \in \mathrm{L}^2(X \otimes_Y X \mid Y)$. The conditional kernel operator $K{*_Y}\colon \mathrm{L}^2(X \mid Y) \to \mathrm{L}^2(X \mid Y)$ is conditionally Hilbert–Schmidt.*

Proof. $\mathrm{L}^\infty(Y)$-linearity follows from Proposition 9.1.4(iii). Let $C > 0$ be such that $\|K\|_{X \otimes_Y X|Y} \leq C$. Let $M \subseteq \mathrm{L}^2(X \mid Y)$ be a conditionally orthonormal set. By Theorem 8.4.3,

$$\sum_{f \in M} \|K *_Y f\|^2_{X|Y} = \sum_{f\in M} \int_X \left| \int_X K(x,x')f(x')\mathrm{d}\mu_{q(x)}(x') \right|^2 \mathrm{d}\mu_y(x)$$

The measure μ_y is supported on the fiber $q^{-1}(\{y\})$ for almost every y, and the measure $\mu_X \otimes_Y \mu_X$ on $X \times X$ is supported on the fiber product of sets $\{(x_1, x_2) \in X \times X : q(x_1) = q(x_2)\}$. Thus for almost every y,

$$\sum_{f\in M} \int_X \left| \int_X K(x,x')f(x')\mathrm{d}\mu_{q(x)}(x') \right|^2 \mathrm{d}\mu_y(x) = \sum_{f\in M} \int_X \left| \int_X K(x,x')f(x')\mathrm{d}\mu_{y}(x') \right|^2 \mathrm{d}\mu_y(x)$$

By Remark 11.3.6, M is at most countable. Furthermore, by Theorem 8.4.3, M is also an ordinary suborthonormal set in $\mathrm{L}^2(X, \mu_y)$ for almost every y. Hence by monotone convergence and Bessel's inequality for almost every y,

$$\sum_{f\in M} \int_X \left| \int_X K(x,x')f(x')\mathrm{d}\mu_{y}(x') \right|^2 \mathrm{d}\mu_y(x) \leq \int_{X\times X} |K(x,x')|^2 \,\mathrm{d}\mu_y \times \mu_y(x,x') \leq C^2.$$

The claim follows from another application of Theorem 8.4.3. □

The following result forms the first step in our strategy to relate the compactness of an extension to the range of certain conditional Hilbert–Schmidt operators:

Proposition 11.3.8. *Let $q\colon (X,\tau) \to (Y,\sigma)$ be a compact extension of concrete measure-preserving systems. Then*

$$M = \{K *_Y f : K \in \mathrm{fix}(U_{\tau\times\tau}) \cap \mathrm{L}^2(X \otimes_Y X \mid Y),\ f \in \mathrm{L}^2(X \mid Y)\}$$

spans a dense linear subspace of $\mathrm{L}^2(X)$.

We use the following lemma.

Lemma 11.3.9. *Let $q\colon (X,\tau)\to(Y,\sigma)$ be a compact extension of concrete measure-preserving systems and $f\in \mathrm{L}^2(X\mid Y)$. For each $\varepsilon>0$ and $\delta>0$ there are finitely many $g_1,\dots,g_n\in\mathrm{L}^2(X\mid Y)$ such that for each $\gamma\in\Gamma$ there is some $A_\gamma\in\Sigma(Y)$ with $\mu_Y(A_\gamma)\geq 1-\delta$ and*

$$\mathbb{1}_{A_\gamma}\inf_{1\leq i\leq n}\|U_{\tau_\gamma}f-g_i\|_{X|Y}\leq\varepsilon\mathbb{1}. \tag{11.4}$$

Proof. Take $\varepsilon>0$ and $\delta>0$. Since, by the definition of compact extensions, the conditionally almost periodic elements in $\mathrm{L}^2(X|Y)$ are dense in $\mathrm{L}^2(X)$, we can find some conditionally almost periodic $g\in\mathrm{L}^2(X\mid Y)$ with $\|f-g\|_{\mathrm{L}^2(X)}\leq\frac{\varepsilon\sqrt{\delta}}{2}$ and set

$$B\coloneqq\left\{y\in Y\colon \|f-g\|_{X|Y}(y)\leq\frac{\varepsilon}{2}\right\}\in\Sigma(Y).$$

Using that $\|f-g\|^2_{\mathrm{L}^2(X)}=\int_Y\|f-g\|^2_{X|Y}$, one can readily check that $\mu_Y(B)\geq 1-\delta$. Using that $g\in\mathrm{L}^2(X\mid Y)$ is conditionally almost periodic, we further find $g_1,\dots,g_n\in\mathrm{L}^2(X\mid Y)$ for some $n\in\mathbb{N}$ such that

$$\inf_{1\leq i\leq n}\|U_{\tau_\gamma}(g)-g_i\|_{X|Y}\leq\frac{\varepsilon}{2}\mathbb{1}$$

for each $\gamma\in\Gamma$.

Then $A_\gamma\coloneqq\sigma_\gamma^{-1}(B)\in\Sigma(Y)$ satisfies $\mu_Y(A_\gamma)=\mu_Y(B)\geq 1-\delta$. Moreover, for almost every $y\in A_\gamma$ we have that

$$\|U_{\tau_\gamma}f-U_{\tau_\gamma}g\|_{X|Y}(y)=\|f-g\|_{X|Y}(\sigma_\gamma(y))\leq\frac{\varepsilon}{2},$$

hence

$$\begin{aligned}\inf_{i=1,\dots,n}\|U_{\tau_\gamma}f-g_i\|_{X|Y}(y)&\leq\|U_{\tau_\gamma}f-U_{\tau_\gamma}g\|_{X|Y}(y)+\inf_{i=1,\dots,n}\|U_{\tau_\gamma}g-g_i\|_{X|Y}(y)\\&\leq\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.\end{aligned}$$

□

Proof of Proposition 11.3.8. It suffices to show that $M^\perp=\{0\}$. For $f\in M^\perp$ we observe that also $\mathbb{1}_{[\|f\|_{X|Y}\leq n]}f\in M^\perp$ for each $n\in\mathbb{N}$ (in fact, $M^\perp$ is a $\mathrm{L}^\infty(Y)$-submodule of $\mathrm{L}^2(X)$). By approximation we therefore only need to consider the case $f\in\mathrm{L}^2(X\mid Y)$.

Let $f\in M^\perp\cap\mathrm{L}^2(X\mid Y)$ and consider $f\odot\bar{f}\in\mathrm{L}^2(X\otimes_Y X\mid Y)$. Let K be the unique element of minimal $\mathrm{L}^2(X\otimes_Y X)$-norm in the closed convex hull of the orbit of $f\odot\bar{f}$

as guaranteed by the abstract ergodic theorem (see Theorem 3.1.5). Then K is an invariant element of $\mathrm{L}^2(X \otimes_Y X|Y)$, and thus, by assumption, f is orthogonal to $K *_Y f$. By Theorem 8.4.3,

$$\begin{aligned} 0 = (f \mid K *_Y f)_{\mathrm{L}^2(X)} &= \int_X f(x) \int_X K(x,x')\overline{f(x')} \mathrm{d}\mu_{q(x)}(x')\,\mathrm{d}\mu_X(x) \\ &= \int_Y \int_{X\times X} K(x,x')f(x)\overline{f(x')} \mathrm{d}(\mu_y \times \mu_y)(x,x')\,\mathrm{d}\mu_Y(y) \\ &= (K \mid f \odot \bar{f})_{\mathrm{L}^2(X\otimes_Y X)}. \end{aligned}$$

Thus, $f \odot \bar{f}$ is orthogonal to K, and since K is invariant, we conclude that $U_{\tau_\gamma \times \tau_\gamma}(f \odot \bar{f})$ is orthogonal to K for every $\gamma \in \Gamma$. Hence, K is orthogonal to itself, and therefore $K = 0$. We therefore find a sequence (K_m) in the convex hull of the orbit of $f \odot \bar{f}$ such that $\|K_m\|_{\mathrm{L}^2(X\otimes_Y X)} \to 0$.

Now let $g_1, \ldots, g_n \in \mathrm{L}^2(X \mid Y)$ be chosen as in Lemma 11.3.9. By the Cauchy–Schwarz inequality,

$$\sum_{i=1}^{n} (K_m \mid g_i \odot \bar{g}_i)_{\mathrm{L}^2(X\otimes_Y X)} \to 0.$$

For a given m, writing K_m as a convex combination

$$K_m = \sum_{j=1}^{l_m} \lambda_{j,m} U_{\tau_{\gamma_{j,m}} \times \tau_{\gamma_{j,m}}}(f \odot \bar{f}),$$

and applying Proposition 8.4.6, we obtain

$$\begin{aligned} \sum_{i=1}^{n} (K_m \mid g_i \odot \bar{g}_i)_{\mathrm{L}^2(X\otimes_Y X)} &= \sum_{i=1}^{n}\sum_{j=1}^{l_m} \lambda_{j,m} \int_Y \mathbb{E}(U_{\tau_{\gamma_{j,m}}}(f)g_i \mid Y)\overline{\mathbb{E}(U_{\tau_{\gamma_{j,m}}}(f)g_i \mid Y)}\,\mathrm{d}\mu_Y \\ &= \sum_{j=1}^{l_m} \lambda_{j,m} \left(\sum_{i=1}^{n} \|(U_{\tau_{\gamma_{j,m}}}(f) \mid g_i)_{X|Y}\|^2_{\mathrm{L}^2(Y)} \right). \end{aligned}$$

Since the latter convex combination converges to zero, we can always find $\gamma \in \Gamma$ such that

$$\sum_{i=1}^{n} \|(U_{\tau_\gamma}(f) \mid g_i)_{X|Y}\|^2_{\mathrm{L}^2(Y)}$$

is arbitrarily small. Thus for any $\varepsilon > 0$ and $\delta > 0$ there is $A \in \Sigma(Y)$ with $\mu_Y(A) \geq 1 - \frac{\delta}{2}$ and $\gamma \in \Gamma$ such that, for all $1 \leq i \leq n$,

$$|(U_{\tau_\gamma}(f) \mid g_i)_{X|Y}|\mathbb{1}_A \leq \varepsilon \mathbb{1}. \tag{11.5}$$

For this $\gamma \in \Gamma$ we then find, by the choice of $g_1, \ldots, g_n$ via Lemma 11.3.9, some $A_\gamma \in \Sigma(Y)$ with $\mu_Y(A_\gamma) \geq 1 - \frac{\delta}{2}$ such that

$$\mathbb{1}_{A_\gamma} \inf_{1 \leq i \leq n} \|U_{\tau_\gamma} f - g_i\|_{X|Y} \leq \varepsilon \mathbb{1}. \tag{11.6}$$

Meanwhile, using the conditional Pythagorean identity in Proposition 9.1.4, for all $1 \leq i \leq n$,

$$\|U_{\tau_\gamma}(f) - g_i\|_{X|Y}^2 = \|U_{\tau_\gamma}(f)\|_{X|Y}^2 - 2\mathrm{Re}\,(U_{\tau_\gamma}(f) \mid g_i)_{X|Y} + \|g_i\|_{X|Y}^2,$$

and thus,

$$\|U_{\tau_\gamma}(f)\|_{X|Y}^2 \leq \|U_{\tau_\gamma}(f) - g_i\|_{X|Y}^2 + 2\mathrm{Re}\,(U_{\tau_\gamma}(f) \mid g_i)_{X|Y}.$$

By combining this inequality with (11.5), we obtain with $B \coloneqq A \cap A_\gamma \in \Sigma(Y)$ that

$$\|U_{\tau_\gamma}(f)\|_{X|Y}^2 \mathbb{1}_B \leq \|U_{\tau_\gamma}(f) - g_i\|_{X|Y}^2 \mathbb{1}_B + 2\varepsilon \mathbb{1}_B.$$

for all $1 \leq i \leq n$. Taking the infimum over $1 \leq i \leq n$, the right-hand side of the last inequality is smaller than $(\varepsilon^2 + 2\varepsilon)\mathbb{1}_B$ due to (11.4). For $C \coloneqq \sigma_\gamma(B) \in \Sigma(Y)$ we thus have

$$U_{\sigma_\gamma}(\|f\|_{X|Y}^2 \mathbb{1}_C) = \|U_{\tau_\gamma}(f)\|_{X|Y}^2 \mathbb{1}_B \leq (\varepsilon^2 + 2\varepsilon)\mathbb{1}_B = U_{\sigma_\gamma}((\varepsilon^2 + 2\varepsilon)\mathbb{1}_C),$$

which implies $\|f\|_{X|Y}^2 \mathbb{1}_C \leq (\varepsilon^2 + 2\varepsilon)\mathbb{1}$. Since $\mu_Y(C) = \mu_Y(B) \geq 1 - \delta$ and $\varepsilon > 0$ as well as $\delta > 0$ can be chosen arbitrarily small, we conclude that $f = 0$. □

By Lemma 11.3.8, to establish Theorem 11.3.3, it suffices to show that the span of the union of invariant finite rank $\mathrm{L}^\infty(Y)$-submodules in the range of any conditional kernel operator $K \in \mathrm{fix}(U_{\tau \times \tau}) \cap \mathrm{L}^2(X \otimes_Y X \mid Y)$ is $\mathrm{L}^2(X)$-dense in the range of K. This will be achieved through a conditional spectral analysis of conditional Hilbert–Schmidt operators in the next lecture.

11.4 Comments and Further Reading

Our proofs of the preservation of the multiple recurrence property under weakly mixing extensions and under taking inverse limits are based on Tao's treatment in [Tao09]. For the original argument by Furstenberg, see [Fur77], or refer to his comprehensive textbook treatment in [Fur14].

In the "Comments and Further Reading" section of the next lecture, we will provide an extended discussion of the background and literature regarding the classification of compact extensions.

11.5 Exercises

Exercise 11.1. Deduce Theorem 11.1.1 from Proposition 11.1.2.

Exercise 11.2. Prove Lemma 11.3.2.

Exercise 11.3. Let $q\colon (X,\tau) \to (Y,\sigma)$ be an extension of concrete measure-preserving systems, and suppose that the union of invariant finite rank $\mathrm{L}^\infty(Y)$-submodules of $\mathrm{L}^2(X \mid Y)$ is dense in $\mathrm{L}^2(X)$. Show that $q\colon (X,\tau) \to (Y,\sigma)$ is a compact extension in the sense of Definition 9.1.5.

Exercise 11.4. Let $q_1\colon (X_1,\tau_1) \to (Y,\sigma)$ and $q_2\colon (X_2,\tau_2) \to (Y,\sigma)$ be compact extensions of concrete measure-preserving systems. Show that $p\colon (X_1 \otimes_Y X_2, \tau_1 \times \tau_2) \to (Y,\sigma)$ is a compact extension.

Exercise 11.5. Show that an extension $q\colon (X,\tau) \to (Y,\sigma)$ of concrete measure-preserving systems is weakly mixing if and only if the extension $p\colon (X \otimes_Y X, \tau \times \tau) \to (Y,\sigma)$ of concrete measure-preserving systems is ergodic in the sense that $\mathrm{fix}(U_{\tau\times\tau}) = U_p(\mathrm{fix}(U_\sigma))$. In particular, if (Y,σ) is an ergodic system, then an extension $q\colon (X,\tau) \to (Y,\sigma)$ is weakly mixing if and only if the system $(X \otimes_Y X, \tau \times \tau)$ is ergodic.

Lecture 12

In this lecture, we complete the proof of Theorem 11.3.3 through a spectral analysis of conditional Hilbert–Schmidt operators. Additionally, we establish a geometric classification of compact extensions due to Mackey and Zimmer, extending the Halmos–von Neumann representation theorem (Theorem 6.2.6) from discrete spectrum systems to compact extensions of measure-preserving systems.

12.1 Classification of Compact Extensions, Part II

To finish the proof of Theorem 11.3.3, we first examine finite rank $\mathrm{L}^\infty(Y)$-submodules in greater detail through the following lemmas, starting with a basic observation.

Lemma 12.1.1. *If $\mathcal{H} \subseteq \mathrm{L}^2(X \mid Y)$ is a finite rank $\mathrm{L}^\infty(Y)$-submodule, then $\mathcal{H}$ is closed in $\mathrm{L}^2(X \mid Y)$ with respect to the $\mathrm{L}^2(X)$-norm.*

Proof. Assume that $\mathcal{H}$ is generated by the finite conditionally orthonormal subset $M = \{f_1, \dots, f_n\} \subseteq \mathrm{L}^2(X \mid Y)$. Let $(g_m)_{m\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ converging to $g \in \mathrm{L}^2(X \mid Y)$ with respect to the $\mathrm{L}^2(X)$-norm. By passing to a subsequence, we may assume that the convergence holds almost surely and that there exists $h \in \mathrm{L}^2(X)$ such that $|g_m| \le h$ for all $m \in \mathbb{N}$.

By Lemma 11.3.2, we can write $g_m = \sum_{i=1}^n (g_m \mid f_i)_{X|Y} f_i$ for every $m \in \mathbb{N}$. Using the conditional Cauchy–Schwarz inequality, we obtain

$$|(g_m \mid f_i)_{X|Y} - (g \mid f_i)_{X|Y}|^2 \le \|g_m - g\|_{X|Y}^2 = \mathbb{E}(|g_m - g|^2 \mid Y),$$

for each $m \in \mathbb{N}$ and $i \in \{1, \dots, n\}$.

Since $(g_m)_{m\in\mathbb{N}}$ converges to g almost surely and $|g_m| \le h$ for all $m \in \mathbb{N}$ for some $h \in \mathrm{L}^2(X)$, we conclude that $\mathbb{E}(|g_m - g|^2 \mid Y) \to 0$ almost surely as $m \to \infty$ by Lemma 8.1.1. Therefore, $(g_m \mid f_i)_{X|Y} \to (g \mid f_i)_{X|Y}$ almost surely for each i.

Substituting back, we see that

$$g = \lim_{m\to\infty} g_m = \lim_{m\to\infty} \sum_{i=1}^{n} (g_m \mid f_i)_{X|Y} f_i = \sum_{i=1}^{n} (g \mid f_i)_{X|Y} f_i,$$

which implies $g \in \mathcal{H}$. Thus, $\mathcal{H}$ is closed in $\mathrm{L}^2(X \mid Y)$ with respect to the $\mathrm{L}^2(X)$-norm. □

We now aim to identify a condition that provides a converse to Lemma 12.1.1: Under what circumstances is an $\mathrm{L}^\infty(Y)$-submodule $\mathcal{H} \subseteq \mathrm{L}^2(X \mid Y)$, which is closed in $\mathrm{L}^2(X \mid Y)$ with respect to the $\mathrm{L}^2(X)$-norm, of finite rank?

Before we answer this question, we need to “normalize” elements with respect to their conditional norm.

Proposition and Definition 12.1.2. *Let $f \in \mathrm{L}^2(X \mid Y)$. Then the equivalence class $f/\|f\|_{X|Y}$ of measurable functions defined by*

$$(f/\|f\|_{X|Y})(x) := \begin{cases} \frac{f(x)}{\|f\|_{X|Y}(q(x))} & \text{if } x \in q^{-1}(\{\|f\|_{X|Y} \neq 0\}) \\ 0 & \text{else} \end{cases}$$

*is an element of $\mathrm{L}^2(X \mid Y)$, and called the **conditional normalization** of f. Moreover,*

(i) $f/\|f\|_{X|Y} = \lim_{n\to\infty} \frac{1}{\|f\|_{X|Y} + \frac{1}{n}\mathbb{1}} \cdot f$ *in $\mathrm{L}^2(X)$-norm,*

(ii) $\|f/\|f\|_{X|Y}\|_{X|Y} = \mathbb{1}_{\{\|f\|_{X|Y} \neq 0\}}$, *and*

(iii) $\|f\|_{X|Y} \cdot f/\|f\|_{X|Y} = f$.

Proof. Set $f_n := \frac{1}{\|f\|_{X|Y} + \frac{1}{n}\mathbb{1}} \cdot f$ and $g_n := \|f_n\|_{X|Y} = \frac{\|f\|_{X|Y}}{\|f\|_{X|Y} + \frac{1}{n}\mathbb{1}}$ for $n \in \mathbb{N}$. Clearly, $\lim_{n\to\infty} g_n = \mathbb{1}_{\{\|f\|_{X|Y} \neq 0\}}$ almost surely and then also in $\mathrm{L}^2(Y)$. A short computation reveals that

$$\|f_n - f_m\|_{\mathrm{L}^2(X)}^2 = \int_Y \|f_n - f_m\|_{X|Y}^2 = \int_Y |g_n - g_m|^2 = \|g_n - g_m\|_{\mathrm{L}^2(Y)}^2$$

for $n, m \in \mathbb{N}$. Thus, $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathrm{L}^2(X)$, hence converges to some $h \in \mathrm{L}^2(X)$. Since $f = \mathbb{1}_{\{\|f\|_{X|Y} \neq 0\}} f$, we obtain that

$$f/\|f\|_{X|Y} = \lim_{n\to\infty} \frac{1}{\|f\|_{X|Y} + \frac{1}{n}\mathbb{1}} \mathbb{1}_{\{\|f\|_{X|Y} \neq 0\}} \cdot f = \lim_{n\to\infty} \frac{1}{\|f\|_{X|Y} + \frac{1}{n}\mathbb{1}} \cdot f$$

almost surely. This implies $h = f/\|f\|_{X|Y}$ by passing to an almost surely convergent subsequence of $(f_n)_{n\in\mathbb{N}}$. Moreover, one can check that $\|h\|_{X|Y} = \lim_{n\to\infty} \|f_n\|_{X|Y}$ in

$\mathrm{L}^2(Y)$, hence $\|h\|_{X|Y} = \mathbb{1}_{\{\|f\|_{X|Y}\neq 0\}}$, in particular $h \in \mathrm{L}^2(X \mid Y)$. Finally, we obtain that

$$\|f\|_{X|Y} \cdot f/\|f\|_{X|Y} = \lim_{n\to\infty} \frac{\|f\|_{X|Y}}{\|f\|_{X|Y} + \frac{1}{n}\mathbb{1}} \cdot f = \mathbb{1}_{\{\|f\|_{X|Y}\neq 0\}} f.$$

□

The following lemma now ensures the existence of elements $f \in \mathcal{H}$ in such submodules that maximize the support of the conditional norm $\|f\|_{X|Y}$ in Y.

Lemma 12.1.3. *Assume that $\mathcal{H} \subseteq \mathrm{L}^2(X \mid Y)$ is an $\mathrm{L}^\infty(Y)$-submodule that is closed in $\mathrm{L}^2(X \mid Y)$ with respect to the $\mathrm{L}^2(X)$-norm. Then there exist $B^* \in \Sigma(Y)$ and $f \in \mathcal{H}$ such that $\|f\|_{X|Y} = \mathbb{1}_{B^*}$ and $\mathbb{1}_{B^*} g = g$ for every $g \in \mathcal{H}$. In particular, for every $g \in \mathcal{H}$, we have $\mu_Y(\{\|g\|_{X|Y} \neq 0\} \setminus B^*) = 0$.*

Proof. Define

$$\mathcal{B} = \{B \in \Sigma(Y) \colon \exists f \in \mathcal{H} \text{ such that } \|f\|_{X|Y} = \mathbb{1}_B\}.$$

By the completeness and countable chain condition of the measure algebra $\Sigma(Y)$ (see Exercise 1.6(vi)), the set $B^* = \bigcup \mathcal{B}$ exists in $\Sigma(Y)$, and there is a countable family $(\tilde{B}_n)$ in $\mathcal{B}$ such that $B^* = \bigcup_n \tilde{B}_n$. Define $B_1 = \tilde{B}_1$ and $B_n = \tilde{B}_n \setminus \bigcup_{i<n} \tilde{B}_i$ for all $n \geq 2$. Then the family (B_n) forms a partition of B^*, and one can directly verify that $B_n \in \mathcal{B}$ for all n. Let $f_n \in \mathcal{H}$ satisfy $\|f_n\|_{X|Y} = \mathbb{1}_{B_n}$.

Form $f = \sum_n f_n \mathbb{1}_{B_n} = \lim_{N\to\infty} \sum_{n\leq N} f_n \mathbb{1}_{B_n}$ almost surely. By Lemma 8.1.1, we have

$$\|f\|_{X|Y} = \left\| \sum_n f_n \mathbb{1}_{B_n} \right\|_{X|Y} = \sum_n \|f_n\|_{X|Y} \mathbb{1}_{B_n} = \mathbb{1}_{B^*}.$$

Thus, $B^* \in \mathcal{B}$ is attained. The claim follows from the maximality of B^* and using normalization of elements $f \in \mathcal{H}$. □

We now obtain the following partial converse to Lemma 12.1.1.

Lemma 12.1.4. *Let $\mathcal{H} \subseteq \mathrm{L}^2(X \mid Y)$ be an $\mathrm{L}^\infty(Y)$-submodule satisfying the following two properties:*

(i) *$\mathcal{H}$ is closed in $\mathrm{L}^2(X \mid Y)$ with respect to the $\mathrm{L}^2(X)$-norm.*

(ii) *There exists a constant $C > 0$ such that $\sum_{f\in M} \|f\|^2_{X|Y} \leq C\mathbb{1}$ for every conditionally orthonormal subset $M \subseteq \mathcal{H}$.*

Then $\mathcal{H}$ is a finite rank $\mathrm{L}^\infty(Y)$-submodule.

Proof. Let n be the largest integer such that $n \leq C$. We recursively construct $f_1, \ldots, f_{n+1} \in \mathcal{H}$ and $B_1, \ldots, B_{n+1} \in \Sigma(Y)$ as follows: Assume for some $m \in$

$\{1, \ldots, n+1\}$ that $f_1, \ldots, f_{m-1}$ and $B_1, \ldots, B_{m-1} \in \Sigma(Y)$ have already been constructed. Then, define

$$\mathcal{H}_m := \{f \in \mathcal{H} \colon (f \mid f_i)_{X|Y} = 0 \text{ for every } i \in \{1, \ldots, m-1\}\}.$$

This is a closed $\mathrm{L}^\infty(Y)$-submodule of $\mathrm{L}^2(X \mid Y)$ with respect to the $\mathrm{L}^2(X)$-norm. By Lemma 12.1.3, we can pick $f_m \in \mathcal{H}_m$ and $B_m \in \Sigma(Y)$ such that $\|f_m\|_{X|Y} = \mathbb{1}_{B_m}$ and $\mathbb{1}_{B_m} g = g$ for every $g \in \mathcal{H}_m$.

By construction, $\{f_1, \ldots, f_{n+1}\}$ is a conditionally orthonormal subset with $\|f_j\|_{X|Y} = \mathbb{1}_{B_j} \le \mathbb{1}_{B_i} = \|f_i\|_{X|Y}$ for $i \le j$. Since

$$\sum_{i=1}^{n+1} \|f_i\|_{X|Y}^2 \le C < n+1,$$

we must have $\|f_{n+1}\|_{X|Y} = 0$, hence $f_{n+1} = 0$. By the choice of f_{n+1}, it follows that $f = 0$ for every $f \in \mathcal{H}$ that is conditionally orthogonal to $f_1, \ldots, f_n$.

Now let $\mathcal{K}$ be the finite rank $\mathrm{L}^\infty(Y)$-submodule generated by $\{f_1, \ldots, f_n\}$. Clearly, $\mathcal{K} \subseteq \mathcal{H}$. To finish the argument, consider the closure $\overline{\mathcal{H}}$ in $\mathrm{L}^2(X)$. We claim that $\mathcal{K}$ is dense in $\overline{\mathcal{H}}$, i.e., $\overline{\mathcal{K}} = \overline{\mathcal{H}}$. Then, by Lemma 12.1.1,

$$\mathcal{K} = \overline{\mathcal{K}} \cap \mathrm{L}^2(X \mid Y) = \overline{\mathcal{H}} \cap \mathrm{L}^2(X \mid Y) = \mathcal{H}$$

as desired.

To check the claim take $g \in \overline{\mathcal{H}}$ that is orthogonal in the classical sense to $\mathcal{K}$ in the Hilbert space $\mathrm{L}^2(X)$ and show $g = 0$. Since $\overline{\mathcal{H}}$ is a $\mathrm{L}^\infty(Y)$-submodule of $\mathrm{L}^2(X)$, we obtain that $g_n := \mathbb{1}_{\{\|g\|_{X|Y} \le n\}} g \in \overline{\mathcal{H}} \cap \mathrm{L}^2(X \mid Y) = \mathcal{H}$ for each $n \in \mathbb{N}$. If $f \in \mathcal{K}$, then for each $h \in \mathrm{L}^\infty(Y)$, we have

$$\int_Y h (f \mid g_n)_{X|Y} = (h \mathbb{1}_{[\|g\|_{X|Y} \le n]} f \mid g)_{\mathrm{L}^2(X)} = 0,$$

since $\mathcal{K}$ is an $\mathrm{L}^\infty(Y)$-submodule. This implies $(f \mid g_n)_{X|Y} = 0$ in $\mathrm{L}^\infty(Y)$, i.e., g_n is conditionally orthogonal to $\mathcal{K}$ for each $n \in \mathbb{N}$. Thus, $g_n = 0$ for every $n \in \mathbb{N}$, and since $g = \lim_{n \to \infty} g_n$ in $\mathrm{L}^2(X)$, also $g = 0$. □

Finally, we also need the following observation, which allows to turn finitely generated submodules into finite rank submodules.

Lemma 12.1.5. *Let $\mathcal{H} \subseteq \mathrm{L}^2(X \mid Y)$ be an $\mathrm{L}^\infty(Y)$-submodule which is finitely generated, that is, there exist $f_1, \ldots, f_n \in \mathcal{H}$ such that each $f \in \mathcal{H}$ admits a representation $f = \sum_{i=1}^n g_i f_i$ for some $g_1, \ldots, g_n \in \mathrm{L}^\infty(Y)$. If $\overline{\mathcal{H}}$ denotes the closure in $\mathrm{L}^2(X)$, then $\overline{\mathcal{H}} \cap \mathrm{L}^2(X \mid Y)$ is a finite rank $\mathrm{L}^\infty(Y)$-submodule.*

Proof. Note first that if $f \in \mathcal{H}$, then $f/\|f\|_{X|Y} \in \overline{\mathcal{H}} \cap L^2(X \mid Y)$ by Proposition 12.1.2 (i). We now use the Gram–Schmidt process to recursively define

$$e_i' := f_i - \sum_{j=1}^{i-1} (f_i|e_j)_{X|Y} e_j \in \overline{\mathcal{H}} \cap L^2(X \mid Y) \text{ and}$$
$$e_i := e_i'/\|e_i'\|_{X|Y} \in \overline{\mathcal{H}} \cap L^2(X \mid Y)$$

for every $i \in \{1, \dots, n\}$. By Proposition 12.1.2 (ii) this gives us a conditionally orthonormal set $\{e_1, \dots, e_n\} \subseteq \overline{\mathcal{H}} \cap L^2(X \mid Y)$. Let $\mathcal{K}$ be the generated finite rank $L^\infty(Y)$-submodule. Then $\mathcal{K} \subseteq \overline{\mathcal{H}} \cap L^2(X \mid Y)$. On the other hand, by Proposition 12.1.2 (iii) we have

$$f_i = \sum_{j=1}^{i-1} (f_i|e_j)_{X|Y} e_j + \|e_i'\|_{X|Y} e_i \in \mathcal{K}$$

for each $i \in \{1, \dots, n\}$. Therefore, $\mathcal{H} \subseteq \mathcal{K}$, and by Lemma 12.1.1 this implies $\overline{\mathcal{H}} \cap L^2(X \mid Y) \subseteq \mathcal{K}$. □

We will need the following corollary in a dynamical situation.

Corollary 12.1.6. *Let $p\colon (X, \tau) \to (Y, \sigma)$ be an extension of concrete measure-preserving systems over Γ, with (Y, σ) ergodic. Suppose that $\mathcal{H}$ is a Γ-invariant, finitely generated $L^\infty(Y)$-submodule of $L^2(X \mid Y)$, closed with respect to the $L^2(X)$-norm, and that $\mathcal{H} \neq \{0\}$.*

Then $\mathcal{H}$ admits a conditionally orthonormal basis M, in the sense that $(f \mid g)_{X|Y} = 0$ for all $f, g \in M$ with $f \neq g$, and $(f \mid f)_{X|Y} = \mathbb{1}_Y$ for all $f \in M$. In particular, $U_\tau^\gamma(M)$ is also a conditionally orthonormal basis of $\mathcal{H}$ for all $\gamma \in \Gamma$.

Proof. Let B^* be as in the proof of Lemma 12.1.3. For $\gamma \in \Gamma$ and $g \in \mathcal{H}$, we have

$$\mathbb{1}_{B^*} U_\tau^\gamma(\mathbb{1}_{B^*} g) = \mathbb{1}_{B^*} \mathbb{1}_{\tau_{\gamma^{-1}}(B^*)} U_\tau^\gamma(g) = \mathbb{1}_{\tau_{\gamma^{-1}}(B^*)} U_\tau^\gamma(g).$$

This implies that $\mathbb{1}_{\tau_{\gamma^{-1}}(B^*)} g = g$ for all $g \in \mathcal{H}$. By the maximality of B^*, this implies that $\mu_Y(\tau_{\gamma^{-1}}(B^*) \Delta B^*) = 0$ for all $\gamma \in \Gamma$. By ergodicity and since B^* must have positive measure because $\mathcal{H} \neq \{0\}$, $\mu_Y(B^*) = 1$. All the claims can now be deduced from the proofs of Lemma 12.1.4 and Lemma 12.1.5. □

Lemma 12.1.5 has the following important consequence.

Lemma 12.1.7. *If $\mathcal{H}_1, \mathcal{H}_2 \subseteq L^2(X \mid Y)$ are invariant finite rank $L^\infty(Y)$-submodules, then there is an invariant finite rank submodule $\mathcal{H} \subseteq L^2(X \mid Y)$ with $\mathcal{H}_1 + \mathcal{H}_2 \subseteq \mathcal{H}$.*

Proof. Apply Lemma 12.1.5 to the invariant and finitely generated $L^\infty(Y)$-submodule $\mathcal{H}_1 + \mathcal{H}_2 \subseteq L^2(X \mid Y)$. □

We now finish the proof of Theorem 11.3.3:

Proof of Theorem 11.3.3. In view of Lemma 12.1.7 and Proposition 11.3.8 it suffices to show that the ranges of $K*_Y$ where $K \in \mathrm{fix}(U_{\tau\times\tau}) \cap \mathrm{L}^2(X \otimes_Y X \mid Y)$ are contained in the closed linear hull of the union of all invariant finite rank $\mathrm{L}^\infty(Y)$-submodules of $\mathrm{L}^2(X \mid Y)$.

Let $K \in \mathrm{fix}(U_{\tau\times\tau}) \cap \mathrm{L}^2(X \otimes_Y X \mid Y)$. By decomposing

$$K(x,y) = \frac{K(x,y) + \overline{K(y,x)}}{2} + \mathrm{i}\frac{K(x,y) - \overline{K(y,x)}}{2\mathrm{i}} \text{ for } (x,y) \in X \otimes_Y X,$$

we may reduce to the case that that $K(x,y) = \overline{K(y,x)}$ for almost all $(x,y) \in X \otimes_Y X$.

By Proposition 11.3.4, $K*_Y \colon \mathrm{L}^2(X) \to \mathrm{L}^2(X)$ is a bounded operator and, since $K(x,y) = \overline{K(y,x)}$ for almost all $(x,y) \in X \otimes_Y X$, it is self-adjoint. For $\varepsilon > 0$, consider the spectral projections

$$P_\varepsilon^+ := \mathbb{1}_{[\varepsilon, \|K*_Y\|]}(K*_Y) \text{ and } P_\varepsilon^- := \mathbb{1}_{[-\|K*_Y\|, -\varepsilon]}(K*_Y).$$

For a collection of facts about spectral projections, see Appendix A.3. In particular, we have

$$P_\varepsilon^+ \circ (K*_Y) = (K*_Y) \circ P_\varepsilon^+ \geq \varepsilon P_\varepsilon^+, \tag{12.1}$$
$$P_\varepsilon^- \circ (K*_Y) = (K*_Y) \circ P_\varepsilon^- \leq -\varepsilon P_\varepsilon^-, \tag{12.2}$$

where the inequalities mean that

$$(K *_Y P_\varepsilon^+ f \mid P_\varepsilon^+ f)_{\mathrm{L}^2(X)} \geq \varepsilon (P_\varepsilon^+ f \mid P_\varepsilon^+ f)_{\mathrm{L}^2(X)} \text{ and}$$
$$(K *_Y P_\varepsilon^- f \mid P_\varepsilon^- f)_{\mathrm{L}^2(X)} \leq -\varepsilon (P_\varepsilon^- f \mid P_\varepsilon^- f)_{\mathrm{L}^2(X)}$$

for all $f \in \mathrm{L}^2(X)$.

Since $P_\varepsilon^\pm$ arises as a limit of polynomials in $K*_Y$ in the strong operator topology, $P_\varepsilon^\pm$ is Γ-equivariant, i.e., $U_{\tau_\gamma} P_\varepsilon^\pm = P_\varepsilon^\pm U_{\tau_\gamma}$ for every $\gamma \in \Gamma$. Additionally, $P_\varepsilon^\pm$ is $\mathrm{L}^\infty(Y)$-linear since $\mathrm{L}^\infty(Y)$-linearity is preserved when passing to strong operator limits. We further show that for $f \in \mathrm{L}^2(X)$ we have $\|P_\varepsilon^\pm f\|_{X|Y} \leq \|f\|_{X|Y}$.

For

$$E := \{y \in Y \colon \|P_\varepsilon^\pm f\|_{X|Y}(y) > \|f\|_{X|Y}(y)\} \in \Sigma(Y)$$

consider $g := (U_q \mathbb{1}_E) \cdot f \in \mathrm{L}^2(X)$. Since $P_\varepsilon^\pm$ is an orthogonal projection, we obtain $\|P_\varepsilon^\pm g\|_{\mathrm{L}^2(X)}^2 \leq \|g\|_{\mathrm{L}^2(X)}^2$. On the other hand, since $P_\varepsilon^\pm$ is $\mathrm{L}^\infty(Y)$-linear, we obtain

$$\begin{aligned} 0 \leq \|g\|_{\mathrm{L}^2(X)}^2 - \|P_\varepsilon^\pm g\|_{\mathrm{L}^2(X)}^2 &= \int_Y \|g\|_{X|Y}^2 - \|P_\varepsilon^\pm g\|_{X|Y}^2 \,\mathrm{d}\mu_Y \\ &= \int_Y \mathbb{1}_E (\|f\|_{X|Y}^2 - \|P_\varepsilon^\pm f\|_{X|Y}^2) \,\mathrm{d}\mu_Y \leq 0. \end{aligned}$$

This implies $\mathbb{1}_E(\|f\|^2_{X|Y} - \|P^{\pm}_\varepsilon f\|^2_{X|Y}) = 0$, and consequently $\mu_Y(E) = 0$. Thus, $\|P^{\pm}_\varepsilon f\|_{X|Y} \leq \|f\|_{X|Y}$ as desired.

We claim that the images $\mathcal{H}^{\pm}_\varepsilon := P^{\pm}_\varepsilon(\mathrm{L}^2(X \mid Y))$ are invariant finite rank $\mathrm{L}^\infty(Y)$-submodules of $\mathrm{L}^2(X \mid Y)$. By what we have just shown, we indeed have that $\mathcal{H}^{\pm}_\varepsilon \subseteq \mathrm{L}^2(X \mid Y)$. Since $P^{\pm}_\varepsilon$ is $\mathrm{L}^\infty(Y)$-linear and equivariant, it follows that $\mathcal{H}^{\pm}_\varepsilon$ is an invariant $\mathrm{L}^\infty(Y)$-submodule of $\mathrm{L}^2(X \mid Y)$.

We now show that it has finite rank. By Properties (12.1) and (12.2), we obtain

$$|(K *_Y f \mid f)_{\mathrm{L}^2(X)}| \geq \varepsilon (f \mid f)_{\mathrm{L}^2(X)}$$

for all $f \in \mathcal{H}^{\pm}_\varepsilon$. Using a similar argument as above, we even obtain

$$|(K *_Y f \mid f)_{X|Y}| \geq \varepsilon (f \mid f)_{X|Y}$$

for each $f \in \mathcal{H}^{\pm}_\varepsilon$. By the conditional Cauchy–Schwarz inequality we thus have

$$\|K *_Y f\|_{X|Y} \cdot \|f\|_{X|Y} \geq \varepsilon \|f\|^2_{X|Y}$$

and this implies $\|f\|_{X|Y} \leq \frac{1}{\varepsilon}\|K *_Y f\|_{X|Y}$ for $f \in \mathcal{H}^{\pm}_\varepsilon$. Using that $K *_Y \colon \mathrm{L}^2(X \mid Y) \to \mathrm{L}^2(X \mid Y)$ is a conditional Hilbert–Schmidt operator (see Proposition 11.3.7), we find $C > 0$ such that

$$\sum_{f \in M} \|K *_Y f\|^2_{X|Y} \leq C$$

for every conditionally orthonormal subset $M \subseteq \mathrm{L}^2(X \mid Y)$. In particular, if $M \subseteq \mathcal{H}^{\pm}_\varepsilon$ is a conditionally orthonormal subset of $\mathcal{H}^{\pm}_\varepsilon$, then

$$\sum_{f \in M} \|f\|^2_{X|Y} \leq \frac{1}{\varepsilon} \sum_{f \in M} \|K *_Y f\|^2_{X|Y} \leq \frac{C}{\varepsilon}.$$

Applying Lemma 12.1.4 we obtain that the $\mathrm{L}^\infty(Y)$-submodules $\mathcal{H}^{\pm}_\varepsilon$ are of finite rank.

To conclude the proof notice that, since $K *_Y \colon \mathrm{L}^2(X) \to \mathrm{L}^2(X)$ is a bounded operator and $\mathrm{L}^2(X \mid Y)$ is dense in $\mathrm{L}^2(X)$, the image $K *_Y (\mathrm{L}^2(X \mid Y))$ is dense in $K *_Y (\mathrm{L}^2(X))$. Now if $f \in \mathrm{L}^2(X \mid Y)$ we obtain by the properties of spectral projections that

$$K *_Y f = \lim_{n \to \infty} (P^{+}_{\frac{1}{n}}(K *_Y f) + P^{-}_{\frac{1}{n}}(K *_Y f)).$$

Therefore, the image of $K *_Y$ is contained in the $\mathrm{L}^2(X)$-closure of the the union of all invariant finite rank $\mathrm{L}^\infty(Y)$-submodules. In view of Proposition 11.3.8, this concludes the proof of the theorem. □

12.2 Classification of Compact Extensions, Part III

In this section, we establish an analogue of the Halmos–von Neumann representation theorem (Theorem 6.2.6) for compact extensions. To this end, we introduce the following definition of measurable cocycles (cf. the definition of continuous cocycles in Definition 7.2.4).

Definition 12.2.1. Let (Y, σ) be a concrete measure-preserving system, and let $K = (K, \cdot)$ be a compact metrizable group (not necessarily abelian), equipped with the Borel σ-algebra. A measurable[1] function $\rho\colon \Gamma \times Y \to K$ is said to be a **cocycle** if the identity

$$\rho(\gamma_1 + \gamma_2, y) = \rho(\gamma_1, \sigma_{\gamma_2}(y)) \cdot \rho(\gamma_2, y) = \rho(\gamma_2, \sigma_{\gamma_1}(y)) \cdot \rho(\gamma_1, y)$$

holds for almost every $y \in Y$ and every pair $\gamma_1, \gamma_2 \in \Gamma$.[2]

Two cocycles ρ_1 and ρ_2 are called **cohomologous** if there exists a measurable map $U\colon Y \to K$ such that

$$\rho_1(\gamma, y) = U(\sigma_\gamma(y)) \cdot \rho_2(\gamma, y) \cdot U(y)^{-1}$$

for almost every y and every $\gamma \in \Gamma$.

Proposition and Definition 12.2.2. *Let (Y, σ) be a concrete measure-preserving system, let K be a compact metrizable group, let $L \subseteq K$ be a closed subgroup, and let $\rho\colon \Gamma \times Y \to K$ be a cocycle. We define the* **homogeneous skew-product extension** $Y \rtimes_\rho K/L = (X, \tau)$ *by the following data:*

(i) The product set $X := Y \times K/L$;

(ii) The product σ-algebra $\Sigma_Y \otimes \Sigma_{K/L}$, where $\Sigma_{K/L}$ is the Borel σ-algebra on K/L;

(iii) The product probability measure $\mu_Y \otimes \mathrm{m}_{K/L}$, where $\mathrm{m}_{K/L}$ is the Haar measure on K/L (see Definition 6.1.3 and Remark 6.1.4);

(iv) The action $\tau_\gamma(y, kL) = (\sigma_\gamma(y), (\rho(\gamma, y) \cdot k)L)$, defined for every $\gamma \in \Gamma$, almost every $y \in Y$, and every $kL \in K/L$.

Then (X, τ) is a concrete measure-preserving system, and the Y-coordinate projection defines a factor map $p\colon (X, \tau) \to (Y, \sigma)$.

A homogeneous skew-product system is called a **group skew-product extension** *if L is the trivial subgroup of K.*

[1] Since Γ is a countable discrete (abelian) group, it is naturally equipped with the power set algebra. Thus, the measurability requirement is equivalent to the condition that the map $\rho(\gamma, \cdot)$ is measurable from Y to K for each $\gamma \in \Gamma$.

[2] Since Γ is assumed to be a countable group, the order of the quantifiers for y and γ_1, γ_2 does not matter.

Moreover, suppose that ρ_1, ρ_2 are cohomologous cocycles, then the respective homogeneous skew-product extensions $Y \rtimes_{\rho_1} K/L$ and $Y \rtimes_{\rho_2} K/L$ are isomorphic extensions in the sense that there is an isomorphism $q\colon Y \rtimes_{\rho_1} K/L \to Y \rtimes_{\rho_2} K/L$ in the sense of Definition 1.1.5 such that $p_2 \circ q = p_1$ where $p_i\colon Y \rtimes_{\rho_i} K/L \to (Y, \sigma)$ are the Y-coordinate projections for $i = 1, 2$.

Proof. Exercise. □

In Exercise 9.5, we showed that in the case where $\Gamma = \mathbb{Z}$, a homogeneous skew-product extension is a compact extension. The same proof generalizes to show that this conclusion holds for any homogeneous skew-product extension with respect to an arbitrary countable abelian group Γ. In this section, we prove the converse statement for general countable abelian groups under the additional assumption of ergodicity:

Theorem 12.2.3 (Mackey–Zimmer). *Let $p\colon (X, \tau) \to (Y, \sigma)$ be a compact extension of concrete measure-preserving systems such that (X, τ) (and therefore also (Y, σ)) is ergodic. Then there exist a compact metrizable group K, a closed subgroup $L \subseteq K$, and a cocycle $\rho\colon \Gamma \times Y \to K$ such that the homogeneous skew-product $Y \rtimes_\rho K/L$ and (X, τ) are isomorphic extensions of (Y, σ).*

The proof of Theorem 12.2.3 relies on the characterization of compact extensions through invariant finite rank $\mathrm{L}^\infty(Y)$-submodules in Theorem 11.3.3 and a technical, non-trivial result, commonly referred to as the Mackey–Zimmer theorem as well. We now pause to state this result before returning to the proof of Theorem 12.2.3.

Definition 12.2.4. An extension $p\colon (X, \tau) \to (Y, \sigma)$ of concrete measure-preserving systems is called **homogeneous** if there are a compact metrizable group K, a closed subgroup $L \subseteq K$, a cocycle $\rho\colon \Gamma \times Y \to K$, and a measurable map $\theta\colon X \to K/L$ such that:

(i) $X = Y \times K/L$ is equipped with the product σ-algebra $\Sigma_Y \otimes \Sigma_{K/L}$;

(ii) $(\theta \circ \tau_\gamma)(x) = (\rho_\gamma \circ p)(x) \cdot \theta(x)$ for almost every $x \in X$ and all $\gamma \in \Gamma$.

A homogeneous extension is called a **group extension** if L is the trivial subgroup of K.

By possibly changing the homogeneous space K/L and considering a cohomologous cocycle, a homogeneous extension of ergodic systems is isomorphic to a homogeneous skew-product extension:

Lemma 12.2.5 (Mackey–Zimmer). *Let $p\colon (X, \tau) \to (Y, \sigma)$ be a homogeneous extension by the data $(K/L, \rho, \theta)$ of ergodic concrete measure-preserving systems. Then there exist a closed subgroup $H \subseteq K$, a closed subgroup $M \subseteq H$, and a cocycle $\tilde{\rho}\colon \Gamma \times Y \to H$ that is cohomologous to ρ (when viewed as a K-valued cocycle), such* *that* $p\colon (X,\tau) \to (Y,\sigma)$ *and* $\tilde{p}\colon Y \rtimes_{\tilde{\rho}} H/M \to (Y,\sigma)$ *are isomorphic extensions.*

A proof of Lemma 12.2.5 can be found in the next section at the end of this lecture. Here we prove Theorem 12.2.3 with it:

Proof. From Theorem 11.3.3 we know that the invariant finite rank $\mathrm{L}^\infty(Y)$-submodules of $\mathrm{L}^2(X \mid Y)$ are dense in $\mathrm{L}^2(X)$.

Let $\mathcal{M} \neq \{0\}$ be such an invariant finite rank $\mathrm{L}^\infty(Y)$-submodule of $\mathrm{L}^2(X \mid Y)$, generated by a conditionally orthonormal basis $\{f_1, \dots, f_n\}$ as in Corollary 12.1.6.

For every $\gamma \in \Gamma$ and $1 \leq i \leq n$, we have for almost every x,

$$U_{\tau_\gamma}(f_i)(x) = \sum_{j=1}^{n} \lambda_{i,j}(\gamma, q(x)) f_j(x),$$

for some suitable $\lambda_{i,j}(\gamma, \cdot) \in \mathrm{L}^\infty(Y)$. Define the following:

$$\theta(x) \coloneqq \frac{1}{\sqrt{n}}(f_1(x), \dots, f_n(x)) \quad \text{for almost every } x,$$
$$\Lambda(\gamma, y) \coloneqq (\lambda_{i,j}(\gamma, y))_{1 \leq i,j \leq n} \quad \text{for almost every } y.$$

Then θ can be viewed as an element of the measurable maps from X to the sphere $S^{2n-1} = \{z \in \mathbb{C}^n : \|z\|_2 = 1\}$ (modulo almost everywhere equality in x), which we identify[3] with the homogeneous space $\mathrm{U}(n)/\mathrm{U}(n-1)$, where $\mathrm{U}(n)$ is the group of unitary $n \times n$ matrices. For almost every y and each $\gamma \in \Gamma$, the matrix $\Lambda(\gamma, y)$ represents an orthonormal basis change in $\mathbb{C}^n$ (by Corollary 12.1.6), and thus is an element of $\mathrm{U}(n)$. One can then verify that the function $\Lambda\colon \Gamma \times Y \to \mathrm{U}(n)$ is a cocycle.

Let $\mathcal{M}_k$ be a sequence of invariant finite rank $\mathrm{L}^\infty(Y)$-submodules of $\mathrm{L}^2(X \mid Y)$, such that the closure of their union is dense in $\mathrm{L}^2(X)$. For each $\mathcal{M}_k$, we are given the homogeneous space $\mathrm{U}(n_k)/\mathrm{U}(n_k - 1)$, the cocycle Λ_k, and the measurable function $\theta_k\colon X \to \mathrm{U}(n_k)$ as described above.

Define $K \coloneqq \prod_k \mathrm{U}(n_k)$ and $L \coloneqq \prod_k \mathrm{U}(n_k - 1)$. Consider the homogeneous space K/L, the cocycle $\Lambda \coloneqq (\Lambda_k)_k$, and the measurable function $\theta\colon X \to K/L$ defined by $\theta = (\theta_k)_k$.

Let $\pi\colon X \to Y \times K/L$ be the measurable map defined by $\pi(x) \coloneqq (p(x), \theta(x))$ for almost every x. Consider the pushforward measure $\pi_*\mu_X$ and the cocycle action induced by τ_Λ on $Y \times K/L$. Then $(Y \times K/L, \tau_\Lambda)$ is a homogeneous extension of (Y, σ).

[3]The unitary group $\mathrm{U}(n)$ acts transitively on the sphere S^{2n-1}, and the subgroup $\mathrm{U}(n-1)$, embedded inside $\mathrm{U}(n)$, stabilizes the north pole $(1, 0, 0, \dots, 0)$. By the Orbit-Stabilizer Theorem, we have the identification $S^{2n-1} \cong \mathrm{U}(n)/\mathrm{U}(n-1)$.

We claim that $\pi\colon (X,\tau) \to (Y \times K/L, \tau_\Lambda)$ is an isomorphism of measure-preserving systems. To prove this, we first show that $\bigcup_k \mathcal{M}_k \subseteq U_\pi(\mathrm{L}^2(Y \times K/L))$.

Viewing $\mathrm{L}^2(Y)$ as a subspace of $\mathrm{L}^2(Y\times K/L)$, we observe that $U_\pi(\mathrm{L}^2(Y)) = U_p(\mathrm{L}^2(Y))$. Therefore, it suffices to show that for every k and every conditional orthonormal basis element $f \in \mathcal{M}_k$, there exists $g \in \mathrm{L}^2(Y \times K/L)$ such that $U_\pi(g) = f$. To construct such a g, we can choose g as the corresponding coordinate projection from K/L to $\mathbb{C}$.

Furthermore, the relationship $\theta \circ \tau_\gamma = (\Lambda_\gamma \circ p) \cdot \theta$ holds almost surely for all $\gamma \in \Gamma$. Thus, π defines an isomorphism of measure-preserving systems, proving the claim.

Theorem 12.2.3 now follows directly from Lemma 12.2.5. □

12.3 Proof of Lemma 12.2.5

Lemma 12.2.5 follows from the following proposition (this deduction is left as an exercise).

Proposition 12.3.1. *Let $p\colon (X,\tau) \to (Y,\sigma)$ be a group extension of ergodic concrete measure-preserving systems given by the data (K,ρ,θ). Then there exist a closed subgroup $H \subseteq K$ and an H-valued cocycle $\tilde{\rho}$ that is cohomologous to ρ (when viewed as a K-valued cocycle), such that $p\colon (X,\tau) \to (Y,\sigma)$ and $\tilde{p}\colon Y \rtimes_{\tilde{\rho}} H \to (Y,\sigma)$ are isomorphic extensions.*

We need the following three auxiliary results. In the following proposition, for Hilbert spaces $\mathcal{H}_1 \subseteq \mathcal{H}_2 \subseteq \mathcal{H}$, we denote by $\mathcal{H}_2 \ominus \mathcal{H}_1$ the orthogonal complement of $\mathcal{H}_1$ within $\mathcal{H}_2$.

Proposition 12.3.2. *Let $p\colon (X,\tau) \to (Y,\sigma)$ be an extension of concrete measure-preserving systems. Identify the spaces $\mathrm{L}^2(Y)$, $\mathrm{L}^2(Y_{\mathrm{inv}})$, and $\mathrm{L}^2(X_{\mathrm{inv}})$, where X_{inv} and Y_{inv} denote the invariant factors of (X,τ) and (Y,σ) (see Example 2.2.8), with their respective subspaces of $\mathrm{L}^2(X)$. Then, the subspaces*

$$\mathrm{L}^2(Y) \ominus \mathrm{L}^2(Y_{\mathrm{inv}}) \quad \textit{and} \quad \mathrm{L}^2(X_{\mathrm{inv}})$$

of $\mathrm{L}^2(X)$ are orthogonal.

Proof. Exercise. □

The following result is a special case of Proposition 12.3.1.

Proposition 12.3.3. *Let K be a compact metrizable group. Then K acts on itself by translations from the right preserving the Haar measure m_K. Denote this concrete measure-preserving system by (K,τ). Let $p\colon (K,\tau) \to (Y,\sigma)$ be a factor map of* *concrete measure-preserving K-systems. Then there exists a closed subgroup H of K such that (Y, σ) is isomorphic to $(H\backslash K, \tau)$.*

Proof. Using Lemma 4.1.7, one can show that $A = \mathrm{C}(K) \cap U_p(\mathrm{L}^2(Y))$ is dense in $U_p(\mathrm{L}^2(Y))$. Let H be the set of all elements $h \in K$ such that $f(gh) = f(g)$ for all $g \in K$ and for all $f \in A$. Then H is a closed subgroup of K, and A may be identified with a subalgebra of $\mathrm{C}(H\backslash K)$. By construction, A separates points in $H\backslash K$, and is thus by Theorem 6.1.18 dense in $\mathrm{C}(H\backslash K)$. The claim follows by applying Proposition 2.2.15, Proposition 2.2.13, and Corollary 1.3.9. □

Finally, the following criterion allows us to determine when a homogeneous extension is a homogeneous skew-product extension:

Proposition 12.3.4. *Let $p\colon (X, \tau) \to (Y, \sigma)$ be a homogeneous extension given by the data $(K/L, \rho, \theta)$. Suppose that*

$$\int_X U_p(f) \cdot g \circ \theta \,\mathrm{d}\mu_X = \left(\int_Y f \,\mathrm{d}\mu_Y\right)\left(\int_{K/L} g \,\mathrm{dm}_{K/L}\right)$$

for all $f \in \mathrm{L}^\infty(Y)$ and $g \in \mathrm{C}(K/L)$. Then the homogeneous extension $p\colon (X, \tau) \to (Y, \sigma)$ is equal to the homogeneous skew-product extension $q\colon Y \rtimes_\rho K/L \to (Y, \sigma)$.

Proof. Exercise. □

We start with the proof of Proposition 12.3.1.

Let $X \rtimes_{\mathrm{id}} K$ be a group-skew extension of (X, τ) by the trivial cocycle. Consider a K-action on $X \rtimes_{\mathrm{id}} K$ and $Y \rtimes_\rho K$ given by $(x, k)k' := (x, kk')$ and $(y, k)k' := (y, kk')$, respectively. Since the K-action and the Γ-action commute, both $X \rtimes_{\mathrm{id}} K$ and $Y \rtimes_\rho K$ can be viewed as concrete $\Gamma \times K$-systems. By Fubini's theorem, there exists a factor map $\pi\colon X \rtimes_{\mathrm{id}} K \to Y \rtimes_\rho K$, defined by $\pi((y, k), k') = (y, kk')$.

Let Z be the invariant factor of $Y \rtimes_\rho K$ as a Γ-system, and let Z' be the invariant factor of $X \rtimes_{\mathrm{id}} K$ as a Γ-system. Then Z is a factor of Z' as K-systems since the K-action and the Γ-action commute.

We claim that Z' is isomorphic to K as a K-system. To do so, write $\mathrm{pr}_K\colon X \times K \to K$ be the factor map to K and write η for the dynamics on $X \times K$. We show that $U_{\mathrm{pr}_K}(\mathrm{L}^2(K)) = \mathrm{fix}(U_\eta)$. Translating this back onto the level of systems (cf. Section 2.2), this shows the claim.

For $g \in \mathrm{L}^\infty(K)$ we obtain that $U_{\mathrm{pr}_K} g = \mathbb{1} \odot g \in \mathrm{fix}(U_\eta)$ by definition of η. This implies the inclusion $U_{\mathrm{pr}_K}(\mathrm{L}^2(K)) \subseteq \mathrm{fix}(U_\eta)$. For the converse direction, let $P \in \mathscr{L}(\mathrm{L}^2(X \times K))$ be the orthogonal projection onto $\mathrm{fix}(U_\eta)$. By linearity and continuity it suffices to show that $P(f \odot g) \in U_{\mathrm{pr}_K}(\mathrm{L}^2(K))$ for all $f \in \mathrm{L}^2(X)$ and $g \in \mathrm{L}^2(Y)$.

However, for such f and g we readily obtain, by definition of η, that

$$\overline{\text{co}}\,\{U_{\eta_\gamma}(f \odot g) \mid \gamma \in \Gamma\} = \overline{\text{co}}\,\{U_{\tau_\gamma}(f) \mid \gamma \in \Gamma\} \odot g.$$

Thus, by the mean ergodic theorem (see Theorem 3.1.5) we find some some $h \in \mathrm{L}^2(X)$ with $P(f \odot g) = h \odot g$. Since $P(f \odot g)$ is invariant, h has to be invariant as well, hence $h = c\mathbb{1}$ for some $c \in \mathbb{C}$ by ergodicity of (X, τ). Therefore $P(f \odot g) = \mathbb{1} \odot cg = U_{\mathrm{pr}_K}(cg)$, which yields the claim.

Now, by Proposition 12.3.3, Z is isomorphic to $H\backslash K$ as a K-system, for some closed subgroup H of K. The group H is known as the **Mackey range** of the cocycle ρ.

Before continuing with the proof of Proposition 12.3.1, we establish the following auxiliary result:

Lemma 12.3.5. *Let V be a symmetric neighborhood of the identity in K (that is, $V = V^{-1}$). Then there exists a measurable function $U\colon Y \to K$ such that*

$$U(\sigma_\gamma(y))\rho_\gamma(y)U(y)^{-1} \in V^2HV^2$$

for almost every y and every $\gamma \in \Gamma$.

Proof. By the definition of the Haar measure, $H \cdot V$ has positive measure in K/H. We can identify $H \cdot V$ with a positive measure subset of Z, and therefore with a positive measure subset E of $Y \times K$. Note that for all $k \in K$ such that $k \notin VHV$, we have $HV \cap HVk^{-1} = \emptyset$, and therefore $\mu_Y \otimes \mathrm{m}_K(E \cap Ek) = 0$.

Since E is invariant and has positive measure, it follows from the ergodicity that $\mathbb{E}(\mathbb{1}_E \mid Y) \equiv C > 0$. Since K is compact, $K \subseteq \bigcup_{i=1}^n k_iV$ for some $k_1, \ldots, k_n \in K$, which implies $\mathbb{1}_E \leq \sum_{i=1}^n \mathbb{1}_E \mathbb{1}_{Y \times k_iV}$. Therefore,

$$\sum_{i=1}^n \mathbb{E}(\mathbb{1}_E \mathbb{1}_{Y \times k_iV} \mid Y) > 0.$$

Now define

$$i(y) \coloneqq \min\{i \in \{1, \ldots, n\}\colon \mathbb{E}(\mathbb{1}_E \mathbb{1}_{Y \times k_iV} \mid Y)(y) > 0\}$$

for almost every y, and let $U(y) \coloneqq k_{i(y)}^{-1}$. Then $U\colon Y \to K$ is measurable, and for almost every y

$$\mathbb{E}(\mathbb{1}_E \mathbb{1}_{Y \times U^{-1}(y)V} \mid Y)(y) > 0. \tag{12.3}$$

Since $\mu_Y \otimes \mathrm{m}_K(E \cap Ek) = 0$ for all $k \notin VHV$, we have

$$(\mathbb{1}_E(y,h)\mathbb{1}_{Y \times U^{-1}(y)V}(y,h)) \cdot k(\mathbb{1}_E(y,h)(\mathbb{1}(y,h) - \mathbb{1}_{Y \times U^{-1}(y)V^2HV}(y,h))) = 0$$

for almost every (y, h) and every $k \in K$. Integrating the last identity with respect to k and then taking conditional expectations over Y, we obtain:

$$\mathbb{E}(\mathbb{1}_E(\mathbb{1} - \mathbb{1}_{Y\times U^{-1}V^2HV}) \mid Y) = 0. \tag{12.4}$$

Using the Γ-invariance of E in (12.3) and cocycle property for $0 = \gamma + (-\gamma)$, for all $\gamma \in \Gamma$, we have

$$\mathbb{E}(\mathbb{1}_E \mathbb{1}_{Y\times\rho(\gamma,\sigma_\gamma(\cdot))U^{-1}(\sigma_\gamma(\cdot))V} \mid Y) > 0.$$

Comparing this with the identity (12.4), we conclude:

$$\mathbb{E}(\mathbb{1}_{Y\times U^{-1}V^2HV}\mathbb{1}_{Y\times\rho(\gamma,\sigma_\gamma(\cdot)U^{-1}(\sigma_\gamma(\cdot))V} \mid Y) > 0.$$

This implies that for almost every y, and every $\gamma \in \Gamma$,

$$U^{-1}V^2HV \cap \rho(\gamma, \sigma_\gamma(y))U^{-1}(\sigma_\gamma(y))V \neq \emptyset.$$

In particular, for almost every y, and every $\gamma \in \Gamma$,

$$U(\sigma_{\gamma^{-1}}(y))\rho_{\gamma^{-1}}(y)U(y)^{-1} \in V^2HV^2.$$

□

We now proceed to construct a cocycle cohomologous to ρ that takes values in H:

Proposition 12.3.6. *There exists a measurable function $U \colon Y \to K$ such that the cocycle $\tilde{\rho}(\gamma, y) := U(\sigma_\gamma(y))\rho(\gamma, y)U(y)^{-1} \in H$ for almost every y and every $\gamma \in \Gamma$.*

Proof. Exercise. □

We continue with the proof of Proposition 12.3.1. Consider a cocycle $\tilde{\varrho}$ as in Proposition 12.3.6. We claim that the system $Y \rtimes_{\tilde{\rho}} H$ is ergodic.

By what we have already shown and since ϱ is cohomologous to $\tilde{\rho}$ as a cocycle to K, we can identify the invariant factor Z of the system $Y \rtimes_{\tilde{\rho}} K$ with $H\backslash K$ with the corresponding factor map given by $r := \pi\circ \mathrm{pr}_K \colon Y\times K \to H\backslash K$ where $\pi\colon K \to H\backslash K$ is the canonical factor map and $\mathrm{pr}_K \colon Y \times K \to K$ is the projection onto the second component. For $g \in \mathrm{C}(K)$ we have the quotient integral formula

$$\int_K g(k)\,\mathrm{dm}_K(k) = \int_{H\backslash K}\int_H g(hk)\,\mathrm{dm}_H(h)\,\mathrm{dm}_{H\backslash K}(Hk),$$

see, e.g., [DE09, Theorem 1.5.2]. By Fubini–Tonelli theorem,

$$\int_{Y\times K} f\odot g\,\mathrm{d}(\mu_Y\otimes \mathrm{m}_K) = \int_{H\backslash K}\int_{Y\times H}(f\odot g)(y, hk)\,\mathrm{d}(\mu_Y\otimes \mathrm{m}_H)(y,h)\,\mathrm{d}(Hk)$$

for all $f \in \mathrm{L}^\infty(Y)$ and $g \in \mathrm{C}(K)$.

Write $\tilde{\mathrm{m}}_H$ for the regular Borel probability on K obtained by trivially extending m_H to K, and $r_k \colon K \to K$, $k' \mapsto k'k$ for the right rotation with $k \in K$. Then the pushforward measures $(r_{k_1})_*\tilde{\mathrm{m}}_H$ and $(r_{k_2})_*\tilde{\mathrm{m}}_H$ for $k_1, k_2 \in K$ agree if $Hk_1 = Hk_2$. We can thus define $\mu_{Hk} := \mu_Y \otimes (r_k)_*\tilde{\mathrm{m}}_H$ for $Hk \in H\backslash K$. The above formula yields

$$\int_K f \odot g \, \mathrm{d}(\mu_Y \otimes \mathrm{m}_K) = \int_{K/H} \int_K f \odot g \, \mathrm{d}\mu_{Hk} \, \mathrm{d}(Hk)$$

for all $f \in \mathrm{L}^\infty(Y)$ and $g \in \mathrm{C}(K)$. As usual, we may assume that Y is a compact metric space. Using that the elements $f \otimes g$ for $f \in \mathrm{C}(Y)$ and $g \in \mathrm{C}(K)$ span a dense subset of $\mathrm{C}(Y \times K)$, a moment's thought reveals that $(\mu_{Hk})_{Hk \in H\backslash K}$ defines a disintegration of $Y \times K$ over the invariant factor $H\backslash K$. By Exercise 8.5 we obtain that the system (X_{Hk}, τ_{Hk}) given by the measurable space $Y \times K$, the measure $\mu_{Hk} = \mu_Y \otimes (r_k)_*\tilde{\mathrm{m}}_H$ and the action defined by the cocycle $\tilde{\varrho}$ is ergodic for almost every $Hk \in H\backslash K$. In particular, there is some $Hk \in H\backslash K$ such that (X_{Hk}, τ_{Hk}) is ergodic. But then, since

$$Y \times H \to Y \times K, \quad (y, h) \mapsto (y, hk^{-1})$$

defines an isomorphism between $Y \rtimes_{\tilde{\rho}} H$ and (X_{Hk}, τ_{Hk}), we finally obtain that $Y \rtimes_{\tilde{\rho}} H$ is ergodic.

To conclude the proof, apply Proposition 12.3.2 to deduce that $\mathrm{L}^2(Y \rtimes_{\tilde{\rho}} H) \ominus \mathbb{C}\mathbb{1}$ and the invariant Markov sublattice of $\mathrm{L}^2(X \rtimes_{\mathrm{id}} H)$ are orthogonal subspaces in $\mathrm{L}^2(X \times H)$ (when properly identified). Therefore, for $f \in \mathrm{L}^\infty(Y)$ and $g \in \mathrm{C}(H)$, we have for all $h' \in H$ that

$$\int_X f(y)g(hh') \, \mathrm{d}\mu_X(y, h) = \int_{Y \times H} f(y)g(h) \, \mathrm{d}(\mu_Y \otimes \mathrm{m}_H)(y, h).$$

We plug in $h' = 1$ and apply Proposition 12.3.4. This completes the proof of Proposition 12.3.1.

12.4 Comments and Further Reading

The definition of compact extensions by invariant finite rank modules and its geometric characterization, as presented in Theorem 12.2.3, originates from the foundational work of Mackey [Mac66] and Zimmer [Zim76a, Zim76b]. Our initial definition of compact extensions in terms of conditional almost periodicity, as well as its relation to conditional Hilbert–Schmidt operators, follows Furstenberg's classical book [Fur14], though we adopt the conditional Hilbert space formalism introduced by Tao in [Tao09]. The compilation of the proof of the equivalence between the definitions via conditional almost periodicity and invariant finite rank modules, as presented here, is new, although it draws inspiration and ideas from our work on uncountable Furstenberg–Zimmer structure theory [Jam23, EHK24]. Our proof of Proposition 12.3.1 is based on Tao's blog post [Tao14a], while the proof of Theorem 12.2.3 follows [Jam23], which was in turn inspired by the proof given in Glasner's book [Gla03]. A number of works generalize various aspects of the classification results from the previous two lectures, see [Aus10, Ell87, EK22a, JT22, Jam23, EHK24, EJK23].

12.5 Exercises

Exercise 12.1. Prove Proposition and Definition 12.2.2.

Exercise 12.2. Prove Proposition 12.3.2.

Exercise 12.3. Prove Proposition 12.3.4.

Exercise 12.4. Prove Proposition 12.3.6. *Hint: Using Lemma 12.3.5, find a sequence of* U_n *such that* $\rho^n_\gamma := U_n \circ U_\sigma \rho_\gamma U_n^{-1}$ *has distance at most* $1/n$ *to* H *for all* $\gamma \in \Gamma$*. Then build measurably* U *out of the* U_n *and verify that it satisfies the desired property.*

Exercise 12.5. Prove Lemma 12.2.5. *Hint: Use the Krein–Milman Theorem (Theorem 4.1.8) to locate an ergodic group extension by* K *from the ergodic homogeneous extension by* K/L*, apply Proposition 12.3.1, and finally show that* $M = H \cap L$ *is the right choice to obtain the homogeneous skew-product extension by* H/M*.*

Lecture 13

In the remaining lectures of this ISem, we provide an introduction to Host–Kra structure theory, a significant and technically challenging refinement of Furstenberg–Zimmer structure theory. Host–Kra theory was originally motivated by the difficult problem of establishing L^2-convergence of non-conventional ergodic averages, which arise in Furstenberg's multiple recurrence theorem. Since its inception, the theory has grown into an active and broad area of research with applications to problems in combinatorics and number theory. At its core, Host–Kra theory studies the classification of so-called *characteristic factors* of ergodic measure-preserving systems for abelian group actions. These factors are for example relevant for understanding the behavior of non-conventional ergodic averages and for advancing the inverse theory of the Gowers norms in additive combinatorics (an ergodic theoretic variant of these norms will be introduced in this lecture). Host–Kra structure theory investigates a hierarchy of these factors, similar to the hierarchy in the Furstenberg–Zimmer structure theorem (Theorem 9.1.12). Our introduction will focus on the first non-trivial case in this hierarchy. However, even in this case, due to space and time constraints, we will omit a few technical details, for which references will be provided.

13.1 Motivation

Motivated by Furstenberg's multiple recurrence theorem, it is natural to ask the following question: Let (X, τ) be a concrete ergodic measure-preserving system over $\Gamma = \mathbb{Z}$, let $k \geq 1$ be an integer, and let $f \in \mathrm{L}^\infty(X)$. Do the averages

$$\frac{1}{N}\sum_{n=0}^{N-1}\prod_{i=1}^{k} U_\tau^{in} f$$

converge in $\mathrm{L}^2(X)$, and if so, what is their limit? To simplify notation, henceforth, we use $\mathbb{E}_N$ as an abbreviation for $\frac{1}{N}\sum_{n=0}^{N-1}$.

For $k = 1$, this is the content of von Neumann's mean ergodic theorem (Theorem 3.1.1): The $\mathrm{L}^2(X)$-limit of $\mathbb{E}_N U_\tau^n f$ exists and is equal to the conditional expectation of f with respect to the fixed space $\mathrm{fix}(U_\tau)$. In the case of ergodic systems, this is simply the expectation of f - a constant. We say that the invariant factor is *characteristic* for the averages $\mathbb{E}_N U_\tau^n f$, that is, to study the $\mathrm{L}^2(X)$-limit of these averages, we can substitute f with its conditional expectation $\mathbb{E}(f \mid \mathrm{fix}(U_\tau))$ in the averages.

For $k = 2$, this is the content of Theorem 8.2.2. In this case, the results of Section 8.2 show that the Kronecker subsystem $(X_{\mathrm{kro}}, \tau_{\mathrm{kro}})$ of (X, τ) (cf. Definition 7.1.14) is characteristic for the averages $\mathbb{E}_N U_\tau^n f U_\tau^{2n} f$. By applying the van der Corput inequality (Lemma 8.2.3), it suffices to study these averages whenever $f \in \mathrm{L}^\infty(X_{\mathrm{kro}})$ (cf. Lemma 8.2.4). In this setting, the classification of the Kronecker subsystem as a compact abelian group rotation, provided by the Halmos–von Neumann representation theorem (Theorem 6.2.6), was crucial in proving the convergence result (cf. proof of Theorem 8.2.2).

In Section 13.2, we introduce a hierarchy of subsystems/factors $\mathcal{Z}_0, \mathcal{Z}_1, \ldots$ of the system (X, τ) such that $\mathcal{Z}_{k-1}$ is characteristic for the averages

$$\mathbb{E}_N \prod_{i=1}^{k} U_\tau^{in} f$$

in the sense that the $\mathrm{L}^2(X)$-limit of these averages can be studied by replacing f with its conditional expectation $\mathbb{E}(f \mid \mathcal{Z}_{k-1})$. Thus, it suffices to study the $\mathrm{L}^2(X)$-convergence of these averages whenever f is $\mathcal{Z}_{k-1}$-measurable.

The heart of Host–Kra structure theory lies in classifying these characteristic factors $\mathcal{Z}_k$ (for $\Gamma = \mathbb{Z}$) as inverse limits of homogeneous systems formed on k-nilpotent Lie groups. This algebro-geometric classification can be used to demonstrate $\mathrm{L}^2(X)$-convergence.

In Section 13.3, we identify the first characteristic factor of an ergodic system with its Kronecker subsystem. For the remainder of these lectures, our primary focus will be on the second characteristic factor $\mathcal{Z}_2$, commonly referred to as the Conze–Lesigne factor.

13.2 The Characteristic Factors

Throughout this lecture, we fix a countable discrete abelian group Γ and restrict our attention to concrete measure-preserving systems over Γ defined on a Lebesgue probability space. For this section, let (X, τ) denote an *ergodic* system.

We will obtain the characteristic factors of (X, τ) as coordinate projections from cubic systems constructed from (X, τ). We begin by introducing these cubic systems.

For any integer $k \geq 1$, we denote by $[k]$ the k-dimensional cube $\{0,1\}^k$. Thus, $[k]$ consists of ordered tuples $\varepsilon = \varepsilon_1 \dots \varepsilon_k$, where $\varepsilon_i \in \{0,1\}$ for all $i = 1, \dots, k$. For example, $[2]$ consists of the tuples $\{00, 01, 10, 11\}$.

It will be convenient to fix an ordering on $[k]$, and we choose the lexicographic ordering. For instance, the lexicographic ordering on $[3]$ is

$$000, 001, 010, 011, 100, 101, 110, 111.$$

Let $\mathbf{0}$ denote the first element in $[k]$ under this ordering, for any k.

A **facet** α of $[k]$ is a subset of vertices that forms a $(k-1)$-dimensional cube within $[k]$. For example, the facets of $[2]$ are its edges, while the facets of $[3]$ are its sides. An **upper facet** is a facet that does not contain the vertex $\mathbf{0}$.

A **cube symmetry** $\vartheta\colon [k] \to [k]$ is a bijective map which maps facets to facets. The **symmetry group** $\mathrm{Sym}([k])$ is the set of all such cube symmetries equipped with the composition of maps.

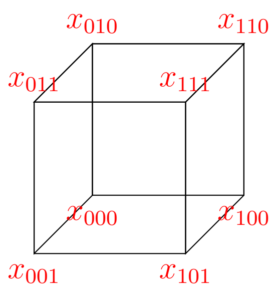
For $\alpha \subseteq [k]$, we denote by $\gamma^\alpha = (\gamma^\alpha_\varepsilon)_{\varepsilon \in [k]}$ the element of $\Gamma^{[k]}$ defined by the entries

$$\gamma^\alpha_\varepsilon := \begin{cases} \gamma & \text{if } \varepsilon \in \alpha, \\ 0 & \text{otherwise.} \end{cases}$$

The **facet group** associated with the group Γ is the smallest subgroup of the product group $\Gamma^{[k]}$ that contains γ^α for all $\gamma \in \Gamma$ and all facets $\alpha \subseteq [k]$. We denote the k^{th} facet group associated with Γ by $\mathcal{F}^k$, and we denote by $\mathcal{F}^k_*$ the subgroup of $\mathcal{F}^k$ generated by elements γ^α, where α is an upper facet.

We define the cubic systems constructed from (X, τ) recursively. Let $(X^{[0]}, \tau^{[0]}) = (X, \tau)$. Suppose we have constructed the cubic system $(X^{[k-1]}, \tau^{[k-1]})$ for some $k \geq 1$. Then, the cubic system $(X^{[k]}, \tau^{[k]})$ is the relatively independent product (see Definition 8.4.5) of two copies of $X^{[k-1]}$ over its invariant factor $(X^{[k-1]})_{\mathrm{inv}}$, where $\tau^{[k]}$ is the diagonal action $\tau \times \cdots \times \tau$. Note that $X^{[k]}$ does *not* carry the 2^k-fold product measure. We denote by $\mu^{[k]} = \mu_{X^{[k]}}$ the relatively independent probability measure of the cubic system $(X^{[k]}, \tau^{[k]})$.

The elements of $X^{[k]}$ are of the form $(x_\varepsilon)_{\varepsilon \in [k]}$ where $x_\varepsilon \in X$ for $\varepsilon \in [k]$, see the Diagram 13.2.



For $k = 1$, the cubic system $(X^{[1]}, \tau^{[1]})$ is the product system $(X \times X, \tau \times \tau)$ due to the ergodicity of (X, τ). By Theorem 7.1.16 and Proposition 7.1.10, the product system $(X \times X, \tau \times \tau)$ is not ergodic unless the system (X, τ) is weakly mixing.

With some effort one can show that the symmetry group of $[k]$ acts in a measure-preserving way on the kth cubic system (see [HK18, Proposition 8 of Chapter 8]):

Theorem 13.2.1. *For each cube symmetry $\vartheta \colon [k] \to [k]$ the map $\vartheta^* \colon X^{[k]} \to X^{[k]}$, defined by $\vartheta^*((x_\varepsilon)_{\varepsilon\in[k]}) := (x_{\vartheta(\varepsilon)})_{\varepsilon\in[k]}$ for $(x_\varepsilon)_{\varepsilon\in[k]} \in X^{[k]}$, preserves the cubic measure $\mu^{[k]}$ on $(X^{[k]}, (\Sigma_X)^{[k]})$. Moreover, the induced measure-preserving $\mathrm{Sym}([k])$-action on $(X^{[k]}, \mu^{[k]})$ commutes with the diagonal action of Γ.*

One can now also show that the facet group acts as automorphisms on the cubic systems:

Proposition 13.2.2. *For every facet $\alpha \subseteq [k]$, the map $\sigma_{\gamma^\alpha} \colon X^{[k]} \to X^{[k]}$, defined by $\sigma_{\gamma^\alpha}((x_\varepsilon)_{\varepsilon\in[k]}) := (\tau_{\gamma^\alpha_\varepsilon} x_\varepsilon)_{\varepsilon\in[k]}$ for $(x_\varepsilon)_{\varepsilon\in[k]} \in X^{[k]}$, preserves the cubic measure $\mu^{[k]}$ on $(X^{[k]}, (\Sigma_X)^{[k]})$. Moreover, the induced measure-preserving $\mathcal{F}^k$-action on $(X^{[k]}, \mu^{[k]})$ commutes with the diagonal action of Γ.*

Proof. Exercise. □

As already pointed out, the cubical systems will generally not be ergodic with respect to the action of Γ. However, the facet group acts ergodically:

Proposition 13.2.3. *Restrict the measure-preserving $\mathcal{F}^k$-action on $(X^{[k]}, \mu_{X^{[k]}})$, as defined in Proposition 13.2.2, to the subgroup $\mathcal{F}^k_*$. Denote by $J^{[k]}$ the invariant σ-algebra of $(X^{[k]}, \mu^{[k]})$ with respect to this $\mathcal{F}^k_*$-action. Then, $J^{[k]}$ is, up to $\mu^{[k]}$-equivalence, equal to $\pi_{\mathbf{0}}^{-1}(\Sigma_X)$, where $\pi_{\mathbf{0}} \colon X^{[k]} \to X$ is the first coordinate projection. In particular, the measure-preserving $\mathcal{F}^k$-action on $(X^{[k]}, \mu^{[k]})$ is ergodic.*

Proof. Exercise. □

Using the cubical systems, we now introduce an important definition.

Definition 13.2.4. For $f \in \mathrm{L}^\infty(X)$ and an integer $k \geq 1$, we define the k^{th} *Gowers–Host–Kra seminorm* $\|f\|_k$ of f by the formula

$$\left(\int_{X^{[k]}} \prod_{\varepsilon\in[k]} C^{|\varepsilon|} f(x_\varepsilon)\right)^{1/2^k} = \left(\int_{X^{[k-1]}} \left|\mathbb{E}\left(\prod_{\varepsilon\in[k-1]} C^{|\varepsilon|} f(x_\varepsilon) \middle| (X^{[k-1]})_{\mathrm{inv}}\right)\right|^2\right)^{1/2^k},$$

where $|\varepsilon| = \sum_{i=1}^k \varepsilon_i \mod 2$ and C is the complex conjugation map.

For example, the second Gowers–Host–Kra seminorm $\|f\|_2$ is

$$\left(\int_{X^{[2]}} f(x_{00})\overline{f(x_{01})f(x_{10})}f(x_{11})\right)^{1/4} = \left(\int_{X^{[1]}} \left|\mathbb{E}(f(x_0)\overline{f(x_1)} \mid (X^{[1]})_{\mathrm{inv}})\right|^2\right)^{1/4}.$$

The fact that the Gowers–Host–Kra seminorms are indeed seminorms on $\mathrm{L}^\infty(X)$ is left as an exercise, as well as the following remarkable property of these seminorms:

Proposition 13.2.5 (Gowers–Cauchy–Schwarz inequality)**.** *For $f_\varepsilon \in \mathrm{L}^\infty(X)$, $\varepsilon \in [k]$, we have*

$$\left|\int_{X^{[k]}} \bigotimes_{\varepsilon\in[k]} f_\varepsilon \, d\mu^{[k]}\right| \leq \prod_{\varepsilon\in[k]} \|f_\varepsilon\|_k. \tag{13.1}$$

For any integer $k \geq 1$, we define the $(k-1)^{\text{th}}$ characteristic factor $\mathcal{Z}_{k-1}$ of the given system (X, τ). To do so, we specify a Γ-invariant σ-subalgebra Σ^{k-1} of Σ_X, which then gives rise to a factor. Informally, a measurable subset $A \subseteq X$ belongs to this σ-subalgebra if, after embedding it into the $\mathbf{0}$-th corner of the cube $X^{[k]}$ (i.e., by considering $\pi_{\mathbf{0}}^{-1}(A)$), it can be identified with a set in $X^{[k]}$ that "does not depend on the first coordinate." We now formalize this intuition.

Let $\mu^{[k]^*}$ be the pushforward measure of $\mu^{[k]}$ with respect to the coordinate projection from $X^{[k]}$ to the (2^k-1)-fold product measurable space $X^{[k]^*}$, where $[k]^* = [k] \setminus \{\mathbf{0}\}$. Then, the measure-preserving action of $\mathcal{F}_*^k$ on $(X^{[k]}, \mu^{[k]})$ induces a measure-preserving action of $\mathcal{F}_*^k$ on $(X^{[k]^*}, \mu^{[k]^*})$. Let $J^{[k]^*}$ denote the $\mathcal{F}_*^k$-invariant factor on $(X^{[k]^*}, \mu^{[k]^*})$. For $E \in J^{[k]^*}$, $X \times E \in J^{[k]}$, and therefore, by Proposition 13.2.3, there exists $F \in X$ such that $\mu^{[k]}((X \times E)\Delta(F \times X^{[k]^*})) = 0$. Conversely, every $E \in \Sigma^{[k]^*}$ satisfying $\mu^{[k]}((X \times E)\Delta(F \times X^{[k]^*})) = 0$ for some $F \in X$ belongs to $J^{[k]^*}$. With these preliminaries at hand, we can define the characteristic factors:

Definition 13.2.6. Let $k \geq 1$ be an integer. We define the σ-algebra Σ^{k-1} to be the set of $F \in \Sigma_X$ such that there exists $E \in J^{[k]^*}$ satisfying

$$\mu^{[k]}((X \times E)\Delta(F \times X^{[k]^*})) = 0.$$

The correspondence between $E \in J^{[k]^*}$ and $F \in \Sigma^{k-1}$ guarantees that Σ^{k-1} is a Γ-invariant σ-subalgebra of Σ_X, and thus induces an ergodic subsystem $\mathcal{Z}_{k-1}$ of (X, τ). We call this subsystem $\mathcal{Z}_{k-1}$ the $(k-1)^{\text{th}}$ **characteristic factor** of the ergodic system (X, τ).

The 0^{th} characteristic factor is simply the invariant factor and is thus trivial due to the ergodicity of (X, τ). The 1^{st} characteristic factor will be identified with the Kronecker subsystem in the next section. We call the 2^{nd} characteristic factor the

Conze–Lesigne factor, and the remaining part of this ISem lecture will be devoted to its classification.

Finally, it follows from von Neumann's mean ergodic theorem (Theorem 3.1.1) applied to the group $\mathcal{F}_*^k$ that $(X^{[k]}, \mu^{[k]})$ is the relatively independent product of $(X^{[k]^*}, \mu^{[k]^*})$ and (X, μ) over $\Sigma^{k-1} \equiv J^{[k]^*}$. This, together with the Gowers–Cauchy–Schwarz inequality (13.1) and the identification of Σ^{k-1} and $J^{[k]^*}$, leads to the following relation between the Gowers–Host–Kra seminorms and the characteristic factors: For $f \in \mathrm{L}^\infty(X)$,

$$\mathbb{E}(f \mid \mathcal{Z}_{k-1}) = 0 \iff \|f\|_k = 0. \tag{13.2}$$

The relation (13.2), combined with the following proposition, shows that the factor $\mathcal{Z}_{k-1}$ is characteristic for the averages $\mathbb{E}_N \prod_{i=1}^k U_\tau^{in} f$.

Proposition 13.2.7. *Let (X, τ) be a concrete ergodic measure-preserving system over $\Gamma = \mathbb{Z}$. Let $k \geq 1$ be an integer and $f_1, \ldots, f_k \in \mathrm{L}^\infty(X)$ with $\|f_i\|_{\mathrm{L}^\infty(X)} \leq 1$ for all $i = 1, \ldots, k$. Then*

$$\limsup_N \left\| \mathbb{E}_N \prod_{i=1}^k U_\tau^{in} f_i \right\|_{\mathrm{L}^2(X)} \leq \min_{1 \leq l \leq k} (l \|f\|_k). \tag{13.3}$$

Proof. We induct on k. For $k = 1$, the mean ergodic theorem (Theorem 3.1.1) yields that

$$\lim_N \|\mathbb{E}_N U_\tau^n f\|_{\mathrm{L}^2(X)} = \left| \int_X f \, \mathrm{d}\mu_X \right| = \|f\|_1.$$

Suppose that (13.3) holds for some $k \geq 1$. Let $f_1, \ldots, f_{k+1} \in \mathrm{L}^\infty(X)$ with $\|f_i\|_{\mathrm{L}^\infty(X)} \leq 1$ for all $i = 1, \ldots, k+1$. Choose $l \in \{2, \ldots, k+1\}$ (the argument below shows that one can treat the case $l = 1$ similarly). For every $n \geq 1$, denote by

$$g_n = \prod_{i=1}^k U_\tau^{in} f_i.$$

By the van der Corput inequality (Lemma 8.2.3),

$$\limsup_N \|\mathbb{E}_N g_n\|_{\mathrm{L}^2(X)}^2 \leq \limsup_H \mathbb{E}_H \left(\limsup_N \left| \mathbb{E}_N \int_X g_{n+h} g_n \, \mathrm{d}\mu_X \right| \right).$$

Denote the right-hand side of the previous inequality by M. We need to show that

$M \leq l^2\|f_l\|_{k+1}^2$. Let $h \geq 1$. By the Cauchy–Schwarz inequality,

$$\begin{aligned}\left|\mathbb{E}_N \int_X g_{n+h}g_n \,\mathrm{d}\mu_X\right| &= \left|\int_X f_1 U_\tau^h f_1 \,\mathrm{d}\mu_X \mathbb{E}_N \left(\prod_{i=2}^{k+1} U_\tau^{(i-1)n}(f_i U_\tau^{ih} f_i)\right)\right| \\ &\leq \|f_1 U_\tau^h f_1\|_{\mathrm{L}^2(X)} \cdot \left\|\mathbb{E}_N \prod_{i=2}^{k+1} U_\tau^{(i-1)n}(f_i U_\tau^{ih} f_i)\right\|_{\mathrm{L}^2(X)}.\end{aligned}$$

It follows from the mean ergodic theorem, the assumption $\|f_1\|_{\mathrm{L}^\infty} \leq 1$, and the inductive assumption that

$$\limsup_N \left|\mathbb{E}_N \int_X g_{n+h}g_n \,\mathrm{d}\mu_X\right| \leq l\|f_l U_\tau^h f_l\|_k.$$

Now,

$$M \leq \limsup_H l\mathbb{E}_H\|f_l U_\tau^h f_l\|_k \leq l^2 \limsup_H \left(\mathbb{E}_H\|f_l U_\tau^h f_l\|_k^{2^k}\right)^{1/2^k}.$$

Define

$$F((x_\varepsilon)_{\varepsilon\in[k]}) := \prod_{\varepsilon\in[k]} f(x_\varepsilon) \text{ for } (x_\varepsilon)_{\varepsilon\in[k]} \in X^{[k]}.$$

From the definition of the seminorm $\|\cdot\|_k$,

$$\mathbb{E}_H\|f U_\tau^h f\|_k^{2^k} = \mathbb{E}_H \int_{X^{[k]}} U_{\tau^{[k]}}^h F \cdot F \,\mathrm{d}\mu^{[k]}.$$

By another application of the mean ergodic theorem and the definition of the seminorm $\|\cdot\|_{k+1}$,

$$\begin{aligned}\lim_H \mathbb{E}_H \int_{X^{[k]}} U_{\tau^{[k]}}^h F \cdot F \,\mathrm{d}\mu^{[k]} &= \int_{X^{[k]}} \mathbb{E}(F \mid \Sigma_{\mathrm{inv}}(X^{[k]}))^2 \,\mathrm{d}\mu^{[k]} \\ &= \int_{X^{[k+1]}} F \odot F \,\mathrm{d}\mu^{[k+1]} = \|f\|_{k+1}^{2^{k+1}}.\end{aligned}$$

This completes the induction step. □

13.3 The Kronecker Factor

We recall some terminology and results from Lectures 5–7. Let $(X_{\mathrm{kro}}, \tau_{\mathrm{kro}})$ denote the Kronecker subsystem of (X, τ) (Definition 7.1.14). By Theorem 6.2.6, $(X_{\mathrm{kro}}, \tau_{\mathrm{kro}})$ is isomorphic to a rotational system (G, τ_c), where G is a compact abelian group equipped with the Haar measure, and $\tau_c\colon \Gamma \to G$ is a group homomorphism. We also refer to $(X_{\mathrm{kro}}, \tau_{\mathrm{kro}})$ as the **Kronecker factor** of the system (X, τ).

We state the following structure theorem for the Kronecker factor:

Theorem 13.3.1 (Structure of Kronecker factor)**.** *The Kronecker factor $(X_{\mathrm{kro}}, \tau_{\mathrm{kro}})$ is isomorphic to the inverse limit of ergodic rotational systems on compact abelian Lie groups.*

A (real) compact abelian Lie group is isomorphic to a group of the form $\mathbb{T}^d \times G$, where $\mathbb{T}$ is a torus and G is a finite abelian group (see [Tao14b, §1.4] for a reference).

Let (G_n) be a sequence of compact abelian groups such that for each pair $m < n$, there is a surjective and continuous group homomorphism $\pi_{m,n}\colon G_n \to G_m$. The inverse limit of the G_n is defined to be the compact abelian group $G = \varprojlim_n G_n$, consisting of all sequences (x_n), where $x_n \in G_n$ for each n, such that for all pairs $m < n$, we have $\pi_{m,n}(x_n) = x_m$, together with the projection maps $\pi_n\colon G \to G_n$ for each n.

Kolmogorov's extension theorem (see, e.g., [Tao11, §2.4]) allows us to form a translation-invariant probability measure μ_G on the inverse limit G from the Haar measures m_{G_n} on the G_n. By the uniqueness of the Haar measure on compact abelian groups, μ_G must coincide with the Haar measure m_G of G.

Proof. Let (G, τ_c) be the rotational system isomorphic to $(X_{\mathrm{kro}}, \tau_{\mathrm{kro}})$ given by Theorem 6.2.6. By the Gleason–Yamabe theorem for compact abelian groups (see, e.g., [Tao14b, §1.4]), G is isomorphic to the inverse limit of compact abelian Lie groups $G_n = G/H_n$, with projection maps $\pi_n\colon G \to G_n$. The Haar measure m_{G_n} on G_n is the pushforward of the Haar measure m_G on G under the projection map π_n.

We define an ergodic action σ_n on G_n by $(\sigma_n)_\gamma(gH_n) := (\tau_c)_\gamma(g)H_n$ for all $\gamma \in \Gamma$ and $gH_n \in G/H_n$. Using the identification $G = \varprojlim_n G_n$, we recover the action τ_c on G from the inverse limit actions $\tilde{\tau}((g_n)) := (\sigma_n(g_n))$. By construction, the projection maps $\pi_n\colon G \to G_n$ are factor maps. □

Next, we aim to identify the Kronecker factor $(X_{\mathrm{kro}}, \tau_{\mathrm{kro}})$ with the first characteristic factor $\mathcal{Z}_1$. For this, we require the following lemma, which follows from Proposition 5.1.10 and Theorem 5.3.1.

Lemma 13.3.2. *The fixed subspace* $\mathrm{fix}(U_{\tau\times\tau})$ *of the product system $(X \times X, \tau \times \tau)$ is spanned by tensors $f \otimes g$, where f, g are eigenfunctions of the system (X, τ).*

Proposition 13.3.3. *The Kronecker subsystem $(X_{\mathrm{kro}}, \tau_{\mathrm{kro}})$ of (X, τ) is isomorphic to the first characteristic factor $\mathcal{Z}_1$ of (X, τ).*

Proof. Let (G, τ_c) be the rotational system isomorphic to $(X_{\mathrm{kro}}, \tau_{\mathrm{kro}})$ given by Theorem 6.2.6, and let $\varphi\colon G \to X_{\mathrm{kro}}$ denote the isomorphism. Let $q_{\mathrm{kro}}\colon X \to X_{\mathrm{kro}}$ be the factor map. Define $q\colon X \times X \to G$ by $q(x_1, x_2) \coloneqq \varphi(q_{\mathrm{kro}}(x_1))\varphi(q_{\mathrm{kro}}(x_2))^{-1}$. We claim that $q^{-1}(\Sigma_G) = \Sigma_{\mathrm{inv}}(X \times X)$ modulo almost sure equality. By definition, $q^{-1}(\Sigma_G) \subseteq \Sigma_{\mathrm{inv}}(X \times X)$. By Lemma 13.3.2, we also have $\Sigma_{\mathrm{inv}}(X \times X) \subseteq q^{-1}(\Sigma_G)$, modulo almost sure equality.

We now show that

$$\mu^{[2]} = \int_G \mu_s \times \mu_s \,\mathrm{dm}_G(s),$$

where μ_s is a probability measure on $X \times X$, defined by

$$\int_{X\times X} f \odot g \,\mathrm{d}\mu_s = \int_G \mathbb{E}(f \mid G)(z) \cdot \mathbb{E}(g \mid G)(z+s) \,\mathrm{dm}_G(z), \tag{13.4}$$

for $f, g \in \mathrm{L}^\infty(X)$.

It is straightforward to verify that

$$\mu \times \mu = \int_G \mu_s \,\mathrm{dm}_G(s). \tag{13.5}$$

Since disintegrations of measures are essentially unique, it follows that $(\mu_s)_{s\in G}$ defines a disintegration of $\mu \times \mu$ over its invariant factor which we can identify with G by the first part of the proof. This proves the claim.

By the above and rotation invariance of the Haar measure, for $f \in \mathrm{L}^\infty(X)$,

$$\begin{aligned}\|f\|_2^4 = \int_G \int_G \int_G &\mathbb{E}(f \mid G)(z_1) \cdot \mathbb{E}(f \mid G)(z_2 + s)\\ &\times \mathbb{E}(f \mid G)(z_1 + z_2) \cdot \mathbb{E}(f \mid G)(z_1 + z_2 + s) \,\mathrm{dm}_G(z_1) \,\mathrm{dm}_G(z_2) \,\mathrm{dm}_G(s).\end{aligned} \tag{13.6}$$

The expression on the right-hand side of (13.6) is the fourth power of the 2nd Gowers–Host–Kra seminorm of $\mathbb{E}(f \mid G)$. Thus, $\|f\|_2 = 0$ if and only if $\mathbb{E}(f \mid G) = 0$ since the 2nd Gowers–Host–Kra seminorm is a norm on rotational systems by Exercise 13.4.

By (13.2), it follows that $(X_{\mathrm{kro}}, \tau_{\mathrm{kro}})$, the Kronecker factor of (X, τ), is isomorphic to the first characteristic factor $\mathcal{Z}_1$. □

13.4 The Conze–Lesigne Factor

Our next and final aim in this ISem is to establish a structure theorem for the Conze–Lesigne factor $\mathcal{Z}_2$. We begin by showing that while $\mathcal{Z}_2$ is not itself a rotational system, it is isomorphic to a compact abelian group skew-product extension of the Kronecker factor $\mathcal{Z}_1$. Furthermore, we will prove something more precise about the associated cocycle: namely, that it is a cocycle of type 2, satisfying a cubic vanishing cohomology condition. This condition is central to deriving the structure theorem for the Conze–Lesigne factor.

Let us first revisit the notion of measurable cocycles and skew-product extensions as introduced in Definition 12.2.2.

Let (Y, σ) be a concrete measure-preserving system over a countable discrete abelian group Γ, and let $(G, \cdot)$ be a compact metrizable group (not necessarily abelian). A measurable function $\rho\colon \Gamma \times Y \to G$, $(\gamma, y) \mapsto \rho_\gamma(y)$ is called a **cocycle** if it satisfies the cocycle property

$$\rho_{\gamma_1+\gamma_2}(y) = \rho\gamma_1(\sigma_{\gamma_2}(y)) \cdot \rho_{\gamma_2}(y) = \rho_{\gamma_2}(\sigma_{\gamma_1}(y)) \cdot \rho_{\gamma_1}(y)$$

for almost every $y \in Y$ and for every pair $\gamma_1, \gamma_2 \in \Gamma$.

A cocycle ρ is said to be a **coboundary** if there exists a measurable function $\varphi\colon Y \to G$ such that

$$\rho_\gamma(y) = \varphi(\sigma_\gamma(y)) \cdot \varphi(y)^{-1},$$

for almost every $y \in Y$ and every $\gamma \in \Gamma$.

Given a cocycle ρ, we can define the notion of a **group skew-product** $Y \rtimes_\rho G$. If $H \leq G$ is a closed subgroup, we also define the **homogeneous skew-product** $Y \rtimes_\rho G/H$. A cocycle ρ defined on an ergodic system (Y, σ) is called **ergodic** if the group or homogeneous skew-product extension it defines is an ergodic system.

Let (X, τ) be a concrete measure-preserving system over Γ. Let $\mathcal{Z}_1 = (Z, \tau_c)$ be the rotational system representing the Kronecker factor of (X, τ). Our next goal is to show that the Conze–Lesigne factor $\mathcal{Z}_2$ is isomorphic to a group skew-product extension $Z \rtimes_\rho G$, where G is a compact metrizable abelian group.

To achieve this, we first introduce some preparatory results and notations in this lecture and prove the aforementioned representation in the next one. We begin with an important result in measurable cohomology that gives a criterion for when a cocycle with values in an abelian group is a coboundary.

Theorem 13.4.1 (Moore–Schmidt theorem)**.** *Let (Y, σ) be a concrete ergodic measure-preserving system over Γ, let G be a metrizable compact abelian group, and let $\rho\colon \Gamma \times Y \to G$ be a cocycle. Then ρ is a coboundary if and only if $\xi \circ \rho\colon \Gamma \times Y \to \mathbb{T}$ is a coboundary for all continuous characters $\xi \in G'$.*

Proof. If ρ is a coboundary, then a direct computation shows that $\xi \circ \rho$ is a coboundary for all $\xi \in G'$. Conversely, assume that for each character $\xi \in \hat{G}'$ there exists a representative φ_ξ of an element in $\mathrm{L}^\infty(Y, \mathbb{T})$ such that for all $\gamma \in \Gamma$,

$$\xi \circ \rho_\gamma(y) = \varphi_\xi(\sigma_\gamma(y))\varphi_\xi(y)^{-1} \tag{13.7}$$

holds for almost every $y \in Y$.

For any $\xi_1, \xi_2 \in \hat{G}$, one sees from comparing (13.7) for $\xi_1, \xi_2, \xi_1 \cdot \xi_2$ that the function $\varphi_{\xi_1 \cdot \xi_2} \cdot \varphi_{\xi_1}^{-1} \cdot \varphi_{\xi_2}^{-1}$ is Γ-invariant, and hence equal in $\mathrm{L}^\infty(Y, \mathbb{T})$ to a constant $c(\xi_1, \xi_2) \in \mathbb{T}$ by ergodicity. As in the proof of Lemma 6.2.12, Lemma 6.2.10 lets us find a homomorphism $w : \mathrm{L}^\infty(Y, \mathbb{T}) \to \mathbb{T}$ with $w(c\mathbb{1}) = c$ for every $c \in \mathbb{T}$. If we define the modified function $\tilde{\varphi}_\xi := \varphi_\xi \cdot w(\varphi_\xi)^{-1}$, then we have $\tilde{\varphi}_{\xi_1 \cdot \xi_2} = \tilde{\varphi}_{\xi_1} \cdot \tilde{\varphi}_{\xi_2}$ for each $\xi_1, \xi_2 \in G'$.

Since U is metrizable, we obtain that $\mathrm{C}(G)$ and hence also $\mathrm{L}^2(G)$ is separable. Since the dual G' defines an orthonormal basis of $\mathrm{L}^2(G)$ (see Proposition 6.1.19), it is therefore at most countable. Hence for almost every $y \in Y$, the map $y \mapsto \tilde{\varphi}_\xi(y)$ is a homomorphism from G' to $\mathbb{T}$, and thus by Pontryagin duality (see, e.g., [DE09, Chapter 3]) takes the form $\tilde{\varphi}_\xi(y) = \xi \circ F(y)$ for some almost everywhere defined map $F \colon Y \to G$, which one can verify to be measurable. Using that G' separates the points of G (see Proposition 6.1.16), one can then check that for all $\gamma \in \Gamma$

$$\rho_\gamma(y) = F(\sigma_\gamma(y)) \cdot F(y)^{-1}$$

for almost every $y \in Y$, giving the claim.

□

The next result gives us a criterion to test when a cocycle with values in an abelian group is ergodic.

Lemma 13.4.2. *Let (Y, σ) be a concrete ergodic measure-preserving system over Γ, let G be a metrizable compact abelian group, and let $\rho \colon \Gamma \times Y \to G$ be a cocycle. Then ρ is not ergodic if and only if there exists a nontrivial character $\xi \in G'$ such that $\xi \circ \rho$ is a coboundary.*

Proof. For $\mathbb{Z}$-systems, the proof is based on Fourier analysis, and can be found in [HK18, Lemma 8 in §5.3]. The same proof extends to systems for arbitrary Γ. □

Proposition 13.4.3. *Let (Y, σ) be a concrete ergodic measure-preserving system over Γ, let G be a metrizable compact abelian group, and let $\rho \colon \Gamma \times Y \to G$ be a cocycle. There exists a closed subgroup $H \leq G$ and an ergodic cocycle $\psi \colon \Gamma \times Y \to H$ such that ρ is cohomologous to ψ if both are viewed as cocycles with values in G.*

Proof. Let $L = \{\xi \in G' : \xi \circ \rho \text{ coboundary}\}$ and H be the annihilator of L in G, that is $\{u \in G \colon \xi(u) = 0 \text{ for all } \xi \in L\}$. Let $\pi : G \to G/H$ be the canonical projection and $\eta = \pi \circ \rho$ be the induced cocycle with values in G/H. As a consequence of Pontryagin duality (see again [DE09, Chapter 3]) the map $(G/H)' \to L$, $\xi \mapsto \xi \circ \pi$ is a group isomorphism. Thus $\xi \circ \eta$ is a coboundary for every $\xi \in (G/H)'$. By Theorem 13.4.1, η is a coboundary and there is a measurable function $F : Y \to G/H$ such that $\eta_\gamma = (f \circ \sigma_\gamma) \cdot f^{-1}$ almost everywhere for all $\gamma \in \Gamma$. Since $G, G/H$ are Polish groups, by [BK96, Theorem 1.2.4], there is a Borel function $l : G/H \to G$ such that $l(uH) \in uH$ for all cosets $uH \in G/H$. Let $F := l \circ f$ and $\psi_\gamma(y) := \rho_\gamma(y) \cdot (F(\sigma_\gamma(y)) \cdot F(y)^{-1})^{-1}$ for $y \in Y$ and $\gamma \in \Gamma$. A straightforward computation shows that, after potential changes on a nullset, ψ maps to H. Thus $\psi \colon \Gamma \times Y \to H$ is a cocycle which is cohomologous to ρ by construction.

Towards a contradiction, assume that ψ is not ergodic. By Lemma 13.4.2, there is a nontrivial $\xi \in H'$ such that $\xi \circ \psi$ is a coboundary. We can extend ξ to some character $\xi' \in G'$ by Lemma 6.2.10. By assumption, $\xi' \in L$ and thus must annihilate H. In particular, its restriction ξ is trivial on H which is contradictory. □

We establish a criterion for when a compact abelian group extension is a group skew-product extension:

Lemma 13.4.4. *Let (Y, σ) be an ergodic concrete measure-preserving system over Γ, let K be a compact metrizable abelian group, and let $\rho \colon \Gamma \times Y \to K$ be an ergodic cocycle. Suppose that (X, τ) is an ergodic group extension of (Y, σ) in the sense of Definition 12.2.4. Then (X, τ) is the group skew-product $Y \rtimes_\rho K$, that is, $\mu_X = \mu_Y \otimes \mathrm{m}_K$.*

Proof. For $\mathbb{Z}$-systems, the proof relies on Fourier analysis and can be found in [HK18, Lemma 4 in §5.2]. The same proof extends to systems for arbitrary Γ. □

The following proposition will help us identify the Conze–Lesigne factor $\mathcal{Z}_2$ as a homogeneous skew-product extension of $\mathcal{Z}_1$.

Proposition 13.4.5. *Let (X, τ) be a concrete ergodic measure-preserving system over Γ, and let $q_{\mathrm{kro}} \colon (X, \tau) \to (Z, c)$ be the factor map to the Kronecker factor (Z, c) of (X, τ). Suppose $q \colon (X, \tau) \to (Y, \sigma)$ is an extension of concrete measure-preserving systems over Γ.*

If Σ_X is generated by the pullback of Σ_Y and Σ_Z (up to almost sure equivalence), then there exists a closed subgroup $H \leq Z$ and a cocycle $\rho \colon \Gamma \times Y \to H$ such that the extension $q \colon (X, \tau) \to (Y, \sigma)$ is isomorphic to the group skew-product extension $p \colon Y \rtimes_\rho H \to (Y, \sigma)$.

Proof. Define a cocycle $\psi \colon \Gamma \times Y \to Z$, $(\gamma, y) \mapsto \psi_\gamma(y)$ by $\psi_\gamma := c_\gamma \mathbb{1} \colon Y \to Z$ for all $\gamma \in \Gamma$. By Proposition 13.4.3, there exists a closed subgroup $H \leq Z$ and an ergodic cocycle $\rho \colon \Gamma \times Y \to H$ that is cohomologous to ψ within Z. That is, there exists a measurable function $f \colon Y \to Z$ such that

$$\psi_\gamma(y) = f(\sigma_\gamma(y)) \cdot f(y)^{-1} \cdot \rho_\gamma(y),$$

for almost every $y \in Y$ and for all $\gamma \in \Gamma$.

Define $\pi(x) \coloneqq q_{\mathrm{kro}}(x) \cdot f(q(x))^{-1}$ for almost every $x \in X$. A direct computation shows that

$$\pi(\tau_\gamma(x)) = \pi(x) \cdot \rho_\gamma(q(x)),$$

for almost every $x \in X$ and for all $\gamma \in \Gamma$.

Let $l\colon Z \to Z/H$ denote the canonical projection. It follows that $l \circ \pi$ is σ-invariant, and thus, by ergodicity, it must be equal to some constant[1] $dH \in Z/H$ μ_Y-almost surely.

By replacing f with df and π with $d\pi$, we preserve the identities they satisfy while ensuring that π takes values in H.

Define $\Phi\colon X \to Y \times H$ by $\Phi(x) \coloneqq (q(x), \pi(x))$, and let $\lambda = \Phi_*\mu_X$ denote the corresponding pushforward measure. Since $\Phi \circ \tau_\gamma = (\tau_\rho)_\gamma \circ \Phi$ for all $\gamma \in \Gamma$ (where τ_ρ denotes the skew-product dynamics given by the cocycle ρ), we obtain that Φ is a factor map from (X, τ) to $Y \rtimes_\rho H$.

Since λ projects onto μ_X, it follows from Lemma 13.4.4 that $\lambda = \mu_Y \otimes \mathrm{m}_H$. Furthermore, since Σ_X is generated by the pullbacks of Σ_Y and Σ_Z (up to almost sure equivalence), Φ establishes an isomorphism. □

[1]The composition of l with any character $\xi \in (Z/H)'$ is also σ-invariant and hence a constant d_ξ. The map $\xi \mapsto d_\xi$ is a group homomorphism, and therefore, by Pontryagin duality, corresponds to an element of Z/H.

13.5 Comments and Further Reading

The L^2-convergence of the double averages $\mathbb{E}_N U_\tau^n f U_\tau^{2n} f$ was first established by Furstenberg [Fur77]. The concept of characteristic factors was explicitly introduced in Furstenberg and Weiss [FW96a]. The existence of limits for $k = 3$, under the additional assumption that the system is totally ergodic, was shown by Conze and Lesigne in a series of papers [CL84, CL88a, CL88b]. The general case for arbitrary k was independently established by Host and Kra [HK05] and Ziegler [Zie07]. Although they defined the characteristic factors differently, Leibman [Ber06] later showed that these definitions are equivalent for $\mathbb{Z}$-actions.

In this lecture, we follow the exposition of Host and Kra as presented in their textbook [HK18]. While the results in [HK18] are stated for $\Gamma = \mathbb{Z}$, the proofs of the results we needed easily adapt to the general case of arbitrary countable abelian Γ.

The proof of the Moore–Schmidt theorem is taken from [JT23b] by Tao and the first author. This result originally goes back to the work of Moore and Schmidt [MS80].

13.6 Exercises

Exercise 13.1. Prove Proposition 13.2.2.

Exercise 13.2. Prove Proposition 13.2.3.

Exercise 13.3. Prove that the Gowers–Host–Kra seminorms satisfy the properties of a seminorm.

Exercise 13.4. Prove that if (X, τ) is a rotational system (G, τ_c), then the Gowers–Host–Kra seminorms are norms for all $k \geq 2$.

Exercise 13.5. Prove Proposition 13.2.5.

Exercise 13.6. Establish (13.2) by filling in the details of the proof sketch in the paragraph before (13.2).

Lecture 14

We continue the study of the Conze–Lesigne factor $\mathcal{Z}_2$ of an ergodic measure-preserving system (X, τ) over a countable abelian group Γ.

Our first goal in this lecture is to show that $\mathcal{Z}_2$ is isomorphic to a compact abelian group skew-product extension $Z \rtimes_\rho G$ of the Kronecker factor (Z, τ_c) of (X, τ), where the cocycle ρ satisfies a specific cohomological condition of type 2.

In the second part, we provide a detailed outline of the proof of the structure theorem for Conze–Lesigne systems, which characterizes these systems as inverse limits of translational systems formed on locally compact 2-nilpotent groups.

14.1 Abelian Extensions

To prove that $\mathcal{Z}_2$ is isomorphic to a compact abelian group skew-product extension $Z \rtimes_\rho G$, we first identify $\mathcal{Z}_2$ as a compact extension of the Kronecker factor (Z, τ_c). By the Mackey–Zimmer representation theorem (Theorem 12.2.3), $\mathcal{Z}_2$ is isomorphic to a homogeneous skew-product extension $Z \rtimes_\rho G/H$ for some compact (not necessarily abelian) group G. In the second part, we use the properties of the characteristic factor $\mathcal{Z}_2$ as a certain projection of the cubic system $X^{[2]}$ to show that G/H is actually abelian.

We start with some auxiliary results.

Lemma 14.1.1. *Let (Y, σ) be an ergodic concrete measure-preserving system over Γ, let G be a metrizable compact group, let $H \leq G$ be a closed subgroup, and let $\rho \colon \Gamma \times Y \to G$ be an ergodic cocycle. Let $p \colon (X, \tau) = Y \rtimes_\rho G/H \to (Y, \sigma)$ be the associated ergodic homogeneous skew-product extension. Then every $\tau \times \tau$-invariant measurable set $E \subseteq X \times X$ is also $V \times V$-invariant, where V is the vertical translation action defined by $V_g(y, hH) = (y, g \cdot hH)$ for $g \in G$, $y \in Y$, and $hH \in G/H$.*

Proof. Let (Z, τ_c) be the Kronecker factor of (X, τ) with factor map $p_{\mathrm{kro}} \colon (X, \tau) \to (Z, \tau_c)$. Define W as the subsystem of (X, τ) corresponding to the invariant σ-

subalgebra generated by p_{kro} and p.

Thus, we have a chain of extensions $(X,\tau) \to W \to (Y,\sigma)$. The conditional expectation operator $\mathbb{E}(\cdot \mid W)$ from $L^2(X)$ to $L^2(W)$ commutes with the Γ-action and with multiplication by $L^\infty(Y)$. Therefore, the image of an invariant finite rank $L^\infty(Y)$-module in $L^2(X \mid Y)$ under the conditional expectation operator $\mathbb{E}(\cdot \mid W)$ is an invariant finite rank $L^\infty(Y)$-module in $L^2(W \mid Y)$. By Exercise 9.5, $p\colon (X,\tau) \to (Y,\sigma)$ is a compact extension. By Theorem 11.3.3, the union of all invariant finite rank $L^\infty(Y)$-modules in $L^2(X \mid Y)$ is dense in $L^2(X)$, and thus the union of their images under the conditional expectation map $\mathbb{E}(\cdot \mid W)$ is dense in $L^2(W)$. An inspection of the proof of Theorem 12.2.3 reveals that W is of the form $Y \rtimes_\rho G/K$ for some closed subgroup $K \leq G$ (for details, see the proof of Lemma 2 and the "References and Further Comments" section in [HK18, Chapter 5]).

The Kronecker factor of W is also Z. By the first part of the proof of Proposition 13.3.3, $\Sigma_{\text{inv}}(X \times X) = \Sigma_{\text{inv}}(W \times W)$ (almost surely).

Proposition 13.4.5 implies that $W = Y \rtimes_\psi L$ for some compact abelian group L. It follows that $G/K = L$, and in particular, K is normal (for more details, see the proof of Lemma 2 in [HK18, Chapter 5]). Since W is an abelian group skew-product extension, $(\tau_\psi)_\gamma \circ V_g = V_g \circ (\tau_\psi)_\gamma$ for all $\gamma \in \Gamma$ and $g \in G$, where τ_ψ denotes the cocycle action on $W = Y \rtimes_\psi L$. Let $f \in L^2(W)$ be an eigenfunction with eigenvalue $\chi \in \Gamma^*$. Then $U_\tau^\gamma(f \circ V_g) = V_g(U_\tau^\gamma f) = \chi_\gamma f \circ V_g$ for all $\gamma \in \Gamma$ and $g \in G$. Thus, $L^2(Z) \subseteq L^2(W)$ is V_g-invariant for every $g \in G$. This implies that V_g is an automorphism of the Kronecker system (Z,τ_c), i.e., a rotation on Z (see, e.g., Corollary 9 of [HK18, Chapter 4]).

By the first part of the proof of Proposition 13.3.3, $\Sigma_{\text{inv}}(W \times W) = \pi^{-1}(\Sigma_Z)$, where $\pi(w_0,w_1) = p_{\text{kro}}(w_0) \cdot p_{\text{kro}}(w_1)^{-1}$. Hence, for $E \in \Sigma_{\text{inv}}(W \times W)$, we have

$$(V_g \times V_g)(E) = (V_g \times V_g)(\pi^{-1}(F)) = \pi^{-1}(F).$$

Thus, E is $V_g \times V_g$-invariant for all $g \in G$. □

For the next proposition, we need the following observation. Let $p\colon (X,\tau) \to (Y,\sigma)$ be a factor map of concrete ergodic measure-preserving systems over Γ. By induction, using the definition of the cubic measures and noting that $\Sigma_{\text{inv}}(X^{[k]}) = (p^{[k]})^{-1}(\Sigma_{\text{inv}}(Y^{[k]}))$, one can show that $p^{[k]}\colon X^{[k]} \to Y^{[k]}$, given by $p^{[k]}((x_\varepsilon)_{\varepsilon\in[k]}) = (p(x_\varepsilon))_{\varepsilon\in[k]}$ for $(x_\varepsilon)_{\varepsilon\in[k]} \in X^{[k]}$, is a factor map of the respective kth cubic systems.

Proposition 14.1.2. *Let (X,τ) be an ergodic concrete measure-preserving system over Γ, let $k \geq 1$, and let $\mathcal{Z}_k$ be the kth characteristic factor of (X,τ). Then $\Sigma_{\text{inv}}(X^{[k]}) \subseteq (q_k^{[k]})^{-1}(\Sigma_{\mathcal{Z}_k^{[k]}})$ modulo $\mu^{[k]}$-null sets, where $q_k\colon (X,\tau) \to \mathcal{Z}_k$ is the factor map.*

Proof. Let $f_\varepsilon \in \mathrm{L}^\infty(X)$, $\varepsilon \in [k]$, be such that $\mathbb{E}(f_\varepsilon|\mathcal{Z}_k) = 0$ for at least one $\varepsilon \in [k]$. By the definition of $\mu^{[k+1]}$ and the Gowers–Cauchy–Schwarz inequality (Proposition 13.4),

$$\int_{X^{[k]}} \left| \mathbb{E}\left(\bigotimes_{\varepsilon\in[k]} f_\varepsilon \middle| (X^{[k]})_{\mathrm{inv}} \right) \right|^2 \mathrm{d}\mu^{[k]} = \int_{X^{[k+1]}} \bigotimes_{\varepsilon\in[k]} f_\varepsilon \bigotimes_{\varepsilon\in[k]} \bar{f}_\varepsilon \, \mathrm{d}\mu^{[k+1]} = 0.$$

By a telescoping sum argument for arbitrary $f_\varepsilon \in \mathrm{L}^\infty(X)$, where $\varepsilon \in [k]$, we have

$$\mathbb{E}\left(\bigotimes_{\varepsilon\in[k]} f_\varepsilon \middle| (X^{[k]})_{\mathrm{inv}} \right) - \mathbb{E}\left(\bigotimes_{\varepsilon\in[k]} U_{q_k}\mathbb{E}(f_\varepsilon|\mathcal{Z}_k) \middle| (X^{[k]})_{\mathrm{inv}} \right) = 0.$$

Since the elements $\bigotimes_{\varepsilon\in[k]} f_\varepsilon \in \mathrm{L}^\infty(X^{[k]})$ where $f_\varepsilon \in \mathrm{L}^\infty(X)$ for $\varepsilon \in [k]$ span a dense linear subspace of $\mathrm{L}^2(X^{[k]})$, this implies that the fixed space of the system $(X^{[k]}, \tau^{[k]})$ is contained in the image of $U_{q_k^{[k]}} \colon \mathrm{L}^2(\mathcal{Z}_k^{[k]}) \to \mathrm{L}^2(X^{[k]})$. Switching to the level of σ-algebras (cf. Section 2.2), this gives us the claim. □

Theorem 14.1.3. *Let (X, τ) be an ergodic measure-preserving system over Γ, and let $k \geq 1$. Then the kth characteristic factor $\mathcal{Z}_k$ is a compact extension of the $(k-1)$th characteristic factor $\mathcal{Z}_{k-1}$.*

Proof. The proof of this result relies on Furstenberg–Zimmer structure theory as developed in Lectures 9-12 and properties of the cubic systems. A proof can be found in [HK18, Lemma 2 §18.2] for $\Gamma = \mathbb{Z}$ which can be adapted to the general case of arbitrary countable abelian groups. □

An ergodic measure-preserving system (X, τ) over Γ is said to be of **order** k if (X, τ) is isomorphic to its kth characteristic factor $\mathcal{Z}_k$.

Proposition 14.1.4. *Let (X, τ) be an ergodic measure-preserving system over Γ of order 2. Then (X, τ) is isomorphic to a group skew-product extension $Z \rtimes_\rho G$, where (Z, τ_c) is the Kronecker factor of (X, τ) and G is a compact abelian group.*

Proof. By Theorem 14.1.3 and Theorem 12.2.3, (X, τ) is isomorphic to a homogeneous skew-product extension $Z \rtimes_\rho G/H$, where (Z, τ_c) is the Kronecker factor of (X, τ), and we have $\mu_X = \mathrm{m}_Z \otimes \mathrm{m}_{G/H}$. By Fubini's theorem, $\mu_X^{[1]} = \mathrm{m}_Z^{[1]} \times \mathrm{m}_{(G/H)^{[1]}}$ where $(G/H)^{[1]} := (G \times G)/(H \times H)$.

It follows from (13.4),(13.5), and Fubini's theorem that $\mu_X^{[1]}$ is relatively independent with respect to $\mathrm{m}_Z^{[1]}$ in the sense that for $f, g \in \mathrm{L}^\infty(X)$,

$$\int_{X^{[1]}} f \odot g \, \mathrm{d}\mu_X^{[1]} = \int_{Z^{[1]}} \mathbb{E}(f \mid Z) \odot \mathbb{E}(g \mid Z) \, \mathrm{dm}_Z^{[1]}. \tag{14.1}$$

Let $((\mathrm{m}_Z^{[1]})_\omega)_{\omega\in\Omega}$ denote the ergodic decomposition of $\mathrm{m}_Z^{[1]}$ (cf. Exercise 8.5), where for notational convenience, we write Ω for the invariant factor of $Z^{[1]}$, and denote its probability measure by P. By (14.1), for $f, g \in \mathrm{L}^\infty(X)$,

$$\begin{aligned}\int_{X^{[1]}} f\odot g\,\mathrm{d}\mu_X^{[1]} &= \int_\Omega\int_{Z^{[1]}} \mathbb{E}(f\mid Z)\odot\mathbb{E}(g\mid Z)\,\mathrm{d}(\mathrm{m}_Z^{[1]})_\omega\,\mathrm{d}P(\omega)\\ &= \int_\Omega\int_{X^{[1]}} f\odot g\,\mathrm{d}(\mu_X^{[1]})_\omega\,\mathrm{d}P(\omega),\end{aligned}$$

where $(\mu_X^{[1]})_\omega := (\mathrm{m}_Z^{[1]})_\omega \otimes \mathrm{m}_{(G/H)^{[1]}}$ for almost every $\omega\in\Omega$. Thus, by Proposition 14.1.2, $((\mu_X^{[1]})_\omega)$ is the ergodic decomposition of $\mu_X^{[1]}$. By definition, we also have

$$\mu^{[2]} = \int_\Omega (\mu_X^{[1]})_\omega \times (\mu_X^{[1]})_\omega dP(\omega) \tag{14.2}$$

For almost every $\omega\in\Omega$, consider the ergodic systems $(Z^{[1]}, \tau_c^{[1]})$ equipped with the measure $(\mathrm{m}_Z^{[1]})_\omega$, and consider the ergodic homogeneous skew-product extension $Z^{[1]} \rtimes_{\rho^{[1]}} (G/H)^{[1]}$ equipped with the measure $(\mu_X^{[1]})_\omega$.

Let $E\subseteq X^{[2]}$ be a $\tau^{[2]}$-invariant measurable set for $\mu_X^{[2]}$, let $g\in G$, and let $\varepsilon=0$. By the decomposition in (14.2), $\mu_X^{[1]})_\omega \times \mu_X^{[1]})_\omega((\tau^{[2]})^{-1}(E)\Delta E) = 0$ for almost every ω, recall that Δ denotes set symmetric difference. Thus E is invariant in the product of $X^{[1]}$, equipped with the measure $(\mu_X^{[1]})_\omega$, with itself. On $(G/H)^{[2]} := G^{[2]}/H^{[2]}$, we can define the translation $g^{(\varepsilon)}\times g^{(\varepsilon)}$ by translating $(g_{00}H, g_{01}H, g_{10}H, g_{11}H)$ to $(gg_{00}H, g_{01}H, gg_{10}H, g_{11}H)$, and denote by $V_g^{(\varepsilon)}\times V_g^{(\varepsilon)}$ the corresponding vertical translation on the product of $Z^{[1]}\times(G/H)^{[1]}$ with itself. Since this system is the ergodic homogeneous skew-product extension $Z^{[1]} \rtimes_{\rho^{[1]}} (G/H)^{[1]}$ equipped with the measure $(\mu_X^{[1]})_\omega$, we can apply 14.1.1, to have that E is $V_g^{(\varepsilon)}\times V_g^{(\varepsilon)}$-invariant up to $(\mu_X^{[1]})_\omega\times\mu_X^{[1]})_\omega$-null sets for almost every ω. By (14.2), E is $V_g^{(\varepsilon)}\times V_g^{(\varepsilon)}$-invariant up to $\mu_X^{[2]}$-null sets. Now $V_g^{(\varepsilon)}\times V_g^{(\varepsilon)}$ corresponds to a transformation V_g^α of $X^{[2]}$, where $\alpha\subseteq[2]$ is an edge. Similarly, this holds when replacing $\varepsilon=0$ by $\varepsilon=1$. Since $\mu_X^{[2]}$ is preserved by the facet group $\mathcal{F}^2$ by Proposition 13.2.2 and by symmetry (see 13.2.1), it follows that V_g^α acts trivially on the invariant factor of $X^{[2]}$ for any edge $\alpha\subseteq[2]$.

Let $\varepsilon\in[2]$, $g,h\in G$, and α,β be two edges of $[2]$ such that $\alpha\cap\beta=\{\varepsilon\}$. By the previous, V_g^α, V_h^β are $\mu_X^{[2]}$-preserving transformations and act trivially on the invariant factor of $X^{[2]}$. Thus, the commutator $[V_g^\alpha, V_h^\beta] = V_{[g,h]}^\varepsilon$ also preserves the measure $\mu_X^{[2]}$, where $V_{[g,h]}^\varepsilon$ is the vertical translate by $[g,h]$ at the vertex ε in $X^{[2]}$. By symmetry, $V_{[g,h]}^\varepsilon$ preserves the measure $\mu_X^{[2]}$ for every $\varepsilon\in[2]$. Thus, for any

$f \in \mathrm{L}^\infty(X)$,

$$\|f \circ V_{[g,h]} - f\|_2^4 = \int_{X^{[2]}} \bigotimes_{\varepsilon \in [2]} C^{|\varepsilon|}(f \circ V_{[g,h]} - f)\, \mathrm{d}\mu_X^{[2]} = 0.$$

By (13.2), $f = f \circ V_{[g,h]}$, and therefore $V_{[g,h]}$ acts trivially on X. But this can only be the case if $[g,h] = 1$, and thus G is abelian. □

Next, we will show that $\mathcal{Z}_2$ is an abelian skew-product extension of $\mathcal{Z}_1$ by a special type of cocycles:

Definition 14.1.5 (Order of cocycles)**.** Let (Y, σ) be a concrete ergodic measure-preserving system over Γ, let $\rho\colon \Gamma \times Y \to G$ be a cocycle, where G is a compact abelian group, and let $k \geq 1$ be an integer. We denote by $\Delta^{[k]}\rho\colon \Gamma \times Y^{[k]} \to G$ the cocycle defined by

$$(\Delta^{[k]}\rho)_\gamma(y) \coloneqq \prod_{\varepsilon \in [k]} \rho_\gamma(y_\varepsilon)^{(-1)^{|\varepsilon|}} \text{ for } y = (y_\varepsilon)_{\varepsilon \in [k]} \in Y^{[k]} \text{ and } \gamma \in \Gamma,$$

where $|\varepsilon| = \sum_{i=1}^k \varepsilon_i$, as before.

A cocycle ρ is said to be **of type** k if $\Delta^{[k]}\rho$ is a coboundary over the cubic system $Y^{[k]}$.

We need the following lemma before proving that the cocycle ρ appearing in the representation of the Conze–Lesigne factor in Proposition 14.1.4 is of order 2.

Lemma 14.1.6. *Let (Y, σ) be a concrete measure-preserving system over Γ, and let $\rho\colon \Gamma \times Y \to \mathbb{T}$ be a cocycle, where $\mathbb{T}$ is the torus. Let $((\mu_Y)_\omega)_{\omega \in \Omega}$ be the ergodic decomposition of μ_Y (cf. Exercise 8.5). Then*

$$C = \{\omega \in \Omega\colon \rho \text{ coboundary of } (Y, \sigma) \text{ with measure } (\mu_Y)_\omega\}$$

is a measurable set, and the cocycle ρ is a coboundary of (Y, σ) if and only C has full measure.

Proof. For a proof, see [HK18, Lemma 11 §5.3]. The proof provided there is for systems over $\Gamma = \mathbb{Z}$; however, the same argument works for systems over arbitrary countable abelian Γ, provided one replaces Birkhoff's pointwise ergodic theorem with a suitable generalization, such as the one given in [Lin01]. □

Proposition 14.1.7. *Let (X, τ) be an ergodic measure-preserving system over Γ of order 2 and let $X = Z \rtimes_\rho G$ be the its abelian skew-product representation given by Proposition 14.1.4. Then the cocycle ρ is of type 2.*

Proof. By Theorem 13.4.1, it suffices to show that $\xi \circ \Delta^{[2]}\rho$ is a coboundary of $Z^{[2]}$ for all $\xi \in G'$. Fix $\xi \in G'$.

Define $\psi\colon Z \times G \to \mathbb{T}$ by $\psi(z,u) = \xi(u)$ and consider

$$\Psi := \bigotimes_{\varepsilon\in[2]} C^{|\varepsilon|}\psi \in \mathrm{L}^\infty(X^{[2]}, \mathbb{T}).$$

Let $J\colon \mathrm{L}^2(Z^{[2]}) \to \mathrm{L}^2(X^{[2]})$ be the operator defined by $J(f) = \Psi \cdot f \circ \pi^{[2]}$, where $\pi\colon X \to Z$ is the first-coordinate projection. Since $\Psi \circ (\tau^{[2]})_\gamma = \Psi \cdot \xi(\Delta^{[2]}\rho_\gamma) \circ \pi^{[2]}$, the range $\mathcal{H}_\xi$ of J is a closed $U_{\tau^{[2]}}$-invariant subspace of $\mathrm{L}^2(X^{[2]})$. Therefore, by the mean ergodic theorem, $\mathcal{H}_\xi$ contains $\mathbb{E}_{\mu^{[2]}}(J(f) \mid (X^{[2]})_{\mathrm{inv}})$ for all $f \in \mathrm{L}^2(Z^{[2]})$.

In particular, since $\Psi = J(\mathbb{1}) \in \mathcal{H}_\xi$, there exists $f \in \mathrm{L}^2(Z^{[2]})$ such that $J(f) = \mathbb{E}(\Psi \mid (X^{[2]})_{\mathrm{inv}})$. Since X is a system of order 2, by (13.2),

$$\|\mathbb{E}(\Psi \mid (X^{[2]})_{\mathrm{inv}})\| = \|\psi\|_3^8 \neq 0.$$

Thus, there exists $f \neq 0$ such that $J(f)$ is $U_{\tau^{[2]}}$-invariant. By the definition of J, this implies that for all $\gamma \in \Gamma$,

$$\xi(\Delta^{[2]}\rho_\gamma) \cdot f \circ (\tau_c^{[2]})_\gamma = f.$$

Thus, the set $E = \{|f| \neq 0\}$ is τ_c-invariant and $\mathrm{m}_Z^{[2]}(E) > 0$.

Let $((\mathrm{m}_Z^{[2]})_\omega)_{\omega\in\Omega}$ be the ergodic decomposition of $\mathrm{m}_Z^{[2]}$. By Proposition 14.1.6, the set C of ω such that $\xi \circ \Delta^{[2]}(\rho)$ is a coboundary of $(Z^{[2]}, (\tau_c)^{[2]})$ equipped with the measure $(\mathrm{m}_Z^{[2]})_\omega$ is measurable. Since $\mathrm{m}_Z^{[2]}(E) > 0$, the measurable set C has positive measure.

We show that C is invariant under the group $\mathcal{F}^2$ of facet transformations of $Z^{[2]}$. Let $\alpha \subseteq [2]$ be an edge, $\gamma_0 \in \Gamma$, and $(\tau_c)_{\gamma_0}^\alpha \in \mathcal{F}^2$. Suppose $\omega \in C$ such that $(\tau_c)_{\gamma_0}^\alpha(\omega) \in C$. Then there exists a measurable function $\varphi\colon Z^{[2]} \to \mathbb{T}$ such that

$$\xi \circ \Delta^{[2]}(\rho_\gamma) = \varphi \circ ((\tau_c)_\gamma)^{[2]} \cdot \varphi^{-1}$$

holds $(\mathrm{m}_Z^{[2]})_{(\tau_c)_{\gamma_0}^\alpha(\omega)}$-almost everywhere for all $\gamma \in \Gamma$.

Rewriting this coboundary equation, we have

$$\xi \circ \Delta^{[2]}(\rho_\gamma) \circ (\tau_c)_{\gamma_0}^\alpha = (\varphi \circ (\tau_c)_{\gamma_0}^\alpha) \circ ((\tau_c)_\gamma)^{[2]} \cdot (\varphi \circ (\tau_c)_{\gamma_0}^\alpha)^{-1}$$

$(\mathrm{m}_Z^{[2]})_\omega$-almost everywhere for all $\gamma \in \Gamma$. By the cocycle property,

$$\prod_{\varepsilon\in\alpha} \rho_{\gamma_0}((\tau_c)_\gamma y_\varepsilon)^{(-1)^{|\varepsilon|}} \cdot \left(\prod_{\varepsilon\in\alpha} \rho_{\gamma_0}(y_\varepsilon)^{(-1)^{|\varepsilon|}}\right)^{-1} = ((\Delta^{[2]}(\rho)_\gamma \circ ((\tau_c)_{\gamma_0})^\alpha) \cdot (\Delta^{[2]}(\rho)_\gamma)^{-1})(y)$$

for $(\mathrm{m}_Z^{[2]})_\omega$-almost every $y = (y_\varepsilon)_{\varepsilon\in[2]} \in Z^{[2]}$ and for all $\gamma \in \Gamma$. Combining this with the coboundary equation above, we find that $\xi \circ \Delta^{[2]}(\rho)$ is a coboundary of $(Z^{[2]}, (\tau_c)^{[2]})$ equipped with the measure $(\mathrm{m}_Z^{[2]})_\omega$.

Since the action of $\mathcal{F}^2$ on $\Sigma_{\mathrm{inv}}(Z^{[2]})$ is ergodic by Proposition 13.2.2, we conclude that C has full measure. By Lemma 14.1.6, $\xi \circ \Delta^{[2]}(\rho)$ is a coboundary of $Z^{[2]}$. $\square$

14.2 The Structure Theorem for Conze–Lesigne Systems

Let (X, τ) be a concrete ergodic measure-preserving system over Γ of order 2. So far, we have represented (X, τ) as a skew-product system $Z \rtimes_\rho G$, where (Z, τ_c) is the Kronecker factor of (X, τ), represented as a rotational system on a compact abelian group Z. Here, G is another compact metrizable abelian group, and $\rho\colon Z \times \Gamma \to G$ is a cocycle of type 2. In this section, we will use these properties to deduce a structure theorem for systems of order 2. We provide an outline of the arguments, while complete details can be found in [JST24].

We begin by reducing to skew-product systems of the form $Z \rtimes_\rho G$, where G is a compact abelian Lie group, i.e., a group of the form $\mathbb{T}^d \times F$, where $d \geq 0$ and F is a finite abelian group.

By the Gleason–Yamabe theorem, any compact abelian metrizable group G can be expressed as the inverse limit of a countable inverse system (G_n) of compact abelian Lie groups, with projections $\pi_n\colon G \to G_n$. This enables us to view the order 2 system $X = Z \rtimes_\rho G$ as the inverse limit of the family of order 2 systems $X_n = Z \rtimes_{\rho_n} G_n$, where $\rho_n = \pi_n \circ \rho$.

We have the following structure theorem after this reduction.

Theorem 14.2.1. *Let (X, τ) be a concrete ergodic measure-preserving system over Γ of order 2, represented as a skew-product system $Z \rtimes_\rho G$, where (Z, τ_c) is the Kronecker factor of (X, τ), represented as a rotational system on a compact abelian group Z, G is a compact abelian Lie group, and $\rho\colon Z \times \Gamma \to G$ is a cocycle of type 2. Then there exists a locally compact Polish group*[1] *$\mathcal{G}$ of nilpotency class*[2] *2, a lattice*[3]

[1]A **Polish group** is a topological group whose topology is separable and metrizable by a complete metric. For example, $(\mathbb{R}^n, +)$ is a Polish group.

[2]A group is **nilpotent** if it admits a series of normal subgroups $G_0 = G$ and $G_{i+1} = [G_i, G]$ for $i \geq 0$, where $[H, K]$ denotes the subgroup of G generated by all commutators $[h, k] := h^{-1}k^{-1}hk$ for $h \in H, k \in K$, such that $G_n = \{1\}$ for some $n \geq 1$. The smallest such n is called the **nilpotency class** of G. For example, a group is nilpotent of nilpotency class 2 if its commutator subgroup $[G, G]$ is abelian.

[3]A closed subgroup H of a locally compact group K is called a **lattice** if it is discrete and the quotient space K/H is compact. For example, the integers $\mathbb{Z}$ form a lattice subgroup of the reals

Λ, and a group homomorphism $\varphi\colon \Gamma \to \mathcal{G}$ such that the translation system $(\mathcal{G}/\Lambda, \tau^\varphi)$, where the underlying probability space is the homogeneous space $\mathcal{G}/\Lambda$ equipped with a translation-invariant Borel probability measure, which we call the Haar measure of $\mathcal{G}/\Lambda$, and with the action $\tau^\varphi_\gamma(g\Lambda) := (\varphi(\gamma)\cdot g)\Lambda$, is isomorphic to (X,τ).

We provide an example of a translational system on a homogeneous space of a locally compact 2-nilpotent Polish group.

Example 14.2.2. Let

$$\mathcal{G} = \left\{ \begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} : a,b,c \in \mathbb{R} \right\}$$

be the real Heisenberg group, endowed with the group law of matrix multiplication as inherited from its definition as a subgroup of 3×3 matrices. As a lattice, we consider

$$\Lambda = \left\{ \begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} : a,b,c \in \mathbb{Z} \right\}.$$

We equip $\mathcal{G}$ with the 3-dimensional Lebesgue measure, which preserves the group law. This measure induces a pushforward measure on the homogeneous space $\mathcal{G}/\Lambda$ that is translation-invariant.

Next, we define a group homomorphism $\varphi\colon \mathbb{Z} \to \mathcal{G}$ by sending the generator $1 \in \mathbb{Z}$ to the matrix

$$\begin{bmatrix} 1 & a & 0 \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix}$$

where a, b are linearly independent over $\mathbb{Q}$.

This construction induces an ergodic measure-preserving system $(\mathcal{G}/\Lambda, \tau^\varphi)$ over $\Gamma = \mathbb{Z}$, which is a system of order 2. Verifying these properties are left as an exercise for the interested reader.

From the previous reduction, we then obtain the following structure theorem:

Theorem 14.2.3 (Structure theorem for Conze–Lesigne systems)**.** *Let (X,τ) be an arbitrary concrete ergodic measure-preserving system over Γ of order 2. Then (X,τ) is isomorphic to an inverse limit of translational systems formed on locally compact Polish groups of nilpotency class 2.*

Remark 14.2.4. The difference in the structure theorem for the Conze–Lesigne factor compared to the Halmos–von Neumann representation theorem for the Kro-

---

$\mathbb{R}$.

necker factor lies in the type of systems considered: instead of rotational systems on compact abelian groups (nilpotent groups of nilpotency class 1), as in Theorem 6.2.6, we now focus on inverse limits of translational systems on homogeneous spaces of nilpotent groups of nilpotency class 2. Similarly, the structure theorem for the Kronecker factor in Theorem 13.3.1 considers inverse limits of rotational systems on compact abelian Lie groups.

The advantage of working with compact abelian Lie groups is the following important result that provides a "linearization" for cocycles of order 2:

Theorem 14.2.5. *Let (Z, τ_c) be an ergodic rotational system over Γ, let G be a compact abelian Lie group, and let $\rho\colon \Gamma \times Z \to G$ be a cocycle of type 2. Then for every $z \in Z$, the derivative map $\partial_z\rho\colon \Gamma \times Z \to G$, defined by*

$$\partial_z\rho_\gamma(w) = \rho_\gamma(z \cdot w) \cdot \rho_\gamma(w)^{-1}$$

for almost every $w \in Z$ and each $\gamma \in \Gamma$, is a quasi-coboundary. That is, there exist a group homomorphism $\xi_z\colon \Gamma \to G$ and a measurable function f_z such that for all $\gamma \in \Gamma$ and almost every w,

$$(\partial_z\rho)_\gamma(w) = \xi_z(\gamma) \cdot f_z((\tau_c)_\gamma(w)) \cdot f_z(w)^{-1}.$$

Proof. A proof of this theorem in the generality of arbitrary countable discrete abelian groups Γ is given in [JST24, Theorem 1.13]. □

Now we sketch the proof of Theorem 14.2.1, beginning with the construction of the group $\mathcal{G}$.

Construction of $\mathcal{G}$:

The group $\mathcal{G}$ consists of all pairs $(z, F) \in Z \times \mathrm{L}^\infty(Z, G)$, where where $\mathrm{L}^\infty(Z, G)$ denotes the space of equivalence classes of measurable maps from Z to G, such that there exists a group homomorphism $\xi\colon \Gamma \to G$ satisfying the following condition: for all $\gamma \in \Gamma$ and almost every $w \in Z$,

$$(\partial_z\rho_\gamma)(w) = F(c(\gamma) \cdot w) \cdot F(w)^{-1} \cdot \xi(\gamma).$$

By Fubini's theorem, each $(z, F) \in \mathcal{G}$ induces a measure-preserving transformation $\tau_{z,F}$ on $Z \times G$, defined as

$$\tau_{z,F}(s, u) = (s \cdot z, F(s) \cdot u),$$

where $s \in Z$, $u \in G$, and $F(s)$ is evaluated using a representative of F in $\mathrm{L}^\infty(Z, G)$.

The group law in $\mathcal{G}$ is determined by the composition of such transformations:

$$(z_1, F_1) \cdot (z_2, F_2) = (z_1 \cdot z_2, F_1 \cdot F_2 \circ V_{z_1}),$$

where $F \circ V_z(w) := F(z \cdot w)$ for $w, z \in Z$. The inverse of (z, F) in this group is given by

$$(z, F)^{-1} = (z^{-1}, F^{-1} \circ V_{z^{-1}}).$$

$\mathcal{G}$ is 2-nilpotent:

The commutator subgroup $[\mathcal{G}, \mathcal{G}]$ is identified with the subgroup of pairs $(1, F)$, where $F \in \mathrm{L}^\infty(Z, G)$ satisfies the following condition: there exists a group homomorphism $\xi \colon \Gamma \to G$ such that for all $\gamma \in \Gamma$ and almost every $w \in Z$,

$$F \circ (\tau_c)_\gamma = \xi(\gamma) \cdot F. \tag{14.3}$$

We refer to a function $F \in \mathrm{L}^\infty(Z, G)$ satisfying (14.3) as a G-valued eigenfunction with eigenvalue ξ of the rotational system (Z, τ_c). Denote the collection of all such eigenfunctions by $\mathrm{E}(Z, G)$. The commutator subgroup $[\mathcal{G}, \mathcal{G}]$ can then be identified as $\{1\} \times \mathrm{E}(Z, G)$.

Since $\{1\} \times \mathrm{E}(Z, G)$ lies in the center of $\mathcal{G}$, it follows that $\mathcal{G}$ is nilpotent of nilpotency class 2.

$\mathcal{G}$ is Polish and locally compact:

Denoting by $\mathrm{Hom_c}(Z, G)$ the set of continuous group homomorphisms from Z to G, one can show that $\mathrm{E}(Z, G)$ is isomorphic to $G \times \mathrm{Hom_c}(Z, G)$ as groups.

Equip $\mathrm{Hom}_c(Z, G)$ with the discrete topology. Consequently, $G \times \mathrm{Hom}_c(Z, G)$ becomes a locally compact abelian group, and using the aforementioned isomorphism, we equip $\mathrm{E}(Z, G)$ with the structure of a locally compact abelian group.

Next, we endow $\mathcal{G}$ with a topology. By Koopmanization, we identify $\mathcal{G}$ as a subgroup of the unitary group of $\mathrm{L}^2(Z \times G)$. Using this identification, $\mathcal{G}$ is equipped with the strong operator topology. Since $\mathrm{L}^2(Z \times G)$ is separable, $\mathcal{G}$ is a Polish group.

As G has a countable Pontryagin dual (because it is metrizable), it admits a translation-invariant metric d. Using this metric d, we equip $\mathrm{L}^\infty(Z, G)$ with the topology of convergence in measure, which is metrizable by the metric $\int_Z \min\{1, d(F_1, F_2)\} \mathrm{d}m_Z$, since Z is separable. The restriction of the resulting product topology on $Z \times \mathrm{L}^\infty(Z, G)$ to $\mathcal{G}$ coincides with the strong operator topology, and its restriction to $\{1\} \times \mathrm{E}(Z, G)$ coincides with the locally compact topology introduced on $\mathrm{E}(Z, G)$.

To show that $\mathcal{G}$ is locally compact, let $\pi \colon \mathcal{G} \to Z$ be the projection onto the first coordinate. From Theorem 14.2.5, we know that π is surjective. Since π is also a continuous homomorphism of Polish groups, it is an open map (see, e.g., [BK96]).

The kernel $\ker(\pi) \subseteq \{1\} \times \mathrm{E}(Z, G)$ is closed, and since $\mathrm{E}(Z; G)$ is locally compact, it follows that $\ker(\pi)$ is locally compact. Using the fundamental theorem on homomorphisms of topological groups (see, e.g., [HR79, Theorem 5.25]), we conclude that $\mathcal{G}$ is locally compact.

Transitivity of the action of $\mathcal{G}$ on $Z \times G$:

By Theorem 14.2.5, $\mathcal{G}$ contains a transformation (z, F_z) for each $z \in Z$. Since $\mathcal{G}$ also contains $(1, g)$ for each $g \in G$ (where g is identified with a constant function), $\mathcal{G}$ acts transitively[4] on $Z \times G$.

Stabilizer of $(1, 1)$:

The stabilizer of the point $(1, 1)$ in $\mathcal{G}$ is given by:

$$\Lambda = \{1\} \times \mathrm{Hom}(Z, G),$$

which is a discrete subgroup. To see this, consider $(z, F) \in \mathcal{G}$. If (z, F) stabilizes $(1, 1)$, then $z = 1$ and $(1, F) \in \ker(\pi)$. Since $\ker(\pi) \subseteq \{1\} \times \mathrm{E}(Z; G)$, there exists a group homomorphism $\xi \colon \Gamma \to G$ such that F is a G-valued eigenfunction with eigenvalue ξ. Moreover, since (z, F) stabilizes $(1, 1)$, $F(1) = 1$. We deduce that $F \colon Z \to G$ is a homomorphism.

Homeomorphism between $\mathcal{G}/\Lambda$ and $Z \times G$:

The quotient map $\mathcal{G} \to \mathcal{G}/\Lambda$ is open, as it is a continuous surjection of Polish groups (see [BK96]). Since $\mathcal{G}$ acts transitively on $Z \times G$, $\mathcal{G}/\Lambda$ is homeomorphic to $Z \times G$ (see [MZ55]). Let $j \colon \mathcal{G}/\Lambda \to Z \times G$ denote this homeomorphism.

Translation Action and Haar Measure:

For every $(z, F) \in \mathcal{G}$, the map $j^{-1} \circ \tau_{(z,F)} \circ j$ is a left translation on $\mathcal{G}/\Lambda$. Since $\tau_{(z,F)}$ is measure-preserving on $Z \times G$, the pushforward measure $j_*^{-1}(\mathrm{m}_Z \otimes \mathrm{m}_G)$ is invariant under the translation action of $\mathcal{G}$ on $\mathcal{G}/\Lambda$. As $\mathcal{G}$ is locally compact and nilpotent, it is unimodular, and hence $j_*^{-1}(\mathrm{m}_Z \otimes \mathrm{m}_G)$ must coincide with the Haar measure $\mathrm{m}_{\mathcal{G}/\Lambda}$ of $\mathcal{G}/\Lambda$ (see [Nac65]).

Embedding of Γ into $\mathcal{G}$:

For any $\gamma_1, \gamma_2 \in \Gamma$, the cocycle equation implies that for almost every $z \in Z$,

$$\rho_{\gamma_1}(z) \cdot (\rho_\gamma((\tau_c)_{\gamma_2}(z)))^{-1} = \rho_{\gamma_2}(z) \cdot (\rho_\gamma((\tau_c)_{\gamma_1}(z)))^{-1}.$$

This shows that $(c(\gamma), \rho_\gamma) \in \mathcal{G}$ (with the choice of $\xi \equiv 1$) for all $\gamma \in \Gamma$. Using the translation action described above, we define a measure-preserving action of Γ on $\mathcal{G}/\Lambda$, equipped with the Haar measure $\mathrm{m}_{\mathcal{G}/\Lambda}$.

This construction completes the proof of the structure theorem for Conze—Lesigne systems.

[4]Since the $\mathcal{G}$-action on $Z \times G$ is only defined almost surely and not continuously, a priori we cannot speak of transitivity or the stabilizer of a point, as in the subsequent step. However, these technical problems can be resolved by working with a suitable topological model of (X, τ); see [JST24, Theorem A.4 and Lemma A.6] for details. Henceforth, we assume that the $\mathcal{G}$-action on $Z \times G$ is defined everywhere and is continuous, and all ensuing identities hold everywhere.

14.3 Comments and Further Reading

The fact that systems of order k are abelian skew-product extensions of their $(k-1)$th characteristic factor was established independently in the $\Gamma = \mathbb{Z}$ case by Host and Kra [HK05] and Ziegler [Zie07], albeit using slightly different definitions of these factors, which were later shown to be equivalent by Leibman. The arguments of Host and Kra adapt easily to the general case of arbitrary Γ, as noted in the literature (see, e.g., [BTZ10, JST24]). In this lecture, we provided a proof for general countable abelian Γ in the case of systems of order 2.

The structure of the second characteristic factor, or systems of order 2, was studied in the $\Gamma = \mathbb{Z}$ case by Conze and Lesigne [CL88a], [CL88b] (see also [Rud95], [Mei90], [FW96b], [HK01], [HK02]). In this setting, a more precise structure theorem can be proven, stating that systems of order 2 are inverse limits of nilsystems, which are translational systems formed on nilmanifolds - homogeneous spaces of nilpotent Lie groups. Unfortunately, when Γ is not finitely generated, there are counterexamples showing that this stronger version of the structure theorem fails; see the example presented after [Sha24, Conjecture 2.14] (in the discussion of [Sha24, Theorem 4.3]). The general structure theorem for Conze–Lesigne systems (Theorem 14.2.1) was recently established in [JST24].

Lecture 15

In this final lecture of ISem28, we first provide a brief and condensed overview of the current state of research in Host–Kra structure theory and its application to inverse Gowers theory. In the second part, we offer an anecdotal introduction to the corresponding structure theory for topological dynamical systems.

15.1 Host–Kra theory: What else is known?

Let $k \geq 1$ be an integer, and let (X, τ) be a concrete ergodic measure-preserving system over Γ of order k. One can generalize the arguments in Proposition 14.1.4 and Proposition 14.1.7 to show that (X, τ) is isomorphic to a group skew-product system $\mathcal{Z}_{k-1} \rtimes_\rho G$, where $\mathcal{Z}_{k-1}$ is the subsystem of (X, τ) of order $k-1$, G is a compact metrizable abelian group, and ρ is a cocycle of type k. This result is proved in [HK18, Chapter 18] for $\Gamma = \mathbb{Z}$, but the proof can be adapted to the general case of arbitrary Γ, as has been noted in the literature, see, e.g., [BTZ10, JST24].

Combining these representations for all $k \geq 2$, we obtain that (X, τ) is isomorphic to a chain of skew-product extensions

$$G_1 \rtimes_{\rho_1} G_2 \rtimes_{\rho_2} \cdots \rtimes_{\rho_{k-1}} G_k$$

where all the G_i are compact metrizable abelian groups, the cocycle ρ_i is of type $i+1$ for all i, and the extensions are constructed from left to right. The groups G_i appearing in this representation are called the structure groups of the system (X, τ).

There is a close connection between the topological and algebraic properties of the acting group Γ and the structure groups G_i:

- If $\Gamma = \mathbb{Z}$, Host and Kra established that all G_i for $i \geq 2$ are connected abelian groups for all finite-order systems, see [HK18, §18.5].
- If Γ is an m-torsion group[1], then all the structure groups of finite-order systems

[1] That is, $\gamma^m = 0$ for all $\gamma \in \Gamma$.

are totally disconnected and m-torsion, see [JST23, Theorem 1.4].

In general, we lack a more systematic way to relate the properties of the structure groups to the properties of the acting group, which is important for developing structure theorems. Notice that the previous examples suggest an interplay with Pontryagin duality between compact and discrete groups, where there is a duality between torsion-free (resp. torsion) groups and connectedness (resp. total disconnectedness). Thus, the previous examples may be seen as a reflection of a higher-order form of Pontryagin duality. Discovering such a duality and its implications remains an intriguing open problem in the structure theory of finite-order systems.

As far as structure theorems are concerned, the only case where a complete and satisfactory answer has been provided is the case $\Gamma = \mathbb{Z}$, due to the independent work of Host and Kra [HK05, HK18] and Ziegler [Zie07]:

Theorem 15.1.1 (Structure theorem for finite order $\mathbb{Z}$-systems)**.** *Let (X, τ) be a concrete ergodic measure-preserving system of order k for some integer $k \geq 1$ over $\Gamma = \mathbb{Z}$. Then (X, τ) is isomorphic to an inverse limit of translational systems formed on k-nilpotent Lie groups.*

The Conze–Lesigne structure theorem (Theorem 14.2.1) is the only result that holds for arbitrary countable Γ, but it applies only to systems of order 2. Partial results for other specific choices of Γ and general finite-order systems can be found in, e.g., [BTZ10, Sha24, JST23, CS23]. Establishing a general structure theorem for arbitrary countable Γ and systems of arbitrary finite order is a very challenging but important open problem.

15.1.1 Application to Inverse Gowers Theory

An area where Host–Kra structure theory has found applications in recent years is the inverse theory for the Gowers norms, which forms the core of the emerging field of higher-order Fourier analysis (cf. [Tao12]). This field is particularly relevant for various problems in additive combinatorics and analytic number theory, such as obtaining quantitative bounds in Szemerédi's theorem or its polynomial and multidimensional generalizations (see, e.g., [PP24, LSS24]), understanding the asymptotics of primes in arithmetic progressions [GT10], and addressing questions related to Sarnak's conjecture [MRT+23]. A central open problem in the inverse theory for the Gowers norms is Conjecture 15.1.2 below.

Let G be a finite abelian group and $k \geq 1$. The **k-th Gowers uniformity norm** of a complex function $f \colon G \to \mathbb{C}$ is defined by the formula

$$\|f\|_{U^k(G)} \coloneqq (\mathbb{E}_{x,h_1,\ldots,h_k \in G} \Delta_{h_1} \ldots \Delta_{h_k} f(x))^{\frac{1}{2^k}},$$

where $\mathbb{E}_{x \in A} = \frac{1}{|A|} \sum_{x \in A}$ and $\Delta_h f(x) \coloneqq f(x+h) \cdot \overline{f}(x)$.

We have the following conjecture from [JT23c], where we refer to [JT23c] for the theory of polynomial maps on filtered groups.

Conjecture 15.1.2 (Inverse theorem for $U^{k+1}(G)$)**.** *Let G be a finite abelian group, $\delta > 0$, $k \geq 0$, and let $f\colon G \to \mathbb{C}$ be a 1-bounded function with $\|f\|_{U^{k+1}(G)} \geq \delta$. Then there is a degree k filtered nilmanifold H/Λ, drawn from some finite collection of such nilmanifolds that depends only on k, δ but not on G (and each such nilmanifold is endowed arbitrarily with a smooth Riemannian metric), a Lipschitz function $F\colon H/\Lambda \to \mathbb{C}$ of Lipschitz norm $O_{\delta,k}(1)$, and a polynomial map $g\colon G \to H/\Lambda$ such that*

$$|\mathbb{E}_{x\in G} f(x)\overline{F}(g(x))| \gg_{\delta,k} 1.$$

Conjecture 15.1.2 has been verified in several important special cases. The case where we restrict to the family of cyclic groups $\mathcal{F} = (\mathbb{Z}/p_n\mathbb{Z})_{n\to\infty}$, where p_n denotes the nth prime number, was established for arbitrary k in a breakthrough result by Green, Tao, and Ziegler [GTZ12]. The case $\mathcal{F} = (\mathbb{F}^n)_{n\to\infty}$, where $\mathbb{F}$ denotes a fixed finite field, was established for arbitrary k and $\mathbb{F}$ by Bergelson, Tao, and Ziegler in a series of papers [BTZ10, TZ10, TZ12], where they develop and use among other tools Host–Kra structure theory for $\mathbb{F}^\omega$-actions.

In [JT23c], Tao and the second author introduced an approach to Conjecture 15.1.2 that relates it to Host–Kra structure theory for $\mathbb{Z}^\omega$-systems. This approach can be outlined as follows: Let $\mathcal{F}$ be a family of finite abelian groups. In [JT23c], a correspondence principle was introduced that connects Conjecture 15.1.2 for $\mathcal{F}$ to the algebraic classification of characteristic factors of an ergodic $\mathbb{Z}^\omega$-system (X, T) modeled on the ultraproduct group of the family $\mathcal{F}$, where the order of the characteristic factor corresponds to the degree k of the Gowers norm. In the same work, the Conze–Lesigne structure theorem (Theorem 14.2.1) was used to prove Conjecture 15.1.2 in the special case $k = 2$. Conjecture 15.1.2 for $k \geq 3$ and arbitrary finite abelian groups is widely open. For some recent partial progress in certain special cases, see, e.g., [JST23, CGSS23].

15.2 A Glimpse at Topological Structure Theory

In our lectures we have developed the structure theory of measure-preserving systems and discussed some applications. A natural question is to ask for similar structure results in the world of topological dynamics, i.e., the study of continuous transformations on compact spaces. In this final section of our notes, let us briefly outline some of the topological structure theory.

For this exposition, we restrict (as usual) to systems (K, τ) given as actions of a countable (discrete) abelian group Γ on a compact metric space K. A **morphism**

$q\colon (K,\tau) \to (L,\sigma)$ between such topological dynamical systems is given by a continuous map $q\colon K \to L$ such that $q \circ \tau_\gamma = \sigma_\gamma \circ q$, i.e., the diagram

$$\begin{array}{ccc} K & \xrightarrow{\tau_\gamma} & K \\ {\scriptstyle q}\downarrow & & \downarrow{\scriptstyle q} \\ L & \xrightarrow{\sigma_\gamma} & L \end{array}$$

commutes for all $\gamma \in \Gamma$. It is

(i) a **factor map** or **extension** if q is surjective, and

(ii) an **isomorphism** if q is bijective (which implies that q^{-1} also defines a morphism of topological dynamical systems).

Structured Systems. What does it mean for such a topological dynamical system (K,τ) to be structured? One possible answer: We demand that there is no "sensitive dependence on initial conditions", i.e., if two points $x, y \in K$ are close to each other, then their orbit points $\tau_\gamma(x)$ and $\tau_\gamma(y)$ should also be close for all $\gamma \in \Gamma$. More precisely, given any $\varepsilon > 0$, we find some $\delta > 0$ such that the following implication holds.

- If $(x,y) \in K \times K$ satisfies $d(x,y) < \delta$, then $d(\tau_\gamma(x), \tau_\gamma(y)) < \varepsilon$ for all $\gamma \in \Gamma$.

This simply means that the family of maps $\{\tau_\gamma \mid \gamma \in \Gamma\} \subseteq \mathrm{Homeo}(K)$ is (uniformly) equicontinuous.[2]

Definition 15.2.1. A topological dynamical system (K,τ) is called **equicontinuous** if the family of maps $\{\tau_\gamma \mid \gamma \in \Gamma\}$ is equicontinuous.

The following rotation system is a typical example.

Example 15.2.2. For $\Gamma = \mathbb{Z}$ consider the rotation $(\mathbb{T}, l_a)$ on the torus defined by $l_a\colon \mathbb{T} \to \mathbb{T}$, $z \mapsto az$ for some $a \in \mathbb{T}$ (cf. Example 6.2.2). Then l_a preserves the Euclidean metric, which implies that the system $(\mathbb{T}, l_a)$ is equicontinuous.

Another idea would be to use a functional analytic definition of structured systems via discrete spectrum (see Definition 6.2.1). To do so, we can for each topological dynamical system (K,τ), just as for measure-preserving systems, consider the Koopman representation

$$U_\tau\colon \Gamma \to \mathscr{L}(\mathrm{C}(K)), \quad \gamma \mapsto U_{\tau_\gamma^{-1}}$$

where $U_{\tau_\gamma^{-1}} f := f \circ \tau_\gamma^{-1}$ for $f \in \mathrm{C}(K)$ and $\gamma \in \Gamma$. Since Γ is abelian, we then assume that the eigenspaces with respect to this representation generate $\mathrm{C}(K)$.

[2] Since K is compact, every equicontinuous family of maps is automatically uniformly equicontinuous.

Definition 15.2.3. A topological dynamical system (K, τ) **has discrete spectrum** if $\mathrm{C}(K) = \overline{\mathrm{lin}} \bigcup_{\chi \in \Gamma^*} \ker(\chi - U_\tau)$ where

$$\ker(\chi - U_\tau) := \{f \in \mathrm{C}(K) \mid U_{\tau_\gamma} f = \chi(\gamma) f \text{ for all } \gamma \in \Gamma\}$$

is the **eigenspace** associated with a character $\chi \in \Gamma^*$.

Once again we obtain rotation systems as examples.

Example 15.2.4. Let $c\colon \Gamma \to G$ be a group homomorphism to any compact abelian metric group G. Consider the induced rotation system system (G, τ_c) via $\tau_c\colon \Gamma \to \mathrm{Homeo}(G)$, $\gamma \mapsto l_{c(\gamma)}$ from Example 6.2.3 (where $l_x\colon G \to G$, $y \mapsto xy$). Using that the continuous characters $\chi \in G'$ span a dense linear subspace of $\mathrm{C}(G)$ by Proposition 6.1.19 and Theorem 6.1.18, one can check that the topological dynamical system (G, τ_c) has discrete spectrum.

It turns out that both previous definitions of structured topological dynamical systems – via equicontinuity and discrete spectrum – are equivalent. To show this, it is helpful to consider yet another characterization based on the following topological algebraic concept introduced by Robert Ellis (see [Ell60]). For a compact space K equip the set of all self-maps $K^K = \{\tau\colon K \to K\}$ with the product topology (i.e., the topology of pointwise convergence). By Tychonoff's theorem (see Theorem 3.2.4) this is a compact space.

Proposition and Definition 15.2.5. *For a topological dynamical system* (K, τ) *the closure*

$$\mathrm{E}(K, \tau) := \overline{\{\tau_\gamma \mid \gamma \in \Gamma\}} \subseteq K^K$$

in K^K *is a semigroup when equipped with the composition of maps, i.e.,* $\vartheta \circ \varrho \in \mathrm{E}(K, \tau)$ *for all* $\vartheta, \varrho \in \mathrm{E}(K, \tau)$. *We call* $\mathrm{E}(K, \tau)$ *the* ***Ellis semigroup*** *of* (K, τ).

We refer to [Aus88, Chapter 3] for a proof and more details, but highlight that in general $\mathrm{E}(K, \tau)$ may contain discontinuous maps. Furthermore, the multiplication (given by composition) on $\mathrm{E}(K, \tau)$ is generally not jointly continuous. However, $\mathrm{E}(K, \tau)$ is still a compact **right-topological semigroup**, i.e., for a fixed element $\varrho \in \mathrm{E}(K, \tau)$ the right multiplication

$$\mathrm{E}(K, \tau) \to \mathrm{E}(K, \tau), \quad \vartheta \mapsto \vartheta \circ \varrho$$

is continuous. One can apply the structure theory of such semigroups (see, e.g., [Rup84], [BJM89]) to the Ellis semigroup to study the topological dynamical system (K, τ).

We compute the Ellis semigroup in the situation of Example 15.2.2.

Example 15.2.6. Consider the torus rotation $(\mathbb{T}, l_a)$ for $a \in \mathbb{T}$. If $a^k = 1$ for some $k \in \mathbb{N}$, then $\{l_a^n \mid n \in \mathbb{Z}\} = \{l_{a^n} \mid n = 0, \ldots, k-1\} \subseteq \mathbb{T}^{\mathbb{T}}$ is discrete, hence closed, and we therefore have

$$\mathrm{E}(\mathbb{T}, l_a) = \{l_{a^n} \mid n = 0, \ldots, k-1\}.$$

On the other hand, if a is not a root of unity, then $\{a^n \mid n \in \mathbb{Z}\}$ is dense in $\mathbb{T}$ by Kronecker's Theorem (see Theorem 6.1.22), and this implies

$$\mathrm{E}(\mathbb{T}, l_a) = \{l_b \mid b \in \mathbb{T}\} \subseteq \mathbb{T}^{\mathbb{T}},$$

i.e., the Ellis semigroup consists of all rotations $l_b \colon \mathbb{T} \to \mathbb{T}$, $z \mapsto bz$ for $b \in \mathbb{T}$ in this case.

In this example the Ellis semigroup is actually a compact toplogical abelian group consisting of homeomorphisms. With the help of the Arzelà–Ascoli theorem (see, e.g., [Sin19, Theorem 11.3.12]), one can check that this is still the case for all equicontinuous systems (see again [Aus88, Chapter 3] for the details). The representation theory of compact groups from Section 6.1 (and some further arguments then yield the following satisfying characterizations of structured systems (see, e.g., [Aus88, Chapters 3 and 4] and [HK23, Theorem 1.11])[3].

Theorem 15.2.7. *For a topological dynamical system (K, τ) the following assertions are equivalent.*

(a) (K, τ) *is equicontinuous.*

(b) $\mathrm{E}(K, \tau)$ *is an abelian metrizable compact topological group of homeomorphisms.*

(c) (K, τ) *has discrete spectrum.*

As a corollary, we obtain a version of the Halmos–von Neumann representation theorem (see Theorem 6.2.6) for topological dynamical systems (see, e.g., [Aus88, Theorem 3.6], [Wal75, Theorem 5.8], [DNP87, Chapter VIII], and [HK23, Theorem 4.4 (iii)]). Recall the notion of minimal systems from Definition 10.2.1. One can readily check that every rotation system (G, τ_c) defined by a group homomorphism $c \colon \Gamma \to G$ with dense range to an abelian metrizable compact topological group G is minimal, and it has discrete spectrum by Example 15.2.4. We obtain the following converse result.

Theorem 15.2.8 (Topological Halmos–von Neumann Representation Theorem)**.** *Let (K, τ) be a minimal topological dynamical system with discrete spectrum. Then there is a group homomorphism $c \colon \Gamma \to G$ with dense range to a compact metric abelian group G such that (K, τ) is isomorphic to the rotation system (G, τ_c).*

[3] The quoted sources deal with the general case without countability and metrizability assumptions. However, in our setting we obtain metrizable groups via [Aus88, page 52] since the topology of uniform convergence on the space of homeomorphisms $\mathrm{Homeo}(K)$ is metrizable.

Proof. By Theorem 15.2.7 we obtain that $G := \mathrm{E}(K,\tau)$ is an abelian metrizable compact topological group of homeomorphisms. The map $c\colon \Gamma \to G$, $\gamma \mapsto \tau_\gamma$ is a group homomorphism with dense range. Now pick any point $x_0 \in K$. Then the evaluation map

$$q\colon G \to K, \quad \vartheta \mapsto \vartheta(x_0)$$

is continuous, and one can readily check that $q \circ (\tau_c)_\gamma = \tau_\gamma \circ q$ for each $\gamma \in \Gamma$, i.e., q is a morphism of topological dynamical systems.

Since $q(G)$ is a non-empty, compact and invariant subset of K, and (K,τ) is minimal, we obtain that q is surjective. To see that it is injective, take $\vartheta_1, \vartheta_2 \in G$ with $\vartheta_1(x_0) = \vartheta_2(x_0)$. Then $\vartheta := \vartheta_2^{-1} \circ \vartheta_1 \in G$ satisfies $\vartheta(x_0) = x_0$, and thus also

$$\vartheta(q(\varrho)) = \vartheta(\varrho(x_0)) = \varrho(\vartheta(x_0)) = \varrho(x_0) = q(\varrho)$$

since G is abelian. As q is surjective, this implies $\vartheta = \mathrm{id}_K$, hence $\vartheta_1 = \vartheta_2$. □

Remark 15.2.9. One can prove topological versions of all aspects of the Halmos–von Neumann theorem classification result (see, e.g., [Wal75, Theorems 5.8 and 5.9], [DNP87, Chapter VIII.3], and [HK23, Theorem 4.4]), even without any countability and metrizability assumptions. Via the concept of topological models discussed in the "Comments and Further Reading Section" of Lecture 8, these topological results can be used to give a different proof of the ergodic theoretic statements (Theorems 6.2.6, 6.2.7, and 6.2.8), see, e.g., [NW72], [DNP87, Chapter VIII] and [HK23, Section 3].

Furstenberg Structure Theory. Can we also rebuild a topological dynamical system (K,τ) from a tower of suitable extensions similarly to Theorem 9.1.12 for measure-preserving systems? The topological theory is somewhat more intricate. However, Furstenberg actually showed a topological structure theorem before his ergodic theoretic result (see [Fur63]). It deals with minimal systems which are *distal*.

Definition 15.2.10. A topological dynamical system (K,τ) is called **distal** if $\inf_{\gamma\in\Gamma} d(\tau_\gamma(x), \tau_\gamma(y)) > 0$ for all $x, y \in K$ with $x \neq y$.

Thus, loosely speaking, a system is distal if two distinct points cannot come close to each other "in the long run". One can check that every equicontinuous system is distal (see [Aus88, beginning of Chapter 5]), but the converse does not hold as shown by the following example.

Example 15.2.11. For $\Gamma = \mathbb{Z}$ and $a \in \mathbb{T}$, which is not a root of unity, consider the skew-rotation $(\mathbb{T}^2,\tau)$ from Example 7.2.6 given by $\tau\colon \mathbb{T}^2 \to \mathbb{T}^2$, $(x,y) \mapsto (ax, xy)$. Then $(\mathbb{T}^2,\tau)$ is a minimal distal system which is not equicontinuous, see (see [Aus88, page 75 of Chapter 5]).

Once again there is a nice characterization in terms of the Ellis semigroup (see [Aus88, Theorem 5.6]).

Theorem 15.2.12. *For a topological dynamical system* (K,τ) *the following assertions are equivalent.*

(a) (K,τ) *is distal.*

(b) $\mathrm{E}(K,\tau)$ *is a group.*

Furstenberg's structure theorem now tells that any minimal distal system can be rebuilt from a trivial system via "structured extensions". To make this concept precise, we introduce the following definition. Here, given a continuous surjection $q\colon K \to L$ between compact spaces, we write $K_l \coloneqq q^{-1}(\{l\})$ for the **fiber** over $l \in L$ and

$$K \times_L K \coloneqq \{(x,y) \in K^2 \mid q(x) = q(y)\} = \bigcup_{l\in L} K_l \times K_l \subseteq K \times K$$

for the **fiber product**. We then "relativize" the notion of equicontinuous systems from Definition 15.2.1 as follows.

Definition 15.2.13. An extension $q\colon (K,\tau) \to (L,\sigma)$ is called "equicontinuous" if for every $\varepsilon > 0$ there is $\delta > 0$ such that the following condition holds.

- Whenever $(x,y) \in K \times_L K$ satisfies $d(x,y) < \delta$, then $d(\tau_\gamma(x), \tau_\gamma(y)) < \varepsilon$ for all $\gamma \in \Gamma$.

Thus, we only have "no sensitive dependence on initial conditions" for points within the same fiber of q.

Example 15.2.14. For $a \in \mathbb{T}$ consider the skew-rotation system $(\mathbb{T}^2,\tau)$ from Example 15.2.11 and the rotation $(\mathbb{T}, l_a)$ from Example 6.2.2. Then the projection $\mathrm{pr}_1\colon \mathbb{T}^2 \to \mathbb{T}$, $(x,y) \mapsto x$ onto the first component defines an equicontinuous extension $\mathrm{pr}_1\colon (\mathbb{T}^2,\tau) \to (\mathbb{T}, l_a)$: If $(x,y),(x,z) \in \mathbb{T}^2$, we obtain for the Euclidean distance on $\mathbb{T}^2$ that

$$\begin{aligned}\|\tau(x,y) - \tau(x,z)\|_2 &= \|(ax,xy) - (ax,xz)\|_2 = |xy - xz| = |x|\cdot|y-z| \\ &= |y-z| = \|(x,y)-(x,z)\|_2.\end{aligned}$$

By induction, we then obtain $\|\tau^n(x,y) - \tau^n(x,z)\|_2 = \|(x,y)-(x,z)\|_2$ for all $n \in \mathbb{Z}$ and $x,y,z \in \mathbb{T}$. This implies that the extension is equicontinuous.

Remark 15.2.15. As in the case of systems, there is (at least for minimal systems) an equivalent functional analytic definition of structured extensions similar to the notion of "relative discrete spectrum" from Theorem 11.3.3, see [EK22b].

To formally state Furstenberg's structure theorem we also need the concept of projective limits for topological dynamical systems.

Definition 15.2.16. A **directed system of topological dynamical systems** $((K_\alpha, \tau_\alpha)_{\alpha\in A}, (q_{\alpha_1}^{\alpha_2})_{\alpha_1\le\alpha_2})$ consists of

(1) a non-empty directed set A,

(2) a topological dynamical system (K_α, τ_α) for each $\alpha \in A$, and

(3) an extension $q_{\alpha_1}^{\alpha_2}\colon (K_{\alpha_2}, \tau_{\alpha_2}) \to (K_{\alpha_1}, \tau_{\alpha_1})$ for all $\alpha_1, \alpha_2 \in A$ with $\alpha_1 \le \alpha_2$,

such that

(i) $q_{\alpha_1}^{\alpha_3} = q_{\alpha_1}^{\alpha_2} \circ q_{\alpha_2}^{\alpha_3}$ for all $\alpha_1, \alpha_2, \alpha_3 \in A$ with $\alpha_1 \le \alpha_2 \le \alpha_3$.

(ii) $q_\alpha^\alpha = \mathrm{id}_{K_\alpha}$ for each $\alpha \in A$.

In this case, the topological dynamical system (K, τ) given by the closed subset

$$K \coloneqq \left\{ (x_\alpha)_{\alpha\in A} \in \prod_{\alpha\in A} K_\alpha \mid q_{\alpha_1}^{\alpha_2}(x_{\alpha_2}) = x_{\alpha_1} \text{ for all } \alpha_1 \le \alpha_2 \right\}$$

of the product space $\prod_{\alpha\in A} K_\alpha$ equipped with the action given by $\tau_\gamma((x_\alpha)_{\alpha\in A}) \coloneqq ((\tau_\alpha)_\gamma(x_\alpha))_{\alpha\in A}$ for $(x_\alpha)_{\alpha\in A} \in K$ and $\gamma \in \Gamma$ is the **projective limit** of $((K_\alpha, \tau_\alpha))_{\alpha\in A}$.

One can show that the projective limit of minimal systems is again minimal. The following is now the precise formulation of the Furstenberg structure theorem for minimal distal systems. We refer to [Aus88, Chapter 7] for more details and a proof.

Theorem 15.2.17 (Furstenberg Structure Theorem)**.** *For a minimal topological dynamical system (K, τ) the following assertions are equivalent.*

(a) *(K, τ) is distal.*

(b) *There is an ordinal β and a projective system $((K_\alpha, \tau_\alpha)_{\alpha\le\beta}, (q_{\alpha_1}^{\alpha_2})_{\alpha_1\le\alpha_2})$ of minimal topologgical dynamical systems such that*

  (i) *$(K_\beta, \tau_\beta) = (K, \tau)$ and $(K_0, \tau_0) = (\{1\}, \mathrm{id})$,*

  (ii) *$q_{\alpha+1}\colon (K_{\alpha+1}, \tau_{\alpha+1}) \to (K_\alpha, \tau_\alpha)$ is equicontinuous for each $\alpha < \beta$, and*

  (iii) *(K_α, τ_α) is the projective limit of $((K_\alpha, \tau_\alpha)_{\alpha'<\alpha}, (q_{\alpha_1}^{\alpha_2})_{\alpha_1\le\alpha_2<\alpha'})$ for every limit ordinal $\alpha \le \beta$.*

A nice proof of the structure theorem can be obtained by studying the Ellis semigroup $\mathrm{E}(K, \tau)$ of a minimal distal system (K, τ), which is a compact right topological group by Theorem 15.2.12. In fact, Isaac Namioka proved a structure theorem for compact right topological groups, and then applied this to $\mathrm{E}(K, \tau)$ in order to obtain the Furstenberg structure theorem (see [Nam72] and [BJM89, Appendix C]).

Remarks 15.2.18. (i) One can also prove (more complicated) structure theorems for general minimal systems, see, e.g., [Vee77] and [dV93, Chapters V and VI].

(ii) As for systems one can use topological models (cf. Remark 15.2.9 above) to establish connections between structured extensions in topological dynamics and ergodic theory, see, e.g., [Ell87], [EJK23]. For $\Gamma = \mathbb{Z}$ Elon Lindenstrauss proved a relation between topologically distal systems and measurably distal systems (i.e., measure-preserving systems for which the weakly mixing extension in the Furstenberg–Zimmer tower can be omitted), see [Lin99]. Generally speaking, there is a close connection between the topological and the measure-theoretic structure theory.

Host–Kra Structure Theory. Finally, we mention that there is also a Host–Kra structure theory for topological dynamical systems (see, e.g., [HKM10], [HK18, Chapter 7 and Section 17.1]). However, we do not give the definition of the topological characteristic factors here (which again relies on certain constructions within cubic systems). As in the ergodic theoretic framework, one then says that a minimal system (K, τ) has **order** $k \in \mathbb{N}$ if it agrees with its kth characteristic factor. For $\Gamma = \mathbb{Z}$ one can then, as in the measure-preserving case, represent a system of order k as a projective limit of "k-step nilsystems".

We conclude our lectures with a nice characterization of order k systems in terms of the Ellis semigroup which has been established rather recently by Sebastián Donoso, Jiahao Qiu and Jianjie Zhao (see [Don14] and [QZ22]). For a compact right topological group E, set $E_0 \coloneqq E$ and recursively define $E_{i+1} \coloneqq \overline{[E_i, E]} \subseteq E$ for $i \in \mathbb{N}_0$. If E_k is the trivial group for $k \in \mathbb{N}_0$, then E is **k-step topologically nilpotent**.

Theorem 15.2.19. *For a minimal topological dynamical system (K, τ) over $\Gamma = \mathbb{Z}$ and $k \in \mathbb{N}$ the following assertions are equivalent.*

(a) (K, τ) *is of order k.*

(b) $\mathrm{E}(K, \tau)$ *is a k-step topologically nilpotent compact right topological group.*

Appendix A

Some Functional Analysis

In this appendix (which will be updated during the course) we treat basic concepts and results from functional analysis and operator theory to the extent needed for the lectures. Everyone who has already attended a course on functional analysis can skip this appendix (or use it as a reminder). Introductions to functional analysis (with the contents below and much more) can be found, e.g., in [Con85], [Rud87], [Ped89], and [Haa14].

A.1 Banach Spaces and Bounded Linear Operators

In this first part of the appendix we recall the concepts of Banach and Hilbert spaces, and bounded linear maps between them (see, e.g., [Con85, Paragraphs I.1, III.1, and III.3]), [Rud87, Paragraphs 4.1 and 5.1], [Ped89, Sections 2.1, 2.2, and 3.1], and [Haa14, Section I.1, Chapters 2, and 5]).

Normed and Banach Spaces. A **seminorm** on a complex vector space E is a map $\|\cdot\|\colon E \to [0, \infty)$ such that

(i) $\|f+g\| \le \|f\| + \|g\|$ for all $f, g \in E$, and

(ii) $\|\lambda f\| = |\lambda| \cdot \|f\|$ for $\lambda \in \mathbb{C}$ and $f \in E$.

A seminorm then automatically satisfies $\|0\| = 0$ for the zero vector $0 \in E$. It is a **norm** if addition $\|f\| = 0$ for $f \in E$ already implies $f = 0$. Every norm $\|\cdot\|$ induces a metric via $d(f,g) \coloneqq \|f - g\|$ for $f, g \in E$. In particular, in we can speak of topological notions like open sets, closed sets, the closure, the interior, convergent sequences, continuity, etc. Note that addition and scalar multiplication as well as the norm define continuous maps with respect to the above metric.

A complex vector space E equipped with a norm is called a **normed space**, and a **Banach space** if it is complete as a metric space, i.e., every Cauchy sequence converges. Classical examples also featuring in this course are the following.

Examples. (i) The vector space $\mathbb{C}^d$ for $d \in \mathbb{N}$ with the Euclidean norm given by $\|v\|_2 := (\sum_{i=1}^d |v_i|^2)^{\frac{1}{2}}$ for $v = (v_1, \dots, v_d) \in \mathbb{C}^d$ is a Banach space. In particular, $\mathbb{C}$ equipped with the modulus $|\cdot|$ is a Banach space.

(ii) The space $\mathrm{C}(K) := \{f\colon K \to \mathbb{C} \mid f \text{ continuous}\}$ for any compact space K (e.g., $K = [0,1]$) equipped with the **supremum norm** given by $\|f\|_\infty := \sup_{x\in K} |f(x)|$ for $f \in \mathrm{C}(K)$ is a Banach space.

(iii) For a probability space X, the space $\mathrm{L}^p(X)$ of equivalence classes of p-integrable measurable functions $f\colon X \to \mathbb{C}$ with the **p-norm** given by

$$\|f\|_p := \left(\int_X |f|^p\right)^{\frac{1}{p}} \qquad \text{for } f \in \mathrm{L}^p(X)$$

is a Banach space for each $p \in [1, \infty)$. Moreover, the space $\mathrm{L}^\infty(X)$ of equivalence classes of bounded measurable functions $f\colon X \to \mathbb{C}$ equipped with the **essential supremum norm** defined by

$$\|f\|_\infty = \inf\{c \geq 0 \mid |f(x)| \leq c \text{ for almost every } x \in X\} \qquad \text{for } f \in \mathrm{L}^\infty(X)$$

is a Banach space. Since X is a probability space, Hölder's inequality gives us the inclusions

$$\mathrm{L}^\infty(X) \subseteq \mathrm{L}^q(X) \subseteq \mathrm{L}^p(X) \subseteq \mathrm{L}^1(X)$$

with $\|f\|_p \leq \|f\|_q$ for $f \in \mathrm{L}^q(X)$ where $1 \leq p \leq q \leq \infty$.

Hilbert Spaces. In some of the examples above, the norm is defined by an inner product. Recall that on a complex vector space H a map $(\cdot|\cdot)\colon H \times H \to \mathbb{C}$ is a **positive sesquilinear form** if

(i) $(f|f) \geq 0$ for every $f \in H$,

(ii) $(\alpha_1 f_1 + \alpha_2 f_2|g) = \alpha_1(f_1|g) + \alpha_2(f_2|g)$ for all $\alpha_1, \alpha_2 \in \mathbb{C}$ and $f_1, f_2, g \in H$, and

(iii) $(f|\beta_1 g_1 + \beta_2 g_2) = \overline{\beta_1}(f|g_1) + \overline{\beta_2}(f|g_2)$ for all $\beta_1, \beta_2 \in \mathbb{C}$ and $f, g_1, g_2 \in H$.

In this case, the **Cauchy-Schwarz inquality**

$$|(f|g)| \leq (f|f)^{\frac{1}{2}} \cdot (g|g)^{\frac{1}{2}}$$

and the **polarization identity**

$$(f|g) = \frac{1}{4}\sum_{k=1}^{4} \mathrm{i}^k (f + \mathrm{i}^k g|f + \mathrm{i}^k g)$$

hold for all $f, g \in H$. The latter implies that $(\cdot|\cdot)$ is **self-adjoint**, i.e., $(g|f) = \overline{(f|g)}$ for all $f, g \in H$ (see, e.g., [Ped89, Section 3.1]).

One can check that $\|f\| := \sqrt{(f|f)}$ for $f \in H$ defines a seminorm on H. The following version of the **"Pythagorean theorem"** for $f, g \in H$ follows directly from the definition:

$$\|f + g\|^2 = \|f\|^2 + 2\,\mathrm{Re}\,(f|g) + \|g\|^2.$$

A positive sesquilinear form $(\cdot|\cdot)\colon H \times H \to \mathbb{C}$ is an **inner product** if $(f|f) = 0$ for $f \in H$ implies $f = 0$. In this case, the induced seminorm $\|\cdot\|$ is actually a norm, and, as a consequence of the Cauchy-Schwarz inequality, the inner product is continuous with respect to the induced metric. If the norm is complete, we call H a **Hilbert space**.

Examples. (i) $\mathbb{C}^d$ for $d \in \mathbb{N}$ with the Euclidean inner product given by $(v|w) := \sum_{i=1}^d v_i \overline{w_i}$ for $v = (v_1, \ldots, v_d), (w_1, \ldots, w_d) \in \mathbb{C}^d$ is a Hilbert space.

(ii) For a probability space X, the vector space $\mathrm{L}^2(X)$ equipped with the inner product defined by $(f|g) := \int_X f \cdot \overline{g}$ for $f, g \in \mathrm{L}^2(X)$ for $f, g \in \mathrm{L}^2(X)$ is a Hilbert space.

Bounded Linear Operators. In contrast to finite-dimensional vector spaces, linear maps between normed spaces need not be continuous. It turns out that a linear map $U\colon E \to F$ between normed spaces E and F is continuous precisely when it is a **bounded linear operator** in the sense that there is some $c \geq 0$ with $\|Uf\| \leq c \cdot \|f\|$ for all $f \in E$.

The space of all such bounded linear operators from E to F is denoted by $\mathscr{L}(E, F)$. Equipped with pointwise defined addition and scalar multiplication and the **operator norm** given by

$$\|U\| := \sup\{\|Uf\| \mid f \in E \text{ with } \|f\| \leq 1\} \text{ for } U \in \mathscr{L}(E, F),$$

the space $\mathscr{L}(E, F)$ itself becomes a normed space. It is a Banach space if F is a Banach space. The operator norm satisfies $\|Uf\| \leq \|U\| \cdot \|f\|$ for all $f \in E$ and $U \in \mathscr{L}(E, F)$. Moreover, it is submultiplicative: If D, E, F are normed spaces, $U \in \mathscr{L}(E, F)$ and $V \in \mathscr{L}(D, E)$, then $UV := U \circ V \in \mathscr{L}(D, F)$ with $\|UV\| \leq \|U\| \cdot \|V\|$. We abbreviate $\mathscr{L}(E) := \mathscr{L}(E, E)$ and $E' := \mathscr{L}(E, \mathbb{C})$ for any normed space E. The space E' is called the **dual space** of E.

Particularly nice bounded operators are the ones which actually preserve the norm: A linear map $U\colon E \to F$ between normed spaces E and F is a **linear isometry** if $\|Uf\| = \|f\|$ for all $f \in E$. Notice that this automatically implies that U is an injective map as it has a trivial kernel.

If H and K are Hilbert spaces, then a linear isometry $U\colon H \to K$ automatically satisfies $(Uf|Ug) = (f|g)$ for all $f, g \in H$ by the polarization identity. In this case, U is called a **unitary operator** if U is also surjective (and then a bijection). We write $\mathscr{U}(H)$ for the set of all unitary operators $U\colon H \to H$ on a Hilbert space H.

The following simple extension theorem (see, e.g., [Ped89, Proposition 2.1.11]) for bounded linear operators is used in the first lecture.

Proposition A.1.1. *Let E be a normed space and $D \subseteq E$ a dense linear subspace. Every bounded linear operator $U \in \mathscr{L}(D, F)$ to a Banach space F has a unique extension to a bounded linear operator $\overline{U} \in \mathscr{L}(E, F)$ with $\|\overline{U}\| = \|U\|$.*

Note that, by continuity of the norm, the extension of a linear isometry is again a linear isometry.

A.2 Basic Hilbert Space Theory

The following are some basic concepts and tools from Hilbert space theory, see, e.g., [Con85, Chapter 1], [Rud87, Chapter 4], [Ped89, Sections 3.1 and 3.2], and [Haa14, Chapters 1 and 8, and Section 12.2].

Given a Hilbert space H, two elements $f, g \in H$ are called **orthogonal** (in symbols: $f \perp g$) if $(f|g) = 0$. For orthogonal $f, g \in H$ the Pythagorean theorem from above becomes $\|f+g\|^2 = \|f\|^2 + \|g\|^2$. For a subset $M \subseteq H$ we call

$$M^\perp := \{f \in M \mid f \perp g \text{ for every } g \in M\}$$

the **orthogonal complement** of M. This is always a closed subspace of H. Two subsets $M, N \subseteq H$ are **orthogonal** if $f \perp g = 0$ for all $f \in M$ and $g \in N$, i.e., $M \subseteq N^\perp$.

The following geometric result is used in Chapter 3 (with $x = 0$) and is fundamental in Hilbert space theory (see, e.g., [Con85, Theorems I.2.5 and I.2.6]. Recall here that a subset $C \subseteq E$ of a vector space E is convex if $tf + (1-t)g \in C$ for all $f, g \in C$ and $t \in [0,1]$.

Theorem A.2.1. *Let $C \subseteq H$ be a non-empty, closed, convex subset of a Hilbert space H and $x \in H$.*

(i) *There is a unique $y_0 \in C$ with $\|y_0 - x\| = \inf\{\|y - x\| \mid y \in C\}$.*

(ii) *If $C \subseteq H$ is even a closed linear subspace, then y_0 of (i) is the unique element $z \in C$ with $x - z \in C^\perp$.*

The following result makes use of Theorem A.2.1, see, e.g., [Ped89, Theorem 3.1.7].

Theorem A.2.2. *Let $M \subseteq H$ be a closed linear subspace of a Hilbert space H. Then $H = M \oplus M^\perp$ is a decomposition of H into closed and orthogonal subspaces M and $M^\perp$.*

In the situation of Theorem A.2.2, the projection map $P_M\colon H \to H$ induced by this decomposition, i.e., the unique linear map sending elements $x \in M$ to itself, and elements $x \in M^\perp$ to 0, is called the **orthogonal projection onto** M.

By Theorem A.2.2 we obtain that for any subset $M \subseteq H$ of a Hilbert space H, we have the identity $(M^\perp)^\perp = \overline{\operatorname{lin}}\, M$ (see [Ped89, Corollary 3.1.8]). Thus, the linear hull $\operatorname{lin} M$ is dense in H precisely when $M^\perp = \{0\}$.

Another important consequence is the following famous representation theorem for the dual of a Hilbert space, see, e.g., [Con85, Theorem I.3.4].

Theorem A.2.3 (Riesz–Fréchet)**.** *Let H be a Hilbert space. Then the map*

$$H \to H', \quad g \mapsto \overline{g}$$

where $\overline{g}(f) := (f|g)$ *for all* $f, g \in H$ *is a bijection. Moreover,* $\|\overline{g}\| = \|g\|$ *for every* $g \in H$.

The Riesz–Fréchet theorem has an interesting consequence for bounded linear operators: If $U \in \mathscr{L}(H, K)$ for Hilbert spaces H and K, then for every $y \in K$ the map

$$\overline{y} \circ U : H \to \mathbb{C}, \quad x \mapsto (UX \mid Y)$$

is a bounded linear map by the Cauchy–Schwarz inequality, and hence there is a unique vector $z \in H$ with $\overline{y} \circ U = \overline{z}$. We set $U^*y := z$, i.e., U^*y is characterized by the identity $(x|U^*y) = (UX \mid Y)$ for all $x \in H$ and $y \in K$. The map $U^* : K \to H$ is called the **adjoint operator** of U, and is itself a bounded linear operator with $\|U^*\| = \|U\|$. Observe that $(U^*)^* = U$. Moreover, if H, K, L are Hilbert spaces, then $(UV)^* = V^*U^*$ holds for all $U \in \mathscr{L}(K, L)$ and $V \in \mathscr{L}(H, K)$.

One can check that a bounded linear operator $U \in \mathscr{L}(H, K)$ between Hilbert spaces H and K is

(i) a linear isometry precisely when $U^*U = \mathrm{Id}_H$.

(ii) a unitary operator precisely when $U^*U = \mathrm{Id}_H$ and $UU^* = \mathrm{Id}_K$.

A bounded linear operator $U \in \mathscr{L}(H)$ on a Hilbert space H is **self-adjoint** if $U^* = U$. A self-adjoint operator $P \in \mathscr{L}(H)$ is an **orthogonal projection** if in addition $P^2 = P$. For every closed subspace $M \subseteq H$ the orthogonal projection P_M onto M from above is indeed an orthogonal projection in this sense. Conversely, if P is an orthogonal projection, then the image $M := PH$ is a closed linear subspace and $P = P_M$. See [Ped89, Section 3.2] for these and further assertions.

Finally, we will also need the concept of orthonormal bases in some of the lectures. A subset $E \subseteq H$ of a Hilbert space H is **orthonormal** if its elements are pairwise orthogonal and $\|e\| = 1$ for every $e \in E$. An orthonormal subset $E \subseteq H$ which is maximal with respect to set inclusion, i.e., there is no strictly larger orthonormal subset $F \subsetneq H$, is called an **orthonormal basis** of H. An orthonormal subset $E \subseteq H$ is an orthonormal basis of H precisely when $\overline{\operatorname{lin}} E = H$. An application of Zorn's lemma yields the following (see [Ped89, Proposition 3.1.12]):

Theorem A.2.4. *Every orthonormal subset of a Hilbert space* H *is contained in an orthonormal basis of* H.

In particular, each Hilbert space has an orthonormal basis. This allows us to represent vectors in a Hilbert space via the following result, see, e.g., [Con85, Theorem I.4.13]. Note here that the finite subsets $\mathcal{P}_{\mathrm{fin}}(A)$ of any set A are directed by set inclusion, and so we can use the notion of net convergence from Chapter 3.

Theorem A.2.5 (Fourier Series/Parseval Identity). *Let $E \subseteq H$ be an orthonormal basis of a Hilbert space H. Then*

$$f = \sum_{e\in E} (f|e)e := \lim_{\substack{F\subseteq E \\ \textit{finite}}} \sum_{e\in F} (f|e)e \qquad \textit{and}$$

$$\|f\|^2 = \sum_{e\in E} |(f|e)|^2 := \lim_{\substack{F\subseteq E \\ \textit{finite}}} \sum_{e\in F} |(f|e)|^2 = \sup_{\substack{F\subseteq E \\ \textit{finite}}} \sum_{e\in F} |(f|e)|^2$$

for every $f \in H$.

By extending an orthonormal subset to an orthonormal basis we obtain the following corollary of Theorem A.2.5.

Corollary A.2.6 (Representation of Projections/Bessel Inequality). *Let $E \subseteq H$ be an orthonormal subset of a Hilbert space H, $M := \overline{\operatorname{lin}} E$ its closed linear hull and P_M the orthogonal projection onto M. Then*

$$P_M f = \sum_{e\in E} (f|e)f := \lim_{\substack{F\subseteq E \\ \textit{finite}}} \sum_{e\in F} (f|e)e \qquad \textit{and}$$

$$\|P_M f\|^2 = \sum_{e\in E} |(f|e)|^2 := \lim_{\substack{F\subseteq E \\ \textit{finite}}} \sum_{e\in F} |(f|e)|^2 = \sup_{\substack{F\subseteq E \\ \textit{finite}}} \sum_{e\in F} |(f|e)|^2 \le \|f\|^2$$

for every $f \in H$.

A.3 Spectral Projections

In this section, we summarize key results from the spectral theory of bounded self-adjoint operators and the closely related Borel functional calculus. For brevity, we will treat these results as a black box; proofs can be found in [Hal13, Chapter 8] for example.

Let $H \neq \{0\}$ be a Hilbert space and $T \in \mathscr{L}(H)$ be a bounded self-adjoint operator. For a real polynomial $p(t) = \sum_{i=0}^{d} c_i t^i \in \mathbb{R}[t]$, let p also denote its restriction as a function to $[-\|T\|, \|T\|]$. Define

$$p(T) = \sum_{i=0}^{d} c_i T^i.$$

Then $p(T)\colon H \to H$ is again a bounded and self-adjoint operator. Now, consider a compact interval $[a, b] \subseteq [-\|T\|, \|T\|]$, and let $f = \mathbb{1}_{[a,b]}$ denote its characteristic function. By the Bolzano–Weierstraß theorem, there exists a sequence of polynomials $p_n\colon [-\|T\|, \|T\|] \to \mathbb{R}$ such that p_n converges pointwise and boundedly to f.

We define

$$f(T)h \coloneqq \lim_{n\to\infty} p_n(T)h \quad \text{for } h \in H,$$

where this limit exists and is independent of the choice of the approximating sequence $(p_n)_{n\in\mathbb{N}}$.

The operator $f(T)$ has the following properties:

(i) $f(T)$ is an orthogonal projection.

(ii) $f(T) \circ T = T \circ f(T)$.

(iii) Let H_f be the range of $f(T)$. Then, for the restriction $T|_{H_f}$ of T to H_f, we have $a\,\mathrm{id}_{H_f} \leq T|_{H_f} \leq b\,\mathrm{id}_{H_f}$ in the sense that $a\|v\|^2 \leq (T|_{H_f} v \mid v) \leq b\|v\|^2$ for all $v \in H_f$.

We refer to $f(T)$ as a **spectral projection** of T.

Finally, by the continuity properties of the Borel functional calculus we have that

$$Th = \lim_{n\to\infty} (T\mathbb{1}_{[-\|T\|, -\frac{1}{n}]}(T)h + T\mathbb{1}_{[\frac{1}{n}, \|T\|]}(T)h)$$

for each $f \in H$.

Bibliography

[AB40] L. Alaoglu and G. Birkhoff. General ergodic theorems. *Ann. of Math.*, 41:293–309, 1940.

[AB06] Charalambos D. Aliprantis and Kim C. Border. *Infinite Dimensional Analysis. A Hitchhiker's Guide.* Springer, 3rd edition, 2006.

[AE05] H. Amann and J. Escher. *Analysis I.* Birkhäuser, 2005.

[Aus88] J. Auslander. *Minimal Flows and their Extensions.* Elsevier, 1988.

[Aus10] Tim Austin. Extensions of probability-preserving systems by measurably-varying homogeneous spaces and applications. *Fund. Math.*, 210(2):133–206, 2010.

[Bad06] M. Badino. The foundational role of ergodic theory. *Found. Sci.*, 11:323–347, 2006.

[Bar95] R.G. Bartle. *The Elements of Integration and Lebesgue Measure.* Wiley, 1995.

[BBP16] H. Bercovici, A. Brown, and C. Pearcy. *Measure and Integration.* Springer, 2016.

[Ber86] V. Bergelson. A density statement generalizing Schur's theorem. *J. Combin. Theory Ser. A*, 43:338—343, 1986.

[Ber04] V. Bergelson. Some historical comments and modern questions around the ergodic theorem. In *Dynamics of Complex Systems*, pages 1–11. Research Institute for Math. Sciences, Kyoto, 2004.

[Ber06] V. Bergelson. *Combinatorial and Diophantine Applications of Ergodic Theory*, chapter 12, pages 745–869. Elsevier, 2006. Appendix A by A. Leibman, and Appendix B by A. Quas and M. Wierdl.

[BGW24] V. Bergelson, E. Glasner, and B. Weiss. *The work of Hillel Furstenberg and its impact on modern mathematics*, pages 399–431. Springer, 2024.

[Bil95] P. Billingsley. *Probability and Measure.* Wiley, 3rd edition, 1995.

[Bir31] G. D. Birkhoff. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci.*

U.S.A., 17:656–660, 1931.

[BJM89] J. F. Berglund, H. Junghenn, and P. Milnes. *Analysis on Semigroups. Function Spaces, Compactifications, Representations*. Wiley, 1989.

[BK96] Howard Becker and Alexander S. Kechris. *The descriptive set theory of Polish group actions*, volume 232 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1996.

[Bro79] I. U. Bronšteĭn. *Extensions of Minimal Transformation Groups*. Sijthoff & Noordhoff, 1979.

[BS20] Thomas F. Bloom and Olof Sisask. Breaking the logarithmic barrier in Roth's theorem on arithmetic progressions. Preprint, arXiv:2007.03528 [math.NT] (2020), 2020.

[BTZ10] Vitaly Bergelson, Terence Tao, and Tamar Ziegler. An inverse theorem for the uniformity seminorms associated with the action of $\mathbb{F}_p^\infty$. *Geom. Funct. Anal.*, 19(6):1539–1596, 2010.

[Bó06] M. Bóna. *A Walk Through Combinatorics. An Introduction to Enumeration and Graph Theory*. Worlds Scientific, 2nd edition, 2006.

[CGSS23] Pablo Candela, Diego González-Sánchez, and Balázs Szegedy. Free nilspaces, double-coset nilspaces, and Gowers norms. Preprint, arXiv:2305.11233 [math.CO] (2023), 2023.

[CL84] Jean-Pierre Conze and Emmanuel Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France*, 112(2):143–175, 1984.

[CL88a] Jean-Pierre Conze and Emmanuel Lesigne. Sur un théorème ergodique pour des mesures diagonales. *C. R. Acad. Sci. Paris Sér. I Math.*, 306(12):491–493, 1988.

[CL88b] Jean-Pierre Conze and Emmanuel Lesigne. Sur un théorème ergodique pour des mesures diagonales. In *Probabilités*, volume 1987 of *Publ. Inst. Rech. Math. Rennes*, pages 1–31. Univ. Rennes I, Rennes, 1988.

[Con85] J.B. Conway. *A Course in Functional Analysis*. Springer, 1985.

[CS23] Pablo Candela and Balázs Szegedy. *Nilspace factors for general uniformity seminorms, cubic exchangeability and limits*, volume 1425 of *Mem. Am. Math. Soc.* Providence, RI: American Mathematical Society (AMS), 2023.

[CSD21] T. Ceccherini-Silberstein and M. D'Adderio. *Topics in Groups and Geometry Growth, Amenability, and Random Walks*. Springer, 2021.

[DE09] A. Deitmar and S. Echterhoff. *Principles of Harmonic Analysis*. Springer, 2009.

[dLG61] K. de Leeuw and I. Glickberg. Applications of almost periodic compactifications. *Acta Mathematica*, 105:63–97, 1961.

[DNP87] R. Derndinger, R. Nagel, and G. Palm. Ergodic Theory in the Perspective of Functional Analysis. Unpublished book manuscript, 1987.

[Don14] S. Donoso. Enveloping semigroups of systems of order d. *Discrete Contin. Dyn. Syst.*, 34:2729–2740, 2014.

[Dud02] R. M. Dudley. *Real analysis and probability.*, volume 74 of *Camb. Stud. Adv. Math.* Cambridge: Cambridge University Press, 2nd edition edition, 2002.

[dV93] J. de Vries. *Elements of Topological Dynamics.* Springer, 1993.

[EF19] T. Eisner and B. Farkas. 22nd Internet Seminar: Ergodic Theorems, 2018–2019.

[EFHN09] T. Eisner, B. Farkas, M. Haase, and R. Nagel. 12th Internet Seminar: Ergodic Theory – An Operator Theoretic Approach, 2008–2009.

[EFHN15] T. Eisner, B. Farkas, M. Haase, and R. Nagel. *Operator Theoretic Aspects of Ergodic Theory.* Springer, 2015.

[EHK24] N. Edeko, M. Haase, and H. Kreidler. A decomposition theorem for unitary group representations on Kaplansky-Hilbert modules and the Furstenberg-Zimmer structure theorem. *Anal. Math.*, 50:377–454, 2024.

[EJK23] Nikolai Edeko, Asgar Jamneshan, and Henrik Kreidler. A Peter-Weyl theorem for compact group bundles and the geometric representation of relatively ergodic compact extensions. *arXiv:2302.13630*, 2023.

[EK22a] Nikolai Edeko and Henrik Kreidler. Distal systems in topological dynamics and ergodic theory. *Ergodic Theory Dynam. Systems*, page 1–22, 2022.

[EK22b] Nikolai Edeko and Henrik Kreidler. Uniform enveloping semigroupoids for groupoid actions. *J. Anal. Math.*, 148(2):739–796, 2022.

[EKN21] N. Edeko, H. Kreidler, and R. Nagel. A dynamical proof of the van der corput inequality. arXiv:2106.11835, 2021.

[Ell60] R. Ellis. A semigroup associated with a transformation group. *Trans. Am. Math. Soc.*, 94:272–281, 1960.

[Ell87] R. Ellis. Topological dynamics and ergodic theory. *Ergodic Theory Dynam. Systems*, 7:25–47, 1987.

[ET36] P. Erdős and P. Turán. On some sequences of integers. *Journal of the London Mathematical Society*, s1-11:261–264, 1936.

[EW11] M. Einsiedler and T. Ward. *Ergodic Theory with a view towards Number*

Theory. Springer, 2011.

[FKO82] H. Furstenberg, Y. Katznelson, and D. Ornstein. The ergodic theoretical proof of Szemerédi's theorem. *Bull. Amer. Math. Soc.*, 7:527–552, 1982.

[Fol99] G. B. Folland. *Real Analysis. Modern Techniques and Their Applications.* John Wiley and Sons, 2nd edition, 1999.

[Fol15] G. B. Folland. *A Course in Abstract Harmonic Analysis.* CRC press, 2nd edition, 2015.

[Fur63] H. Furstenberg. The structure of distal flows. *Amer. J. Math.*, 85:477–515, 1963.

[Fur77] H. Furstenberg. Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Anal. Math.*, 31:204–256, 1977.

[Fur14] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory.* Princeton University Press, 2014.

[FW78] H. Furstenberg and B. Weiss. Topological dynamics and combinatorial number theory. *J. Anal. Math.*, 34:61–85, 1978.

[FW96a] 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 P. March V. Bergelson and J. Rosenblatt, editors, *Convergence in Ergodic Theory and Probability*, Ohio State University Mathematical Research Institute Publications. Walter de Gruyter, Berlin, New York, 1996.

[FW96b] Hillel Furstenberg and Benjamin Weiss. A mean ergodic theorem for $(1/N)\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$. In *Convergence in ergodic theory and probability (Columbus, OH, 1993)*, volume 5 of *Ohio State Univ. Math. Res. Inst. Publ.*, pages 193–227. de Gruyter, Berlin, 1996.

[GH55] W. H. Gottschalk and G. A. Hedlund. *Topological Dynamics.* American Mathematical Society, 1955.

[Gla03] E. Glasner. *Ergodic Theory via Joinings.* American Mathematical Society, 2003.

[Gow01] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[Gow13] W. Timothy Gowers. *Erdős and Arithmetic Progressions*, pages 265–287. Springer, 2013.

[GT08] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. Math. (2)*, 167:481–547, 2008.

[GT10] Benjamin Green and Terence Tao. Linear equations in primes. *Ann.*

Math. (2), 171(3):1753–1850, 2010.

[GTZ12] Ben Green, Terence Tao, and Tamar Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$-norm. *Ann. of Math. (2)*, 176(2):1231–1372, 2012.

[Haa14] M. Haase. *Functional Analysis: An Elementary Introduction.* American Mathematical Society, 2014.

[Hal56] P. Halmos. *Lectures on Ergodic Theory.* American Mathematical Society, 1956.

[Hal13] Brian C. Hall. *Quantum theory for mathematicians*, volume 267 of *Grad. Texts Math.* New York, NY: Springer, 2013.

[HK01] Bernard Host and Bryna Kra. Convergence of Conze-Lesigne averages. *Ergodic Theory Dynam. Systems*, 21(2):493–509, 2001.

[HK02] Bernard Host and Bryna Kra. An odd Furstenberg-Szemerédi theorem and quasi-affine systems. *J. Anal. Math.*, 86:183–220, 2002.

[HK05] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. Math.*, 161(1):397–488, 2005.

[HK18] Bernard Host and Bryna Kra. *Nilpotent structures in ergodic theory*, volume 236 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2018.

[HK23] P. Hermle and H. Kreidler. A Halmos–von Neumann theorem for actions of general groups. *Appl Categor Struct*, 31, 2023.

[HKM10] Bernard Host, Bryna Kra, and Alejandro Maass. Nilsequences and a structure theorem for topological dynamical systems. *Adv. Math.*, 224(1):103–129, 2010.

[HM20] K. H. Hofmann and S. A. Morris. *The Structure of Compact Groups. A Primer for the Student – A Handbook for the Expert.* De Gruyter, 2020.

[HR79] Edwin Hewitt and Kenneth A. Ross. *Abstract Harmonic Analysis. Vol. I*, volume 115. Springer, 2nd edition, 1979.

[HR02] Edwin Hewitt and Kenneth A. Ross. *Abstract Harmonic Analysis. Vol. II.* Springer, 2nd edition, 2002.

[HS65] E. Hewitt and K. Stromberg. *Real and Abstract Analysis.* Springer, 1965.

[Hun74] T. W. Hungerford. *Algebra.* Springer, 1974.

[HvN42] Paul R. Halmos and John von Neumann. Operator methods in classical mechanics. II. *Ann. of Math. (2)*, 43:332–350, 1942.

[Jac56] K. Jacobs. Ergodentheorie und fastperiodische Funktionen auf Halbgruppen. *Math. Z.*, 64:298–338, 1956.

[Jam03] G. J. O. Jameson. *The Prime Number Theorem.* Cambridge University Press, 2003.

[Jam23] Asgar Jamneshan. An uncountable Furstenberg-Zimmer structure theory. *Ergodic Theory Dyn. Syst.*, 43(7):2404–2436, 2023.

[JST23] Asgar Jamneshan, Or Shalom, and Terence Tao. The structure of totally disconnected Host–Kra–Ziegler factors, and the inverse theorem for the U^k Gowers uniformity norms on finite abelian groups of bounded torsion. Preprint, arXiv:2303.04860 [math.DS] (2023), 2023.

[JST24] Asgar Jamneshan, Or Shalom, and Terence Tao. The structure of arbitrary Conze-Lesigne systems. *Commun. Am. Math. Soc.*, 4:182–229, 2024.

[JT22] A. Jamneshan and T. Tao. An uncountable Mackey-Zimmer theorem. *Studia Math.*, 266(3):241–289, 2022.

[JT23a] A. Jamneshan and T. Tao. Foundational aspects of uncountable measure theory: Gelfand duality, Riesz representation, canonical models, and canonical disintegration. *Fund. Math.*, 261(1):1–98, 2023.

[JT23b] A. Jamneshan and T. Tao. An uncountable Moore-Schmidt theorem. *Ergodic Theory Dynam. Systems*, 43(7):2376–2403, 2023.

[JT23c] Asgar Jamneshan and Terence Tao. The inverse theorem for the U^3 Gowers uniformity norm on arbitrary finite abelian groups: Fourier-analytic and ergodic approaches. *Discrete Anal.*, pages Paper No. 11, 48, 2023.

[Jus22] K. Juschenko. *Amenability of Discrete Groups by Examples.* American Mathematical Society, 2022.

[Kal97] O. Kallenberg. *Foundations of Modern Probability.* Springer, 1997.

[KN74] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences.* Wiley, 1974.

[Koo31] B. O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proc. Natl. Acad. Sci. U.S.A.*, 18:315–318, 1931.

[Kre85] U. Krengel. *Ergodic Theorems.* de Gruyter, 1985.

[Kre18] H. Kreidler. Compact operator semigroups applied to dynamical systems. *Semigroup Forum*, 97:523–547, 2018.

[Lan02] S. Lang. *Algebra.* Springer, 3rd edition, 2002.

[Lin99] Elon Lindenstrauss. Measurable distal and topological distal systems. *Ergodic Theory Dynam. Systems*, 19(4):1063–1076, 1999.

[Lin01] Elon Lindenstrauss. Pointwise theorems for amenable groups. *Invent. Math.*, 146(2):259–295, 2001.

[LSS24] James Leng, Ashwin Sah, and Mehtaab Sawhney. Improved Bounds for Szemerédi's Theorem. Preprint, arXiv:2402.17995 [math.CO] (2024), 2024.

[Mac66] George W. Mackey. Ergodic theory and virtual groups. *Math. Ann.*, 166:187–207, 1966.

[Mei90] David Meiri. Generalized Correlation Sequences. Master's thesis, Tel-Aviv University, 1990.

[MRT$^+$23] Kaisa Matomäki, Maksym Radziwiłł, Terence Tao, Joni Teräväinen, and Tamar Ziegler. Higher uniformity of bounded multiplicative functions in short intervals on average. *Ann. Math. (2)*, 197(2):739–857, 2023.

[MS80] Calvin C. Moore and Klaus Schmidt. Coboundaries and Homomorphisms for Non–Singular Actions and a Problem of H. Helson. *Proc. Lond. Math. Soc.*, 40:443–475, 1980.

[MZ55] Deane Montgomery and Leo Zippin. *Topological transformation groups.* Interscience Publishers, New York-London, 1955.

[Nac65] Leopoldo Nachbin. *The Haar integral.* D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London, 1965.

[Nag73] R. Nagel. Mittelergodische Halbgruppen linearer Operatoren. *Ann. Inst. Fourier*, 23:75–87, 1973.

[Nam72] I. Namioka. Right topological groups, distal flows, and a fixed-point theorem. *Math. Systems Theory*, 6:193–209, 1972.

[NW72] R. Nagel and M. Wolff. Abstract dynamical systems with an application to operators with discrete spectrum. *Arch. Math.*, 23:170–176, 1972.

[Pat88] A. L. T. Paterson. *Amenability.* AMS, 1988.

[Ped89] G. K. Pedersen. *Analysis Now.* Springer, 1989.

[Poi90] H. Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta Math.*, 13:1–270, 1890.

[PP24] Sarah Peluse and Sean Prendiville. Quantitative bounds in the nonlinear Roth theorem. *Invent. Math.*, 238(3):865–903, 2024.

[QZ22] J. Qiu and J. Zhao. Top-nilpotent enveloping semigroups and pronilsystems. *J. Anal. Math.*, 148:595–636, 2022.

[Ran02] I. K. Rana. *An Introduction to Measure and Integration.* American Mathematical Society, 2nd edition, 2002.

[Rot53] K. F. Roth. On certain sets of integers. *Journal of the London Mathematical Society*, s1-28:104–109, 1953.

[Rud87] W. Rudin. *Real and Complex Analysis.* McGraw-Hill, 1987.

[Rud95] Daniel J. Rudolph. Eigenfunctions of $T \times S$ and the Conze-Lesigne algebra. In *Ergodic theory and its connections with harmonic analysis (Alexandria, 1993)*, volume 205 of *London Math. Soc. Lecture Note Ser.*, pages 369–432. Cambridge Univ. Press, Cambridge, 1995.

[Run02] V. Runde. *Lectures on Amenability.* Springer, 2002.

[Rup84] W. A. F. Ruppert. *Compact Semitopological Semigroups: An Instrinsic Theory.* Springer, 1984.

[Sat78] R. Sato. On abstract mean ergodic theorems. *Tôhoku Math. Journ.*, 38:575–581, 1978.

[Sch17] I. Schur. Über die Kongruenz $x^m \cong y^m$ (mod. p). *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 25:114–116, 1917.

[Sch13] M. Schreiber. Uniform families of ergodic operator nets. *Semigroup Forum*, 86:321–336, 2013.

[Sha24] Or Shalom. Host-Kra factors for $\bigoplus_{p\in P}\mathbb{Z}/p\mathbb{Z}$ actions and finite-dimensional nilpotent systems. *Anal. PDE*, 17(7):2379–2449, 2024.

[Sin19] T. B. Singh. *Introduction to Topology.* Springer, 2019.

[SV06] S. Shirali and H. L. Vasudeva. *Metric Spaces.* Springer, 2006.

[Sze75] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta. Arith.*, 27:199–245, 1975.

[Tao09] T. Tao. *Poincaré's Legacies: Pages from Year Two of a Mathematical Blog, Vol I.* American Mathematical Society, 2009.

[Tao11] T. Tao. *An Introduction to Measure Theory.* American Mathematical Society, 2011.

[Tao12] T. Tao. *Higher Order Fourier Analysis.* Graduate studies in mathematics. American Mathematical Society, 2012.

[Tao14a] T. Tao. An abstract ergodic theorem, and the Mackey-Zimmer theorem, 2014. https://terrytao.wordpress.com/2014/06/20/an-abstract-ergodic-theorem-and-the-mackey-zimmer-theorem/ (accessed 02/03/2023).

[Tao14b] Terence Tao. *Hilbert's fifth problem and related topics*, volume 153 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2014.

[TZ10] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields via the correspondence principle. *Anal. PDE*, 3(1):1–20, 2010.

[TZ12] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *Ann. Comb.*, 16(1):121–188, 2012.

[vdW27] B. L. van der Waerden. Beweis einer Baudetschen Vermutung. *Nieuw Archief*, 15:212–216, 1927.

[Vee77] W.A. Veech. Topological dynamics. *Bull. Amer. Math. Soc.*, 83:775–830, 1977.

[vN32a] J. von Neumann. Einige Sätze über messbare Abbildungen. *Ann. Math. (2)*, 33:574–586, 1932.

[vN32b] J. von Neumann. Proof of the quasi-ergodic hypothesis. *Proc. Natl. Acad. Sci. U.S.A.*, 18:70–82, 1932.

[VO16] M. Viana and K. Oliveira. *Foundations of Ergodic Theory.* Cambridge University Press, 2016.

[Wal75] P. Walters. *An Introduction to Ergodic Theory.* Springer, 1975.

[Wei80] J. Weidmann. *Linear Operators in Hilbert Spaces.* Springer, 1980.

[Wey16] H. Weyl. Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.*, 77:313–352, 1916.

[Zie07] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.*, 20:53–97, 2007.

[Zim76a] R. J. Zimmer. Ergodic actions with a generalized spectrum. *Illinois J. Math.*, 20:555–588, 1976.

[Zim76b] R. J. Zimmer. Extensions of ergodic group actions. *Illinois J. Math.*, 20:373–409, 1976.