Szemerédi's Theorem via Ergodic Theory

Yufei Zhao

April 20, 2011

Abstract

We provide an expository account of Furstenberg's ergodic theoretic proof of Szemerédi's theorem, which states that every subset of the integers with positive upper density contains arbitrarily long arithmetic progressions.

Contents

1 Introduction

1.1 van der Waerden and Szemerédi

In 1927, van der Waerden [vdW27] published a famous theorem regarding the existence of arithmetic progressions in any partition of the integers into finitely many parts.

Theorem 1.1 (van der Waerden)**.** *If we colour the integers with finitely many colours, then for any k there exists a monochromatic k-term arithmetic progression.*

A strengthening of this result was later conjectured by Erdős and Turán in 1936 [ET36]. They believed that the true reason for the existence of arithmetic progressions is that some colour class occupies positive density.

For a subset A of integers, we define its *upper density* to be

$$\overline{\delta}(A) := \limsup_{n\to\infty} \frac{1}{2n+1} \left| A \cap \{-n, -n+1, \ldots, n-1, n\} \right| .$$

We can similarly define the *lower density* of A by replacing the lim sup by lim inf. If it is clear that we are only working with nonnegative integers, then we should consider the limit of $\frac{1}{n}|A \cap \{0, 1, \ldots, n-1\}|$ instead.[1]

In 1953, Roth [Rot53] proved that any subset of the integers with positive upper density contains a 3-term arithmetic progression. In 1969, Endre Szemerédi [Sze69] proved that the subset must contain a 4-term arithmetic progression, and then in 1975 proved that the subset must contain arithmetic progressions of any length [Sze75]. Szemerédi's proof use an ingenious yet complicated combinatorial argument, applying what is now known as Szemerédi's regularity lemma.

Theorem 1.2 (Szemerédi)**.** *Let A be a subset of the integers with positive upper density, then A contains arbitrarily long arithmetic progressions.*

Many distinct proofs of Szemerédi's theorem have been given since Szemerédi's original proof. Here we list four notable approaches, each with a rich theory of its own.

1. The original combinatorial proof by Szemerédi [Sze75].
2. The ergodic theoretic proof by Furstenberg [Fur77].
3. The Fourier analytic proof by Gowers [Gow01].
4. The proof using hypergraph removal, independently by Nagle-Rödl-Schacht-Skokan [NRS06, RS04, RS06, RS07b, RS07a], Gowers [Gow07], and Tao [Tao06b, Tao07a].

Although these approaches are distinct, they share some common themes. For instance, every approach involves separating some system into pseudorandom components and structured components, and handling these two cases using different methods.

In this essay, we explain the second approach listed above, namely the ergodic theoretic proof by Furstenberg. Furstenberg's landmark paper [Fur77] connects combinatorial problems with ergodic theory. His work ignited the study of ergodic Ramsey theory, and it has led to many generalisations of Szemerédi's theorem, such as the multidimensional generalisation by Furstenberg and Katznelson [FK78] and the polynomial generalisation by Bergelson and Leibman [BL96] (see Section 3.4). The ergodic approach is the only known approach so far to some of the generalisations of Szemerédi's theorem.

The presentation of this essay is inspired by the following expository works: the AMS Bulletin article of Furstenberg, Katznelson, and Ornstein [FKO82]; the book by Furstenberg [Fur81]; and the lecture notes and blog posts of Tao, which are contained in his book [Tao09]. We provide the necessary background in ergodic theory in Section 2.

1.2 Idea of proof

Let us sketch the main ideas of the ergodic theoretic proof of Szemerédi's theorem. First we need to convert the problem about arithmetic progressions of integers into a problem about arithmetic progressions in dynamical systems. This is done via the correspondence principle, which is explained in detail in Section 3. Instead of working in $\mathbb{Z}$, which has the disadvantage of being non-compact, we work instead in some closed subset X of $\{0,1\}^{\mathbb{Z}}$. It is best to think abstractly about X, forgetting that it lies inside $\{0,1\}^{\mathbb{Z}}$. The space X comes equipped with some homeomorphism $T\colon X \to X$, making (X, T) a dynamical system. In this case T is the "shift" map, sending an element of $\{0,1\}^{\mathbb{Z}}$, represented as a sequence, to the sequence obtained by shifting sequence one position to the right. Again, we shall work abstractly and not be too concerned about the specific forms of X and T.

[1]Actually, Szemerédi's theorem only requires the subset to have positive *upper Banach density*, which is defined to be $\limsup_{n\to\infty} \sup_{N\in\mathbb{Z}} \frac{1}{n} |[N, N+n) \cap A|$. We shall keep things simple by only considering upper density, though it is easy to modify the arguments to deal with upper Banach density.

It turns out that the problem of finding an arithmetic progression in a subset $A \subset \mathbb{Z}$ is equivalent to the problem of finding an arithmetic progression in some subset of X. There is a T-invariant probability measure μ on X (meaning that $\mu(E) = \mu(T^n E)$ for all $n \in \mathbb{Z}$ and measurable set E), and some measurable set $E \subset X$ with $\mu(E) > 0$, such that the existence of a k-term arithmetic progression in A is equivalent to the existence of some $x \in X$ such that $x, T^n x, T^{2n} x, \ldots, T^{(k-1)n} x \in E$. This brings us to the domain of ergodic theory, which is the study of recurrence phenomena in these types of dynamical systems, known as measure preserving systems. In Section 2 we provide the necessary background in ergodic theory for the proof of Szemerédi's theorem.

In fact, we show that

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\left(E \cap T^n E \cap T^{2n} E \cap \cdots \cap T^{(k-1)n} E\right) > 0 \tag{1.1}$$

whenever $\mu(E) > 0$. This implies the previous claim about the existence of an arithmetic progression in E, as it implies that $E \cap T^n E \cap T^{2n} E \cap \cdots \cap T^{(k-1)n} E \neq \emptyset$ for some n.

There are two situations when (1.1) is rather easy to prove. They correspond to the "structured" and "pseudorandom" cases that we briefly mentioned in the introduction.

The structured scenario occurs when the subset E displays some near-periodicity. Indeed, if E is periodic, in the sense that $T^r E = E$ for some $r > 0$, then (1.1) is trivial since the summand is equal to $\mu(E)$ whenever n is a multiple of r. When E is not periodic, but "almost periodic", we can still deduce (1.1) in a very similar manner. When X displays almost periodicity for all subsets E, we say that X is a *compact* system. A typical example of a compact system is one where X is a circle and T is rotation by some fixed angle α on X. In general, a compact system can always be modelled by some compact abelian group, where the map T is translation by some fixed element. Systems that arise from compact abelian groups are known as *Kronecker systems*.

The pseudorandom scenario occurs when T displays some "mixing" phenomenon. Think of E as an event in the probability space X. Suppose that T mixes up the space in a random fashion, so that the events $E, T^1 E, T^2 E, \ldots$ are all independent, then (1.1) is again trivial, as each term is equal to $\mu(E)^k$. It turns out that we can establish (1.1) even if T only satisfies some weaker mixing properties, for instance, if E and $T^n E$ become nearly uncorrelated in some sense for large n. When X displays mixing properties for all subsets E, we say that X is a *weak mixing* system. A typical example of a weak mixing system is the Bernoulli system, where $X = \{0,1\}^{\mathbb{Z}}$, and T is the map that takes each $x \in X$, viewing x as a sequence, to the sequence obtained by shifting x one position to the right.

We shall prove that if X is either compact or weak mixing, then (1.1) always holds. The weak mixing case is analysed in Section 5 and the compact case is analysed in Section 6. Unfortunately, not every system is either weak mixing or compact, as these two cases represent the most extreme scenarios. We shall approach this problem by showing that if X is not completely pseudorandom, i.e., weak mixing, then we can always isolate some structured component K_1 in X, known as a *Kronecker factor*, which gives rise to a Kronecker system, i.e., a compact system. From our analysis on compact systems, we know that (1.1) is satisfied within the Kronecker factor. We are left with analysing what happens between X and K_1.

There is a natural map $X \to K_1$, which we call an *extension*. Just like how systems can display structured or pseudorandom behaviour, so can extensions, and we can generalise the notions of weak mixing and compact systems to weak mixing and compact extensions. If $X \to K_1$ is a weak mixing or compact extension, then the problem of proving (1.1) in X can be projected down to the problem in K_1, which we already know how to solve. Unfortunately, as with systems, not all extensions are weak mixing or compact. This is much akin to the dichotomy that we saw in systems, where we know how to handle systems which are either weak mixing or compact, but not all systems fall in one of the two extremes. If $X \to K_1$ is not a weak mixing extension, then we can find an intermediate factor K_2, giving us a composition of extensions $X \to K_2 \to K_1$, where $K_2 \to K_1$ is a compact extension. As mentioned earlier, the compactness of $K_2 \to K_1$ means that the once we've reduced the problem from X to K_2, we know how to pass it down to K_1, which we know how to handle. So this reduces the problem to $X \to K_2$, i.e., showing that if (1.1) holds in K_2 then it must also hold in X. We can continue this process, building a tower of extensions $X \to K_\alpha \to \cdots \to K_3 \to K_2 \to K_1$, where each $K_{n+1} \to K_n$ is a compact extension. The top extension $X \to K_\alpha$ is weak mixing, for otherwise we can keep growing the tower of extensions from K_α. (We are simplifying things somewhat here, since α may be an ordinal.) Finally we lift the property (1.1) all the way to the top of the tower of extensions, showing that it holds in X.

A key step in the above procedure is in showing that if $X \to Y$ is a weak mixing or compact extension, then knowing (1.1) for Y implies (1.1) for X. The proofs are analogous to the proofs of (1.1) for weak mixing and compact systems. Weak mixing extensions are discussed in Section 9 and compact extensions are discussed in Section 10. We shall put everything together in Section 11 where we analyse the tower of extensions that arise from the process.

In fact, it is not actually necessary to prove that (1.1) holds for weak mixing systems and compact systems, since they follow from the more general results on extensions, as a system is equivalent to an extension of the system over the one point trivial system. However, it is more instructive to present the results for weak mixing systems and compact systems first, since they are technically less complex than their versions for extensions. The relevant concepts for extensions are formed by "relativising" the corresponding concepts for systems, and the proof for extensions are obtained by modifying the proofs for systems. This is the approach taken in all three expositions of the ergodic theoretic proof of Szemerédi's theorem [FKO82, Fur81, Tao09]. We shall highlight the similarity between the approach for weak mixing/compact systems and the approach for weak mixing/compact extensions by presenting their proofs in a parallel manner. One might get a déjà vu feeling while reading the sections on weak mixing and compact extensions after having read the corresponding sections on weak mixing and compact systems. This was done on purpose.

2 Ergodic theory

In this section, we provide a basic introduction to ergodic theory. Ergodic theory is the study of measure preserving dynamical systems. We are interested in the behaviour of spaces under repeated applications of a certain transformation. For instance, will a point always come back to somewhere close to its starting point if we apply the transformation enough times?

Ergodic theory is a deep and well-established theory that we will not be able to cover in great depth in this essay. There are many textbooks on the subject, such as [EW11, Wal82]. We will only present enough background to discuss the proof of Szemerédi's theorem. We will explain the connection to Szemerédi theorem in Section 3.

2.1 Dynamical systems

A *dynamical system* is pair (X, T), where X is a set and $T : X \to X$ is a map, sometimes referred to as the *shift map*, which we will always assume to be invertible in this essay. For any nonnegative integer n, we can define the iterated map $T^n : X \to X$, given by $T^n x = T(T(\cdots(T(x))\cdots))$, where T is repeated n times. Since we are assuming that T is invertible, we can also define T^{-1} as the inverse of T, and we can extend T^n to all integers n via $T^n = (T^{-1})^{-n}$ for $n < 0$. To simplify notation, we will often refer to a system (X, T) by its underlying set X. Intuitively, we are studying the evolution of the system X as it is being transformed by T over time.

Example 2.1. Here are some examples of dynamical systems.

(a) (Finite systems) X is a finite set and T is a permutation of X.

(b) (Group actions) If a group G acts on set X, then every g defines a dynamical system (X, T_g) where T_g is given by $T_g x = gx$.

(c) (Group rotations) If G is a group and $a \in G$, then $(G, x \mapsto ax)$ is a dynamical system.

(d) (Bernoulli systems) Let Ω be a set, and let $\Omega^{\mathbb{Z}}$ be the set of all integer-indexed Ω-valued sequences. Let T denote the right shift operator, sending the sequence $(x_n)_{n\in\mathbb{Z}}$ to the sequence $(x_{n-1})_{n\in\mathbb{Z}}$. Then $(\Omega^{\mathbb{Z}}, T)$ is a dynamical system. This system will play a very important role for us later on in relating ergodic theory to combinatorics.

(e) (Binary Bernoulli systems) As a special case of the above example, let $2^{\mathbb{Z}}$ denote the collection of all subsets of $\mathbb{Z}$. Then T sends every $A \subset \mathbb{Z}$ to $A + 1 = \{a + 1 : a \in A\}$.

So far we have viewed T^n as a map on the points of X. We can abuse notation by letting it act on subsets $E \subset X$ as

$$T^n E := \{T^n x : x \in E\}.$$

If f is a function on X, then define $T^n f$ as

$$T^n f(x) := f(T^{-n}x).$$

Later when we have a measure μ on X, we define $T^n\mu$ to be the measure given by

$$T^n\mu(E) := \mu(T^{-n}E).$$

The sign conventions are chosen so that we have the identities

$$T^n 1_E = 1_{T^n E} \quad \text{and} \quad T^n \delta_x = \delta_{T^n x},$$

where 1_E is the indicator function on of $E \subset X$ and δ_x is the measure defined by $\delta_x(E) = 1$ if $x \in E$ and 0 if $x \notin E$.

So far, without additional structure on dynamical systems, there is very little to say about them. We know that T, being invertible, is a permutation of X, so that X can be decomposed into orbits of T, each of which is either a finite cycle $(\mathbb{Z}/m\mathbb{Z}, x \mapsto x + 1)$ or an infinite path $(\mathbb{Z}/m\mathbb{Z}, x \mapsto x + 1)$. In order to obtain more interesting results, we shall impose additional structure on X. We consider the following two types of structured dynamical systems, with emphasis on the second type.

- **Topological dynamical systems** (X, T), where X is a compact metric space and $T : X \to X$ is a homeomorphism. The study of these systems is known as topological dynamics.

- **Measure preserving systems** $(X, \mathcal{X}, \mu, T)$, where $(X, \mathcal{X}, \mu)$ is a probability space, with X a compact metric space, $\mathcal{X}$ the σ-algebra of measurable sets and μ the probability measure. Here T is a probability space isomorphism, i.e., both T and T^{-1} are measurable and measure preserving (i.e., $\mu(T^n E) = \mu(E)$ for all $E \in \mathcal{X}$ and $n \in \mathbb{Z}$). Ergodic theory is the study of these systems.

In the proof of Szemerédi's theorem, we will be working almost exclusively with measure preserving systems. We will often just call them *systems*. The measure plays an important role, and intuitively we should only care about sets up to measure zero. Equality in functions will general refer to equality almost everywhere (a.e.). So for instance, when we talk about constant functions, we generally mean functions that are constant almost everywhere.

Example 2.2. Examples of measure preserving systems. Each of these example is also a topological dynamical system, and the measurable sets can be chosen to be the Borel σ-algebra.

(a) (Finite systems) X is a finite set under the discrete topology, μ is the uniform measure, and T is a permutation of X.

(b) (Circle rotation) X is the unit circle, represented as $\mathbb{R}/\mathbb{Z}$ with the standard Lebesgue measure, and $Tx = x + a \pmod 1$ with some fixed a.

(c) (Kronecker system) Let G be a compact metrisable group, and $a \in G$. Then we can take $X = G$, μ the Haar measure, and $Tx = ax$.

(d) (Bernoulli systems) Take the Bernoulli system from Example 2.1(d), and assume that Ω is a finite set. Use the product measure on $X = \Omega^{\mathbb{Z}}$. Then X is compact by Tychonoff's theorem, and it is metrisable since it can be given the metric $d(x, y) = \frac{1}{|m|+1}$ for $x \neq y \in \Omega^{\mathbb{Z}}$ where m is the index of the least absolute value such that $x_m \neq y_m$.

Although the discussion of topological dynamical systems is not strictly necessarily for proving Szemerédi's theorem, it is helpful for illustrative purposes. In the next section, we look at some results about topological dynamical systems before moving to measure preserving systems.

2.2 Basic recurrence theorems

In this section we give some results about recurrence in dynamical systems. Roughly speaking, these results guarantee that a dynamical system will always return to a state close to its starting state after some time. We start with topological dynamical systems.

Definition 2.3. Let (X, T) be a topological dynamical system. Say that $x \in X$ is *recurrent* if there exists a sequence $n_j \to \infty$ such that $T^{n_j} x \to x$.

Theorem 2.4 (Birkhoff recurrence theorem)**.** *Every topological dynamical system (X, T) contains some recurrent point.*

Note that the compactness assumption on X plays a crucial role, since the system $(\mathbb{R}, x \mapsto x + 1)$ clearly does not have the above recurrence property.

In order to prove the Birkhoff recurrence theorem, we need to introduce the notion of a minimal system.

Definition 2.5. Let (X, T) be a topological dynamical system. A *subsystem* is a topological system of the form (Y, T), where Y is a closed T-invariant (meaning $TY = Y$) subset of X, and the action of T is now restricted to Y. We say that a topological dynamical system (X, T) is *minimal* if $X \neq \emptyset$ and its only subsystems are $\emptyset$ and X.

Example 2.6. Consider the circle rotation system $X = (\mathbb{R}/\mathbb{Z}, x \mapsto x + \alpha)$. If α is rational, then the orbit of a single point is a discrete subset of $\mathbb{R}/\mathbb{Z}$, and is therefore a subsystem of X. On the other hand, if α is irrational, then the orbit of a point is always dense in X, so X is minimal.

Example 2.7. For any (X, T) and any $x \in X$, the closure of the orbit of x, $\overline{\{T^n x : n \in \mathbb{Z}\}}$, is always a subsystem of X.

The existence of minimal subsystems is a consequence of Zorn's lemma.

Lemma 2.8. *Let (X, T) be a topological dynamical system. Then (X, T) contains some minimal subsystem.*

Proof. Consider the collection of all nonempty T-invariant subsystems $\{Y_\alpha\}$, with a partial order given by inclusion. Every chain $\{Y_\beta\}_{\beta\in B}$ has a lower bound given by $Y' = \bigcap_{\beta\in B} Y_\beta$. Indeed, Y' is nonempty since $\{Y_\beta\}_{\beta\in B}$ is a chain of closed subsets of a compact space satisfying the finite intersection property, and Y' is T-invariant because every Y_β is. So every chain of T-invariant subsystems has a lower bound. By Zorn's lemma, we know that there exists a minimal subsystem. $\square$

It turns out that in a minimal subsystem, every point is recurrent.

Lemma 2.9. *Let (X,T) be a minimal topological dynamical system. Then every point of X is recurrent.*

Proof. Let $x \in X$. Since X is compact, the forward orbit $\{T^n x : n \geq 0\}$ has a limit point y, so there exists a sequence $n_i \to \infty$ such that $T^{n_i}x \to y$. If $y = x$ then we are done. Otherwise, since $\overline{\{T^n y : n \in \mathbb{Z}\}}$ is a subsystem of X, which is minimal, we must have $X = \overline{\{T^n y : n \in \mathbb{Z}\}}$. This implies that there exists some sequence n'_j (not necessarily going to infinity) such that $T^{n'_j}y \to x$. Then $T^{n_i+n'_j}x \to x$ as $i,j \to \infty$, this shows that x is recurrent as long we as let $i,j \to \infty$ in a way so that $n_i + n'_j \to \infty$, which is always possible as $n_i \to \infty$. $\square$

Proof of Birkhoff recurrence theorem (Theorem 2.4). The result follows from Lemmas 2.8 and 2.9, as we can choose any x in a minimal subsystem of X. $\square$

Having gotten a taste of a recurrence theorem in the topological dynamical setting, we now move to a basic result about recurrence in measure preserving systems.

Theorem 2.10 (Poincaré Recurrence Theorem)**.** *Let $(X,\mathcal{X},\mu,T)$ be a measure preserving system. Let $E \in \mathcal{X}$ with $\mu(E) > 0$. Let $E' \subset E$ be the set of $x \in E$ for which $\{n \in \mathbb{N} : T^n x \in E\}$ is infinite. Then $\mu(E \setminus E') = 0$. i.e., almost all points in E recur back to E.*

Proof. Let $A_N = \bigcup_{n\geq N} T^{-n}E$. Note that $A_0 \supset A_1 \supset A_2 \supset \cdots$, and $S = \bigcap_{N>0} A_N$ is the set of all points in X that enter E infinitely often, so that $S \cap E = E'$. Note that $A_i = T^{-1}A_{i-1}$, so by the invariance of the measure, we have $\mu(A_0) = \mu(A_1) = \cdots$. Using Monotone Convergence theorem on the sequence of characteristic functions $\{1_{X\setminus A_i}\}$, we find that $\mu(S) = \mu(A_0)$ and hence $\mu(A_0\setminus S) = 0$. Note that $E \subset A_0$, so $E\setminus E' = E\cap(A_0\setminus S)$, hence $\mu(E\setminus E') \leq \mu(A_0\setminus S) = 0$, as desired. $\square$

Poincaré recurrence theorem is also implied by the apparently stronger result that in a measure preserving system X, whenever $\mu(E) > 0$, we have

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

Indeed, if $\mu(E \setminus E') > 0$, then applying (2.1) to $E \setminus E'$ shows that some element of $E \setminus E'$ must recur, which is impossible. Note that (2.1) is the $k = 2$ case of (1.1), which is what we will ultimately prove. In fact, (2.1) follows easily from the mean ergodic theorem, which we look at next.

2.3 Ergodic theorems

In the previous section, we discussed the notion of minimal topological dynamical systems, where the only T-invariant closed subsets are the empty set and the whole space. In the world of measure preserving systems, as we only care about sets up to measure zero, we have the following analogous notion of minimality.

Definition 2.11 (Ergodicity)**.** We say that a measure preserving system $(X, \mathcal{X}, \mu, T)$ is *ergodic* if all T-invariant sets (i.e., sets $E \in \mathcal{X}$ with $TE = E$) have measure 0 or 1.

It will be useful to consider the action of T on measurable functions in addition to measurable sets. An equivalent definition for X being ergodic is that every T-invariant function, i.e., $Tf = f$, is constant almost everywhere.

Next we discuss a few "ergodic theorems". These are theorems stating that in an ergodic system, the "time averages"

$$\mathbb{E}_{0 \le n < N}(T^n f) := \frac{1}{N} \sum_{n=0}^{N-1} T^n f$$

converge to the "space average"

$$\mathbb{E}(f) := \mathbb{E}_X(f) := \int_X f \, d\mu.$$

There are different versions of ergodic theorems, corresponding to difference types of convergence of functions.

Definition 2.12 (Mode of convergence)**.** Let $(X, \mathcal{X}, \mu)$ be a compact measure space. Let $f_1, f_2, \ldots$ and f be real-valued measurable functions on X. As $n \to \infty$, we say that

(i) $f_n \to f$ *weakly* in $L^2(X)$ if $\langle f_n - f, g \rangle \to 0$ for any $g \in L^2(X)$.

(ii) $f_n \to f$ in $L^2(X)$ if $\|f_n - f\|_2 \to 0$.

(iii) $f_n \to f$ *pointwise almost everywhere* if $f_n(x) \to f(x)$ for all x outside a set of measure 0.

Remark. In order to simplify notation, in this essay, our functions are assumed to be real valued and our Hilbert spaces are assumed to be over the real numbers, unless otherwise specified. It is easy to adapt to the complex setting.

Weak convergence implies L^2 convergence by Cauchy-Schwarz inequality, as $|\langle f_n - f, g \rangle| \le \|f_n - f\|_2 \|g\|_2$. Pointwise convergence implies L^2 convergence by dominated convergence theorem if the sequence of functions is assumed to be uniformly bounded.

To state the L^2 ergodic theorem, it will be helpful to use the language of conditional expectations. Let $\mathcal{X}'$ is a sub-σ-algebra of $\mathcal{X}$. Then $L^2(X, \mathcal{X}', \mu)$ is contained as a subspace in $L^2(X, \mathcal{X}, \mu)$, and the conditional expectation can be given as the orthogonal projection map:

$$\mathbf{E}(\cdot | \mathcal{Y}) : L^2(X, \mathcal{X}, \mu) \to L^2(X, \mathcal{X}', \mu).$$

It is possible to extend this definition to L^1 functions, although we do not need it here. Let

$$\mathcal{X}^T = \{E \in \mathcal{X} : TE = E\}$$

be the maximal T-invariant sub-σ-algebra of $\mathcal{X}$. Since $f \in L^2(X, \mathcal{X}, \mu)$ is T-invariant if and only if f is $\mathcal{X}^T$-measurable, we see that $L^2(X, \mathcal{X}^T, \mu)$ is the set of T-invariant functions in $L^2(X, \mathcal{X}, \mu)$. In particular, X being ergodic is equivalent to $\mathcal{X}^T$ containing only sets of measure 0 and 1, or equivalently $L^2(X, \mathcal{X}^T, \mu)$ consisting of only constant a.e. functions. When X is ergodic, $\mathbf{E}\left(f | \mathcal{X}^T\right) = \mathbf{E}(f)$, the usual expectation in X.

Theorem 2.13 (von Neumann mean ergodic theorem)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system. Let $f \in L^2(X, \mathcal{X}, \mu)$. Then*

$$\mathbf{E}_{0 \le n < N}(T^n f) \to \mathbf{E}\left(f | \mathcal{X}^T\right)$$

in $L^2(X, \mathcal{X}, \mu)$ as $N \to \infty$. In particular, if X is ergodic, then the limit equals to $\mathbf{E}(f)$.

Since convergence in $L^2(X)$ is stronger than weak convergence, the mean ergodic mean implies the following weaker ergodic theorem.

Corollary 2.14 (Weak ergodic theorem). *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system. For any $f \in L^2(X, \mathcal{X}, \mu)$, we have $\mathbf{E}_{0\le n<N}(T^n f) \to \mathbf{E}\left(f \middle| \mathcal{X}^T\right)$ weakly in $L^2(X, \mathcal{X}, \mu)$ as $N \to \infty$.*

The mean ergodic theorem is sufficient to prove Szemerédi's theorem, though most of the time we only need to apply the weak ergodic theorem.

In fact, it turns out that the convergence $\mathbf{E}_{0\le n<N}(T^n f) \to \mathbf{E}\left(f \middle| \mathcal{X}^T\right)$ even holds pointwise. We shall state the result here without proof, and we will not use the pointwise ergodic theorem in the proof of Szemerédi's theorem.

Theorem 2.15 (Birkhoff pointwise ergodic theorem). *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic measure preserving system. For any $f \in L^1(X, \mathcal{X}, \mu)$, we have $\mathbf{E}_{0\le n<N}(T^n f) \to \mathbf{E}\left(f \middle| \mathcal{X}^T\right)$ pointwise almost everywhere as $N \to \infty$.*

Example 2.16. In a finite system, where T acts by permutation on a finite set X with uniform measure, the ergodic theorems say that given any function $f : X \to \mathbb{R}$, the average of f under T, i.e., $\mathbf{E}_{0\le n<N}(T^n f)$, converges to the function found by averaging f over each cycle of T. In particular, when T is transitive, so that it consists of one cycle and the system is ergodic, the limit is just the constant function on X with value being the mean of f.

In the rest of this section, we prove von Neumann's mean ergodic theorem. We shall work in a Hilbert space setting.

Theorem 2.17 (Mean ergodic theorem for Hilbert spaces). *Let H be a Hilbert space. Let $U : H \to H$ be a unitary operator. Let $H^U = \{v \in H : Uv = v\}$ denote the subspace of U-invariant vectors. Let $P : H \to H^U$ be the orthogonal projection onto H^U. Then for every $v \in H$, $\mathbf{E}_{0\le n<N}(U^n v) \to Pv$ in H as $N \to \infty$.*

Proof. If $v \in H^U$, then $\mathbf{E}_{0\le n<N}(U^n v) = v$ for all N and the result is clear. The idea is to approximate the orthogonal complement of H^U using the (possibly non-closed) space $W = \{Uw - w : w \in H\}$. We indeed have $W \subset (H^U)^\perp$, since for any $w \in H$ and $v \in H^U$

$$\langle Uw - w, v\rangle = \langle Uw, v\rangle - \langle w, v\rangle = \left\langle w, U^{-1}v\right\rangle - \langle w, v\rangle = \langle w, v\rangle - \langle w, v\rangle = 0.$$

Since P is an orthogonal projection, $P(Uw - w) = 0$. By telescoping, we have

$$\mathbf{E}_{0\le n<N}(U^n(Uw - w)) = \frac{1}{N}\sum_{n=0}^{N-1} U^n(Uw - w) = \frac{1}{N}(U^N w - w).$$

So $\left\|\mathbf{E}_{0\le n<N}(U^n(Uw - w))\right\| \le \frac{2}{N}\|w\| \to 0$. This shows that $\mathbf{E}_{0\le n<N}(U^n v) \to 0 = Pv$ for all $v \in W$.

By linearity $\mathbf{E}_{0\le n<N}(U^n v) \to Pv$ holds for all $v \in H^U + W$, and a limiting argument shows that it holds for all v in the closure $\overline{H^U + W}$.

Finally, we claim that $H = \overline{H^U + W}$. If not, then there exists some nonzero $v \in H$ orthogonal to the subspace $\overline{H^U + W}$, and

$$\begin{aligned}\|Uv - v\| &= \langle Uv - v, Uv - v\rangle = \langle Uv, Uv\rangle - \langle Uv, v\rangle - \langle v, Uv\rangle + \langle v, v\rangle \\ &= 2\langle v, v\rangle - 2\langle Uv, v\rangle = 2\langle v - Uv, v\rangle = 0.\end{aligned}$$

Hence $Uv = v$, and thus $v \in H^U$, a contradiction as v was supposed to be nonzero vector orthogonal to H^U. Therefore, $H = \overline{H^U + W}$, and hence $\mathbf{E}_{0\le n<N}(U^n v) \to Pv$ holds for all $v \in H$. $\square$

Proof of Theorem 2.13. Note that the operator $T : L^2(X, \mathcal{X}, \mu) \to L^2(X, \mathcal{X}, \mu)$, sending f to Tf, is unitary, as T is measure preserving, so that

$$\langle Tf, Tf \rangle = \int_X f(T^{-1}x)^2 \, d\mu(x) = \int_X f(x)^2 \, d\mu(x) = \langle f, f \rangle .$$

Theorem 2.13 follows from Theorem 2.17 by considering the Hilbert space $L^2(X, \mathcal{X}, \mu)$ and the unitary operator T. $\square$

We can prove Poincaré recurrence (Theorem 2.10) using the mean ergodic theorem following the remarks given at the end of the previous subsection. Indeed, note that $\mu(E \cap T^n E) = \int_X 1_E \; T^n 1_E \; d\mu = \langle 1_E, T^n 1_E \rangle$. By the weak ergodic theorem, $\frac{1}{N} \sum_{n=0}^{N-1} T^n 1_E$ converges to $\mathbf{E}\left(1_E \middle| \mathcal{X}^T\right)$ weakly in $L^2(X)$, so

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\left(E \cap T^n E\right) &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle 1_E, T^n 1_E \rangle = \left\langle 1_E, \mathbf{E}\left(1_E \middle| \mathcal{X}^T\right) \right\rangle \\ &= \left\langle \mathbf{E}\left(1_E \middle| \mathcal{X}^T\right), \mathbf{E}\left(1_E \middle| \mathcal{X}^T\right) \right\rangle = \left\| \mathbf{E}\left(1_E \middle| \mathcal{X}^T\right) \right\|_2^2 > 0. \end{aligned}$$

The last step follows from noting that $\mathbf{E}\left(1_E \middle| \mathcal{X}^T\right)$ is a nonnegative function on X with mean $\mathbf{E}(1_E) = \mu(E) > 0$.

2.4 Ergodic decomposition

Ergodicity is a useful property. For instance, ergodic theorems have very simple forms for ergodic systems. It would be nice if we can somehow decompose an arbitrary system into ergodic components, so that we can focus our study on ergodic systems. This is much akin to studying irreducible representations in representation theory once we know that every representation can be written as a direct sum of irreducible representations.

In the setting of dynamical systems without any additional structure, as discussed at the start of Section 2.1, it is very easy to decompose an arbitrary system into minimal components, which are precisely the orbits. However, in the setting of topological dynamical systems, it is no longer possible to do the same, as the following example shows.

Example 2.18. Consider the one-point compactification $\mathbb{Z} \cup \{\infty\}$ of $\mathbb{Z}$ under the discrete topology. The closed sets are any set containing ∞ or any finite set. Consider the shift map T given by $n \mapsto n+1$ and $\infty \mapsto \infty$. Then $\{\infty\}$ is a closed T-invariant set, so that it is a minimal subsystem. However, the closure of the orbit of any other point is the whole space. Therefore, it is not possible to decompose this topological dynamical system into a union of minimal subsystems.

The situation is more promising in the setting of measure preserving systems. Ergodicity is a property of the measure, and not just that of the the underlying measurable space. It turns out that we should decompose the measure into a sum of ergodic measures, as opposed of breaking the space up into different pieces. As a motivating example, let us consider the finite case.

Example 2.19. Let $X = \{1, 2, \ldots, 6\}$ equipped with the uniform measure μ. Let T denote the permutation $(2\ 3)(4\ 5\ 6)$. Then $(X, \mathcal{X}, \mu, T)$ is not ergodic, since the set $E = \{1\}$ satisfies $TE = E$ and we have $\mu(E) = \frac{1}{6} \notin \{0, 1\}$. Now, consider the following measures:

$$\mu_1(\{x\}) = \begin{cases} 1 & \text{if } x = 1; \\ 0 & \text{otherwise.} \end{cases} \qquad \mu_2(\{x\}) = \begin{cases} \frac{1}{2} & \text{if } x \in \{2, 3\}; \\ 0 & \text{otherwise.} \end{cases} \qquad \mu_3(\{x\}) = \begin{cases} \frac{1}{3} & \text{if } x \in \{4, 5, 6\}; \\ 0 & \text{otherwise.} \end{cases}$$

By changing the measure, the system $(X, \mathcal{X}, \mu_i, T)$ is now ergodic for each $i = 1, 2, 3$. Indeed, the invariant sets, which are some union of $\{1\}$, $\{2,3\}$, $\{4,5,6\}$ all have measure 0 or 1 under each μ_i. Observe that $\mu = \frac{1}{6}\mu_1 + \frac{1}{3}\mu_2 + \frac{1}{2}\mu_3$, so we have decomposed μ as a weighted average of ergodic measures. This is an example of ergodic decomposition.

In general, a system might not admit a decomposition into finitely many, or even countably many ergodic measures. We might have to integrate over a family of ergodic measures.

Theorem 2.20 (Ergodic decomposition)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system. Let $\mathcal{E}(X)$ denote the set of ergodic measures on X. Then there exists a map $\beta : X \to \mathcal{E}(X)$ such that for any $A \in \mathcal{X}$, the map $x \mapsto \beta_x(A) : X \to [0,1]$ is measurable, and*

$$\mu(A) = \int_X \beta_x(A)\ d\mu(x).$$

In Example 2.19, β can be chosen to map $\beta_1 = \mu_1$, $\beta_2 = \beta_3 = \mu_2$, $\beta_4 = \beta_5 = \beta_6 = \mu_3$.

The theorem follows from Choquets theory which concerns writing point inside a convex body as in terms of the extreme points of the body. See [Phe01]. It can also be deduce from results on disintegration of measures which will be discussed later in Section 8.4.

3 Correspondence principle

In this section, we explain how combinatorial results such as Szemerédi's theorem are related to results in dynamical systems. We shall formulate a statement about measure preserving systems that is equivalent to Szemerédi's theorem.

3.1 Bernoulli systems

Which dynamical systems should we consider? One obvious guess is $(\mathbb{Z}, x \mapsto x+1)$. However, this space is non-compact, and furthermore it does not admit a shift-invariant probability measure, so that our tools in ergodic theory cannot be applied. It turns out that the correct dynamical system to consider is the Bernoulli system $(2^{\mathbb{Z}}, B \mapsto B+1)$, where the points are subsets of $\mathbb{Z}$, and the shift map sends each $B \subset \mathbb{Z}$ to $B + 1 = \{b+1 : b \in \Omega\}$. Alternatively, we may view $2^{\mathbb{Z}}$ as the set of $\mathbb{Z}$-indexed $\{0,1\}$-valued sequences, so that the shift map shifts each sequence to the right by one position, sending the sequence $(x_n)_{n\in\mathbb{Z}}$ to $(x_{n-1})_{n\in\mathbb{Z}}$. Equip $2^{\mathbb{Z}}$ with the product topology, where each component $\{0,1\}$ has the discrete topology. The space is compact due to Tychonoff's theorem.

More generally, we can consider the system $(\Omega^{\mathbb{Z}}, T)$, where Ω is some finite set equipped with the discrete topology, $\Omega^{\mathbb{Z}}$ is the set of all Ω-valued sequences $(x_n)_{n\in\mathbb{Z}}$, and T shifts a sequence to the right by one position, i.e., $T(x_n)_{n\in\mathbb{Z}} = (x_{n-1})_{n\in\mathbb{Z}}$.

To illustrate how the Bernoulli system could be used to prove combinatorial results, let us first look at some examples in the topological dynamical setting. These examples are meant for illustrative purposes. They are simpler versions of the argument that we use in the next subsection to relate Szemerédi's theorem to ergodic theory.

Theorem 3.1 (Simple recurrence in open covers)**.** *Let (X, T) be a topological dynamical system and let $(U_\alpha)_{\alpha\in\Omega}$ be an open cover of X. Then there exists an open set U_α in this cover such that $U_\alpha \cap T^n U_\alpha \neq \emptyset$ for infinitely many n.*

We'll argue that Theorem 3.1 is equivalent to the *infinite pigeonhole principle* on $\mathbb{Z}$, which states that any colouring of $\mathbb{Z}$ into finitely colours always contains a colour with infinitely many elements. Although the infinite pigeonhole principle is a rather trivial result, it bares some semblance to Szemerédi's theorem, so it is interesting to see how we can use Theorem 3.1, a result about topological dynamical systems, to deduce the infinite pigeonhole principle.

Proof that the infinite pigeonhole principle on $\mathbb{Z}$ implies Theorem 3.1. Since X is compact, we may assume that $(U_\alpha)_{\alpha\in\Omega}$ is a finite open cover. Pick $x \in X$ arbitrarily. Then by the infinite pigeonhole principle, one of the open sets U_α must contain $T^n x$ for infinitely many $n \in \mathbb{Z}$. Let $S = \{n \in \mathbb{Z} : T^n x \in U_\alpha\}$. Pick $n_0 \in S$ arbitrarily, then $T^{n_0}x \in U_\alpha$ and $T^n x \in U_\alpha$ for all $n \in S$, so that $T^{n_0}x \in U_\alpha \cap T^{n-n_0}U_\alpha$ for all $n \in S$. Hence S has infinitely many elements. □

Proof that Theorem 3.1 implies infinite pigeonhole principle on $\mathbb{Z}$. Let Ω be the set of colours used. Consider the Bernoulli system $(\Omega^{\mathbb{Z}}, T)$, where T is the right shift operator. Every colouring of $\mathbb{Z}$ corresponds to a point in $\Omega^{\mathbb{Z}}$. Suppose that our colouring corresponds to $c = (c_n)_{n\in\mathbb{Z}} \in \Omega^{\mathbb{Z}}$. Consider the subsystem $X = \overline{\{T^k c : k \in \mathbb{Z}\}}$, the closure of the orbit of c. Since X is a closed subset of a compact set, it is also compact. For each $\alpha \in \Omega$, let U_α denote the subset of X consisting of all sequences whose 0-th term is α. Then $\{U_\alpha\}_{\alpha\in\Omega}$ is an open cover of X. Applying Theorem 3.1, there is an α so that $U_\alpha \cap T^n U_\alpha \neq \emptyset$ for infinitely many n. We claim that the colour α occurs infinitely many times in c. Indeed, $T^n U_\alpha$ is precisely the set of points in X whose n-th term is α. Since $U_\alpha \cap T^n U_\alpha$ is open, it contains some $T^k c$ whenever it is nonempty, so that $c_{-k} = c_{n-k} = \alpha$, and hence there are two elements of $\mathbb{Z}$ differing by n both coloured α. Since this is true for infinitely many n, it follows that that infinitely many elements of $\mathbb{Z}$ are coloured α. □

This example demonstrates how the Bernoulli system could be used to reduce problems in combinatorics to problems in dynamical systems. Of course, this is only a toy example for illustrative purposes. The technique of considering subsystems of Bernoulli systems allows us to prove more difficult combinatorial results.

Next, we show that when the simple recurrence in Theorem 3.1 is extended to so-called multiple recurrence, the corresponding statement in combinatorics is van der Waerden's theorem (Theorem 1.1), which states that if the integers $\mathbb{Z}$ are finitely coloured, then one the colour classes contains arbitrarily long arithmetic progressions. This is already a non-trivial result. We argue that van der Waerden's theorem is equivalent to the following result about topological dynamical systems.

Theorem 3.2 (Multiple recurrence in open covers)**.** *Let (X,T) be a topological dynamical system, and let $(U_\alpha)_{\alpha\in\Omega}$ be an open over of X. Then there exists U_α such that for every $k \geq 1$, we have $U_\alpha \cap T^n U_\alpha \cap \cdots \cap T^{(k-1)n}U_\alpha \neq \emptyset$ for some $n > 0$.*

Proposition 3.3. *Theorem 3.2 is equivalent to van der Waerden's theorem.*

Proof that van der Waerden's theorem implies Theorem 3.2. Since X is compact, we may assume that $(U_\alpha)_{\alpha\in\Omega}$ is a finite cover. Pick any $x \in X$. Assign every integer $n \in \mathbb{Z}$ some colour $\alpha \in \Omega$ so that$T^n x \in U_\alpha$. By van der Waerden's theorem, some colour class α contains a k-term arithmetic progression, say $\{a-(k-1)n, a-(k-2)n, \ldots, a\}$, so that we have $T^{a-in}x \in U_\alpha$ for $0 \leq i < k$. It follows that $T^a x \in U_\alpha \cap T^n U_\alpha \cap \cdots \cap T^{(k-1)n}U_\alpha$. □

Proof that Theorem 3.2 implies van der Waerden's theorem. As in the previous proof, suppose that we have a finite colouring of the integers corresponding to some $c = (c_n)_{n\in\mathbb{Z}} \in \Omega^{\mathbb{Z}}$, where Ω is the finite set of colours. Consider the closure of its orbit: $X = \overline{\{T^n c : n \in \mathbb{Z}\}}$. This is a closed subset of the compact space $\Omega^{\mathbb{Z}}$, so it is closed. For each $\alpha \in \Omega$, let U_α denote the subset of X consisting of elements whose 0-th term is α, so that $\{U_\alpha\}_{\alpha\in\Omega}$ is an open cover of X. Then by Theorem 3.2, we have $U_\alpha \cap T^n U_\alpha \cap \cdots \cap T^{(k-1)n}U_\alpha \neq \emptyset$ for some $n > 0$. Note that $U_\alpha \cap T^n U_\alpha \cap \cdots \cap T^{(k-1)n}U_\alpha$ is open, so it contains $T^{-m}c$ for some $m \in \mathbb{Z}$, so $T^{-m}c \in T^{in}U_\alpha$, thus $c \in T^{m+in}U_\alpha$, which means that $c_{m+in} = \alpha$, and this is true for $i = 0, 1, \ldots, k-1$. Therefore $(m+in)_{0\leq i\leq k-1}$ is a monochromatic k-term arithmetic progression. This establishes van der Waerden's theorem. □

We won't prove van der Waerden's theorem in this essay. Its proof can be found in many places, e.g., [GRS90]. In [Tao07b], Tao compares two proofs of van der Waerden's theorem, one in the combinatorial setting and one in the topological dynamical setting. The core ideas of the two proofs are the same, involving what is known as the "colour focusing argument", although the language differs. Tao argues that the dynamical argument is conceptually cleaner than the combinatorial argument once the machinery has been set up, although it has the disadvantage of not immediately giving any quantitative bounds (see Section 3.3).

3.2 Furstenberg multiple recurrence

In the previous subsection, we saw how to relate recurrence in topological dynamical systems to combinatorial results. Now, we add measures to the mix and develop the connection between ergodic theory and combinatorics. In particular, we reduce Szemerédi's theorem to the following result about measure preserving systems.

Theorem 3.4 (Furstenberg multiple recurrence theorem [Fur77])**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system, and k a positive integer. Then for any $E \in \mathcal{X}$ with $\mu(E) > 0$ there exists some $n > 0$ such that*

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

Note the similarity between Theorems 3.2 and 3.4, the former in the setting of topological dynamical systems and the latter in the setting of measure preserving systems. We show that Theorem 3.4 for k is equivalent to Szemerédi's theorem (Theorem 4.2) for k-term arithmetic progressions.

Lemma 3.5. *Let $(X, \mathcal{X}, \mu, T)$ be a system, and let $E \in \mathcal{X}$ with $\mu(E) > 0$. Then there exists a set F with $\mu(F) > 0$ such that $\{n \in \mathbb{Z} : T^n x \in E\}$ has positive upper density for all $x \in F$.*

Proof. For each positive integer N, let $\delta_N : X \to [0, 1]$ be given by

$$\delta_N(x) = \frac{|\{n \in \mathbb{Z} : -N \leq n \leq N, T^n x \in E\}|}{2N + 1}.$$

Note that $\mathbf{E}(\delta_N) = \mu(E)$ for all N, since δ_N is the average of $T^n 1_E$ over $-N \leq n \leq N$, and each such function has expected value $\mu(E)$. Let

$$A_N = \left\{ x \in X : \delta_N(x) \geq \frac{1}{2}\mu(E) \right\}.$$

Then, as $\mu(A_N) \geq \frac{1}{2}\mu(E)$,

$$\mu(E) = \mathbf{E}(\delta_N) \leq \mu(A_N) \cdot 1 + (1 - \mu(A)) \cdot \frac{1}{2}\mu(E) \leq \mu(A_N) + \frac{1}{2}\mu(E).$$

The set $\limsup_{N\to\infty} A_N = \bigcap_{N\geq 1} \bigcup_{n \geq N} A_n$ contains precisely the points $x \in X$ that appear in A_n infinitely often, i.e., the set of x such that the upper density of $\{n \in \mathbb{Z} : T^n x \in E\}$ is at least $\frac{1}{2}\mu(E)$. Since $\mu(A_n) \geq \frac{1}{2}\mu(E)$ for all n, it follows that $\mu\left(\bigcup_{n\geq N} A_n\right) \geq \frac{1}{2}\mu(E)$ for all N, and since $\bigcup_{n\geq N} A_n$ is a decreasing sequence of sets in N, we obtain that

$$\mu\left(\limsup_{N\to\infty} A_N\right) = \mu\left(\bigcap_{N\geq 1} \bigcup_{n\geq N} A_n\right) \geq \frac{1}{2}\mu(E) > 0.$$

So we can choose $F = \limsup_{N\to\infty} A_N$. $\square$

Proof that Szemerédi's theorem implies Theorem 3.4. Let AP_k denote the set of all k-term arithmetic progressions in $\mathbb{Z}$. For each arithmetic progression $a = (a_1, \ldots, a_k) \in A_k$, let B_a denote the set of points $x \in X$ such that $T^{a_i}x \in E$ for each $i = 1, 2, \ldots, k$. Then $B_k = \bigcup_{a \in AP_k} B_a$ is the set of all points $x \in X$ such that $\{n \in \mathbb{Z} : T^n x \in E\}$ contains some arithmetic progression.

Let F as in Lemma 3.5 be a set of positive measure such that $\{n \in \mathbb{Z} : T^n x \in E\}$ has positive upper density for all $x \in F$. By Szemerédi's theorem, $\{n \in \mathbb{Z} : T^n x \in E\}$ contains a k-term arithmetic progression for all $x \in F$, so that $F \subset B_k$, and hence $\mu(B_k) \geq \mu(F) > 0$. Since AP_k is countable, it follows that $\mu(B_a) > 0$ for some $a \in AP_k$. Thus $T^b B_a \subset E \cap T^n E \cap \cdots \cap T^{(k-1)n} E$ for some $n > 0$ and $b \in \mathbb{Z}$, and this set has positive measure. $\square$

Proof that Theorem 3.4 implies Szemerédi's theorem. Let $A \subset \mathbb{Z}$ have positive upper density, represented as a point a in the Bernoulli system $(2^{\mathbb{Z}}, T)$, where T corresponds to the map $B \mapsto B+1$. Let $X = \overline{\{T^n a : n \in \mathbb{Z}\}}$ and $E = \{b \in X : 0 \in b\}$. If we can find a T-invariance measure μ on X such that $\mu(E) > 0$, then Theorem 3.4 would imply that $\mu(E \cap T^n E \cap \cdots \cap T^{(k-1)n} E) > 0$ for some n, so the intersection contains $T^{-m} a$ for some $m \in \mathbb{Z}$. It follows that the k-term arithmetic progression $(m + in)_{0 \leq i < k}$ is contained in A. It remains to prove that existence of a T-invariance measure μ.

Let μ_N denote the measure on X given by

$$\mu_N = \frac{1}{2N+1} \sum_{n=-N}^{N} \delta_{T^n a},$$

where δ_b is the point mass at $b \in X$. Note that $\mu_N(E)$ is density of A in $\{-N, -N+1, \ldots, N\}$. Since A has positive upper density, we can pick a sequence $N_j \to \infty$ such that $\mu_{N_j}(E)$ approaches some positive limit as $j \to \infty$. Let μ be some weak limit point of μ_{N_j}, which exists by sequential compactness via the Banach-Alaoglu theorem. Then $\mu(E) > 0$. Although each μ_N is not T-invariant, it is nearly so, as

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

which has its total mass bounded by $\frac{2}{2N+1} \to 0$. Since μ is a limit point, it must be T-invariant. So we produced a T-invariance measure μ on X such that $\mu(E) > 0$. $\square$

Table 1 summarises the equivalences that we can demonstrated so far between results in combinatorics and results in dynamical systems. In Section 3.4 we state some additional combinatorial results and their ergodic theoretic formulations.

Table 1: Equivalence via the correspondence principle

Combinatorial result	Dynamical system result
Infinite pigeonhole principle	Simple recurrence in open covers (Thm. 3.1)
van der Waerden's theorem (Thm. 1.1)	Multiple recurrence in open covers (Thm. 3.2)
Szemerédi's theorem (Thm. 4.2)	Furstenberg multiple recurrence theorem (Thm. 3.4)

This completes our reduction of Szemerédi's theorem to ergodic theory. In next two subsections, we remark on some variations and generalisations of Szemerédi's theorem. These remarks are not needed for the proof of Szemerédi's theorem.

3.3 Finitary versions

Previously we stated the combinatorial results in their "infinitary" versions. In section, we state their "finitary" versions and prove that the two versions of each result are equivalent. In each case, it is very easy to show that the finitary version implies the infinitary version, while for the converse we use a "compactness-and-contradiction" method.

As a toy example, let us state the infinitary and the finitary versions of the pigeonhole principle.

Theorem 3.6 (Pigeonhole principle)**.**

1. *(Infinitary version) If we colour $\mathbb{Z}$ using c colours, then some colour class contains infinitely many elements.*

2. *(Finitary version) For every positive integers c and k, there exists some $N(c,k)$ such that if $n \geq N(c,k)$ and we colour $\{1,2,\ldots,n\}$ using c colours, then some colour class contains at least k elements.*

Of course, we know that we can choose $N(c,k) = c(k-1)+1$. However, in other results, we do not know what the optimal value of N is.

Similarly, we have two versions of van der Waerden's theorem.

Theorem 3.7 (van der Waerden's theorem)**.**

1. *(Infinitary version) If we colour $\mathbb{Z}$ using c colours, then for every k, there is a monochromatic k-term arithmetic progression.*

2. *(Finitary version) For every positive integers c and k, there is some $N(c,k)$ such that if $n \geq N(c,k)$ and we colour $\{1,2,\ldots,n\}$ using c colours, then there is a monochromatic k-term arithmetic progression.*

Proof that the two versions of van der Waerden's theorem are equivalent. It is clear that the finitary version implies the infinitary version, since given any colouring of $\mathbb{Z}$, we can just look at the colouring of the first $N(c,k)$ positive integers to find a monochromatic k-term arithmetic progression.

To prove the converse, let Ω denote the set of set of c colours used. Consider the Bernoulli system $\Omega^{\mathbb{Z}}$. A point in this space corresponds to a colouring of $\mathbb{Z}$ with colours from Ω. This space is sequentially compact by a diagonalisation argument.

Suppose that $N(c,k)$ doesn't exist for some c and k, so that for every n there is some colouring of $\{-n,-n+1,\ldots,n\}$ containing no monochromatic k-term arithmetic progression. Extend such a colouring arbitrarily to a colouring of $\mathbb{Z}$, represented by $x_n \in \Omega^{\mathbb{Z}}$. By sequential compactness of $\Omega^{\mathbb{Z}}$, some subsequence of x_{n_j} converge to some $x^* \in \Omega^{\mathbb{Z}}$. We claim that x^* also contains no k-term arithmetic progression. Indeed, suppose it did, and suppose that the arithmetic progression is contained in $[-n,n]$. Since $x_{n_j} \to x^*$, it follows that the colouring corresponding to x_{n_j} must eventually agree with x^* on $[-n,n]\cap\mathbb{Z}$. But for $n_j \geq n$, x_{n_j} cannot contain a k-term arithmetic progression in $[-n,n]$ by construction, so x^* cannot contain any monochromatic k-term arithmetic progression. But this contradicts the infinitary version of van der Waerden's theorem. □

And likewise we have two versions of Szemerédi's theorem.

Theorem 3.8 (Szemerédi's theorem)**.**

1. *(Infinitary version) Any subset of $\mathbb{Z}$ with positive upper density contains arbitrarily long arithmetic progressions.*

2. *(Finitary version) Let $\delta > 0$, and let k be a positive integer. Then there is some $N(\delta,k)$ such that if $n \geq N(\delta,k)$ then any subset of $\{1,2,\ldots,n\}$ with at least δn elements contains a k-term arithmetic progression.*

The proof of equivalence is almost identical to the one for van der Waerden's theorem.

Proof that the two versions of Szemerédi's theorem are equivalent. First we deduce the infinitary version from the finitary version. Let $A \subset \mathbb{Z}$ have upper density δ. Let k be a positive integer. Then we can find some n with $2n+1 > N(\frac{\delta}{2}, k)$ such that $\frac{1}{2n+1}|A \cap [-n,n]| > \frac{\delta}{2}$. By the finitary version, $[-n,n] \cap A$ contains a k-term arithmetic progression.

For the converse, we can use a compactness argument as before. Suppose that $N(\delta, k)$ doesn't exist. We can pick a sequence $n_j \to \infty$ such that for each j, there is a subset of $\{-\lfloor n_j/2 \rfloor, -\lfloor n_j/2 \rfloor + 1, \dots, \lceil n_j/2 \rceil\}$ with density at least δ containing no k-term arithmetic progression. Represent this set as $x_j \in 2^{\mathbb{Z}}$. Then, as in the previous proof, we can take a subsequence limit $x^* \in 2^{\mathbb{Z}}$ which represents a subset of $\mathbb{Z}$ with upper density at least δ and containing no k-term arithmetic progression, contradicting the infinitary version of Szemerédi's theorem. $\square$

We can ask for the growth rate of $N(\delta, k)$. One disadvantages of the ergodic theoretic approach to Szemerédi's theorem is that it is unable to give any bounds on $N(\delta, k)$, as we jump straight to the infinitary version when bringing the problem to the domain of measure preserving systems, using the axiom of choice in the process. Szemerédi's original combinatorial proof relied on the finitary formulation, but it gave very poor bounds on $N(\delta, k)$, ones of Ackerman type. Gowers' proof [Gow01] gave much better bounds, showing that we can take $N(\delta, k) = \exp\left(\delta^{-2^{2^{k+9}}}\right)$. Tao [Tao06a] combines Furstenberg's ergodic theoretic ideas and Gowers' machinery to extract some quantitative bounds from the ergodic theoretic approach.

3.4 Generalisations of Szemerédi's theorem

The ergodic theoretic approach has lead to some generalisations of Szemerédi's theorem. By considering multiple transformations on X, Furstenberg and Katznelson [FK78] have obtained a multidimensional generalisation of Szemerédi's theorem.

Theorem 3.9 (Multidimensional Szemerédi theorem [FK78])**.** *Let $d \geq 1$. Let $v_1, \dots, v_k \in \mathbb{Z}^d$. Let A be a subset of $\mathbb{Z}^d$ of positive upper density. Then A contains a pattern of the form $w + nv_1, \dots, w + nv_k$ for some $w \in \mathbb{Z}^d$ and $n > 0$.*

Theorem 3.10 (Recurrence for multiple commuting shifts [FK78])**.** *Let $k \geq 1$ be an integer. Let $(X, \mathcal{X}, \mu)$ be a probability space, and let $T_1, \dots, T_k : X \to X$ be probability space isomorphisms that commute with each other. Let $E \in \mathcal{X}$ with $\mu(E) > 0$. Then there exists $n > 0$ such that*

$$\mu\left(T_1^n E \cap T_2^n E \cap \cdots \cap T_k^n E\right) > 0.$$

Theorems 3.9 and 3.10 are equivalent via a variation of the correspondence principle. Note that Furstenberg multiple recurrence (Theorem 3.4) follows from Theorem 3.10 by setting $T_i = T^{i-1}$.

Bergelson and Leibman [BL96] later generalised the result to polynomials.

Theorem 3.11 (Multidimensional polynomial Szemerédi's theorem [BL96])**.** *Let $d \geq 1$. Let $P_1, \dots, P_k : \mathbb{Z} \to \mathbb{Z}^d$ be polynomials with $P_1(0) = \cdots = P_k(0) = 0$. Let $A \subset \mathbb{Z}^d$ be a set of positive upper density. Then A contains a pattern of the form $w + P_1(n), \dots, w + P_k(n)$ for some $w \in \mathbb{Z}^d$ and $n > 0$.*

Theorem 3.12 (Polynomial recurrence for multiple commuting shifts [BL96])**.** *Let k, $(X, \mathcal{X}, \mu)$, $T_1, \dots, T_k \colon X \to X$, E be as in Theorem 3.10, and let $P_1, \dots, P_k$ be as in Theorem 3.11. Then there is some $n > 0$ such that*

$$\mu\left(T^{P_1(n)} E \cap \cdots \cap T^{P_k(n)} E\right) > 0$$

where we write $T^{(a_1, \dots, a_k)} = T_1^{a_1} \cdots T_k^{a_k}$.

Using a more advanced form of the correspondence principle, Furstenberg and Katznelson [FK91] proved the density Hales-Jewett theorem. To state this result, we need to define the notion of a *combinatorial line*. Let $[k] := \{1, 2, \ldots, k\}$. A combinatorial line is a sequence of the form $(x(1), \ldots, x(k))$ where x is a template, i.e., an element in $([k] \cup \star)^n$ containing at least one $\star$, and $x(i)$ is obtained by substituting each $\star$ by i. For example, when $k = 3$, $n = 8$, $x = 132 \star \star 22\star$, the corresponding combinatorial line is

$$(132\mathbf{11}22\mathbf{1}, 132\mathbf{22}22\mathbf{2}, 132\mathbf{33}22\mathbf{3})$$

Hales-Jewett theorem is a strengthening of van der Waerden theorem, claiming that every r-colouring of $[k]^n$ contains a combinatorial line, provided that n is large enough. Density Hales-Jewett theorem generalises Hales-Jewett theorem in the same way that Szemerédi's theorem generalises van der Waerden's theorem.

Theorem 3.13 (Density Hales-Jewett [FK91])**.** *For every positive integer k and real number $\delta > 0$, there exists some positive integer $N(k, \delta)$ such that whenever $n > N(k, \delta)$, every subset of $[k]^n$ with density at least δ contains a combinatorial line.*

Finally, we cannot help but to mention the recent celebrated result of Green and Tao [GT08], proved in 2004, extending Szemerédi's theorem and combines ergodic theory with other power tools to establish the long standing folklore conjecture about arithmetic progressions of prime numbers.

Theorem 3.14 (Green-Tao [GT08])**.** *The prime numbers contain infinitely many arithmetic progressions of length k for all k.*

4 Examples of structure and randomness

4.1 SZ systems

Using the Furstenberg correspondence principle, we showed that Szemerédi's theorem is equivalent to Furstenberg multiple recurrence theorem (Theorem 3.4), which states that in every measure preserving system $(X, \mathcal{X}, \mu, T)$, for any positive integer k and any $E \in \mathcal{X}$ with $\mu(E) > 0$, there exists some $n > 0$ such that

$$\mu\left(E \cap T^n E \cap T^{2n} E \cap \cdots \cap T^{(k-1)n} E\right) > 0. \tag{4.1}$$

As indicated in the introduction, we will prove the apparently stronger result that

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\left(E \cap T^n E \cap T^{2n} E \cap \cdots \cap T^{(k-1)n} E\right) > 0 \tag{4.2}$$

for all $E \in \mathcal{X}$ with $\mu(E) > 0$. In fact, using an argument of Varnavides [Var59] (see [Tao09, Sec. 2.10.1]), one can also deduce (4.2) from (4.1) for all systems.

It is convenient to state this property in terms of functions.

Definition 4.1 (SZ systems)**.** Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system. We say that X is *SZ* (short for Szemerédi) of level k if

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f \; T^n f \; T^{2n} f \cdots T^{(k-1)n} f \; d\mu > 0. \tag{4.3}$$

whenever $f \in L^\infty(X)$, $f \geq 0$ and $\mathbf{E}(f) > 0$. We say that X is *SZ* if it is SZ of every level.

The conditions (4.2) for all E and (4.3) for all f are equivalent, since in one direction we can take $f = 1_E$ and in the other direction we can bound f from below by $\epsilon 1_{\{f \geq \epsilon\}}$.

Through the Furstenberg correspondence principle, we know that Szemerédi's theorem for k-term arithmetic progressions follows from showing that every measure preserving system is SZ of level k. The goal for the rest of this essay is to prove the following result.

Theorem 4.2. *Every measure preserving system is SZ.*

Szemerédi's theorem for k-term arithmetic progressions is trivial when $k = 1, 2$. Similarly, (4.3) is trivial when $k = 1$. When $k = 2$, (4.3) follows from the weak ergodic theorem (Corollary 2.14). By the ergodic decomposition theorem (Theorem 2.20) we may assume that X is ergodic. The case $k = 3$ is the first non-trivial case, known as Roth's theorem, which we prove in Section 7.

We look at two examples of systems that are easily shown to be SZ, but for different reasons. These two examples represent the two extreme situations, where the system is either "very structured" or "very random".

4.2 Pseudorandomness: a weak mixing system

Our example of a random-like system is the Bernoulli system $\Omega^{\mathbb{Z}}$ from Example 2.2(d). Suppose that we have some probability distribution on Ω, so that $\omega \in \Omega$ is assigned probability p_ω. The σ-algebra of $\Omega^{\mathbb{Z}}$ is generated by the evaluation maps $x \mapsto x_n : \Omega^{\mathbb{Z}} \to \Omega$, and the probability measure μ is the given by the product measure

$$\mu\left\{x : x_{i_1} = \omega_1, x_{i_2} = \omega_2, \ldots, x_{i_m} = \omega_m\right\} = p_{\omega_1} p_{\omega_2} \cdots p_{\omega_m}.$$

Recall that T acts on $\Omega^{\mathbb{Z}}$ by shifting each sequence to the right by one index, i.e., $T(x_n)_{n\in\mathbb{Z}} = (x_{n-1})_{n\in\mathbb{Z}}$.

Let us denote this Bernoulli system by $X = (X, \mathcal{X}, \mu, T)$. Then X exhibits the following mixing behaviour.

Proposition 4.3. *Let $(X, \mathcal{X}, \mu, T)$ be a Bernoulli system. Let $E_0, E_1, \ldots, E_{k-1} \in \mathcal{X}$. Then*

$$\lim_{n\to\infty} \mu\left(E_0 \cap T^n E_1 \cap T^{2n} E_2 \cap \cdots \cap T^{(k-1)n} E_{k-1}\right) \to \mu(E_0)\mu(E_1)\cdots\mu(E_{k-1})$$

It follows from the proposition that X is SZ, as

$$\lim_{n\to\infty} \mu\left(E \cap T^n E \cap T^{2n} E \cap \cdots \cap T^{(k-1)n} E\right) = \mu(E)^k > 0.$$

Proof of Proposition 4.3. A *cylinder set* is an event of the form $E = \left\{x \in \Omega^{\mathbb{Z}} : (x_{i_1}, \ldots, x_{i_m}) \in \overline{E}\right\}$ for some $i_1 < i_2 < \cdots < i_m$ and some $\overline{E} \subset \Omega^m$. In other words, it is an event specified by the values on a fixed finite set of terms. These events form a dense subfamily of $\mathcal{X}$, in the sense that any measurable event can be approximated arbitrarily closely by a cylinder event. Then, it suffices to prove the proposition in the case when each E_i is a cylinder set, as we can obtain the general case by taking a limit.

When $E_0, E_1, \ldots, E_{k-1}$ are all cylinder events, the proposition is rather trivial, as for n large enough, the defining coordinates of $E_0, T^n E_1, T^{2n} E_2, \ldots, T^{(k-1)n} E_{k-1}$ are all disjoint, so that these events are actually independent. So

$$\mu\left(E_0 \cap T^n E_1 \cap T^{2n} E_2 \cap \cdots \cap T^{(k-1)n} E_{k-1}\right) = \mu(E_0)\mu(E_1)\cdots\mu(E_{k-1}).$$

for all n sufficiently large. This establishes the proposition. $\square$

4.3 Structure: a compact system

Our example of a structured system is the circle rotation system $(\mathbb{R}/\mathbb{Z}\mathcal{B}, \mu, T_a)$ from Example 2.2(b). Recall that $a \in \mathbb{R}/\mathbb{Z}$ is fixed and T_a is the rotation given by $T_a x = x + a$. If a is rational, then T is periodic, in the sense that T^p is the identity for some positive integer p. It is easy to see that periodic systems are automatically SZ. So we are interested in the case when a is irrational. In this case, although T is not periodic, it is almost so, since T^n can be made arbitrarily close to the identity as na can be made arbitrarily close to an integer. This contrasts the situation in the previous example, where the transformation mixes the system. In the current example, the transformation tends to preserve the structure of the system.

Proposition 4.4. *The system* $(\mathbb{R}/\mathbb{Z}, \mathcal{B}, \mu, T_a)$ *is SZ.*

Proof. We shall use the set version of SZ as in (4.2). Let $E \in \mathcal{B}$ with $\mu(E) > 0$. Observe that for every $\epsilon > 0$ there exists $\delta > 0$ so that

$$\mu\left(E \cap (E - y) \cap (E - 2y) \cap \cdots (E - (k-1)y)\right) > \mu(E) - \epsilon.$$

whenever $|y| < \delta$. Choose $\epsilon = \frac{1}{2}\mu(E)$. Then

$$\begin{aligned} \mu\left(E \cap T^n E \cap T^{2n} E \cap \cdots \cap T^{(k-1)n} E\right) &= \mu\left(E \cap (E + na) \cap (E + 2na) \cap \cdots \cap (E + (k-1)na)\right) \\ &> \mu(E) - \epsilon = \frac{1}{2}\mu(E). \end{aligned}$$

whenever na lies within $\frac{\delta}{k}$ of an integer, and such n occupies a positive lower density since $\{na : n \in \mathbb{N}\}$ is equidistributed in $\mathbb{R}/\mathbb{Z}$. Hence the system is SZ. $\square$

In the next two sections we explore weak mixing and compact systems in more generality.

5 Weak mixing systems

5.1 Cesàro and density convergence

We need to frequently consider limits of partial averages, so it is convenient to introduce some notation.

In a normed vector space, the usual convergence of a sequence $v_n \to v$ means that $\lim_{n\to\infty} \|v_n - v\| = 0$. It will be useful to consider two other notions of convergence: convergence in the Cesàro sense and convergence in density.

Definition 5.1. Let $v_0, v_1, \ldots$ be a sequence in a normed vector space V. Let $v \in V$.

1. (Usual convergence in norm) We say that $\lim_{n\to\infty} v_n = v$ if $\lim_{n\to\infty} \|v_n - v\| = 0$.
2. (Convergence in density) We say that v_n converges to v *in density*, denoted $\text{D-lim}_{n\to\infty} v_n = v$, if for any $\epsilon > 0$, the set $\{n \in \mathbb{N} : \|v_n - v\| > \epsilon\}$ has upper density zero.
3. (Cesàro convergence) We say that v_n converges to v *in a Cesàro sense* , denoted $\text{C-lim}_{n\to\infty} v_n = v$, if $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} v_n = v$.
4. (Cesàro supremum) Define $\text{C-sup}_{n\to\infty} v_n = \limsup_{N\to\infty} \left\|\frac{1}{N}\sum_{n=0}^{N-1} v_n\right\|$. Note that this is a nonnegative real number.

Note that each in (a), (b), (c), the limit v is unique if it exists. Also $\text{C-lim}_{n\to\infty} v_n = 0$ is equivalent to $\text{C-sup}_{n\to\infty} v_n = 0$.

Convergence in norm implies both convergence in density and Cesàro convergence. When the sequence is uniformly bounded (which is the case for most of our applications), convergence in density implies Cesàro convergence (see Proposition 5.5 below). So we have the following chain of implications for *bounded sequences.*

$$\text{convergence in norm} \implies \text{convergence in density} \implies \text{Cesàro convergence}$$

Counterexamples to converses are given below.

Example 5.2. In $\mathbb{R}$, the sequence $x_n = (-1)^n$ does not convergence in norm or density, but it converges to 0 in the Cesàro sense.

Example 5.3. In $\mathbb{R}$, the sequence defined by setting x_n to be 1 if n is a perfect square and 0 otherwise does not convergence in norm, but convergences to 0 in density as well in a Cesàro sense.

Example 5.4. (Unbounded sequences) In $\mathbb{R}$, the sequence defined by setting x_{k^2} to be $2k+1$ and $x_n = 0$ if n is not a perfect square does not convergence in norm, but converges to 0 in density and converges to 1 in a Cesàro sense. Similarly, the sequence defined by setting $x_n = n$ if x_n is a power of 2 and $x_n = 0$ otherwise does not convergence in norm or in a Cesàro sense, but does converge to 0 in density. These examples show that these notions of convergence behave poorly when the sequence is unbounded. Fortunately, we shall work almost exclusively with bounded sequences.

The following results show how Cesàro convergence and convergence in density are the related.

Proposition 5.5. *Let $v_0, v_1, v_2, \dots$ be a bounded sequence of vectors in a normed vector space V, and let $v \in V$. Then the following are equivalent:*

(a) $\text{C-lim}_{n\to\infty} \|v_n - v\| = 0$.

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

(c) $\text{D-lim}_{n\to\infty} v_n = v$.

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

Proof. Since the sequence v_n is bounded, we may assume by scaling if necessary that $\|v_n - v\| \le 1$ for all n.

(a) $\implies$ (b): We have $\|v_n - v\|^2 \le \|v_n - v\|$, so that $\text{C-lim}_{n\to\infty} \|v_n - v\|^2 \le \text{C-lim}_{n\to\infty} \|v_n - v\| = 0$.

(b) $\implies$ (c): If there is some $\epsilon > 0$ such that $\{n : \|v_n - v\| > \epsilon\}$ has upper density δ, then

$$\underset{n\to\infty}{\text{C-sup}} \|v_n - v\|^2 = \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \|v_n - v\|^2 \ge \delta\epsilon^2 > 0,$$

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

(c) $\implies$ (a): Let $\epsilon > 0$ be arbitrary. We know that $\{n : \|v_n - v\| > \epsilon\}$ has upper density zero. Using the assumption $\|v_n - v\| \le 1$, we get

$$\underset{n\to\infty}{\text{C-sup}} \|v_n - v\| = \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \|v_n - v\| \le \epsilon + \limsup_{N\to\infty} \frac{1}{N} \left|\{n : 0 \le n < N, \|v_n - v\| > \epsilon\}\right| = \epsilon.$$

Since ϵ could be arbitrarily small, we get $\text{C-lim}_{n\to\infty} \|v_n - v\| = 0$.

Finally, (a), (b), (c) all imply that $\text{C-lim}_{n\to\infty} v_n = v$ as

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} v_n - v \right\| \le \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \|v_n - v\| = \underset{n\to\infty}{\text{C-lim}} \|v_n - v\| = 0.$$

$\square$

Although Cesàro convergence usually does not imply convergence in density, we may be able to deduce convergence in density if we can compute the Cesàro limit of the sequence and its square. The following proposition is a handy fact. The same proof works in Hilbert spaces, but we will only need the result for real sequences.

Proposition 5.6. *Let $x_0, x_1, x_2, \ldots$ be a sequence of real numbers. If* $\text{C-lim}_{n\to\infty} x_n = x$ *and* $\text{C-lim}_{n\to\infty} x_n^2 = x^2$ *for some $x \in \mathbb{R}$, then* $\text{D-lim}_{n\to\infty} x_n = x$.

Proof. We have

$$\operatorname*{C-lim}_{n\to\infty} |x_n - x|^2 = \operatorname*{C-lim}_{n\to\infty} \left(x_n^2 - 2xx_n + x^2\right) = x^2 - 2x \cdot x + x^2 = 0.$$

So the result follows from Proposition 5.5. □

5.2 Weak mixing systems

We say that a system $(X, \mathcal{X}, \mu, T)$ is *mixing* if for any two sets $A, B \in \mathcal{X}$, we have $\lim_{n\to\infty} \mu(T^n A \cap B) = \mu(A)\mu(B)$. However, this requirement is often too strong for our needs. We shall consider systems where this limit holds in a weaker sense.

Definition 5.7 (Weak mixing system)**.** A measure preserving system $(X, \mathcal{X}, \mu, T)$ is *weak mixing* if

$$\operatorname*{D-lim}_{n\to\infty} \mu(T^n A \cap B) = \mu(A)\mu(B). \tag{5.1}$$

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

$$\operatorname*{D-lim}_{n\to\infty} \langle T^n f, g\rangle = \mathbf{E}(f)\mathbf{E}(g) \tag{5.2}$$

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

The two definitions are equivalent since in one direction we can take characteristic functions $f = 1_A$, $g = 1_B$, and in the other direction we can approximate f and g by simple measurable functions, noting that $\langle T^n f, g\rangle$ is bilinear in f and g. It is interesting to compare weak mixing with ergodicity. It can be shown that X is ergodic if and only if

$$\operatorname*{C-lim}_{n\to\infty} \langle T^n f, g\rangle = \mathbf{E}(f)\mathbf{E}(g). \tag{5.3}$$

for every $f, g \in L^2(X)$. The "only if" direction is a consequence of the weak ergodic theorem (Corollary 2.14). For the "if" direction, take f to be any T-invariant function, so $\langle f, g\rangle = \mathbf{E}(f)\mathbf{E}(g)$, thus $f - \mathbf{E}(f)$ is orthogonal to every $g \in L^2(X)$, and hence $f = \mathbf{E}(f)$ a.e. Since convergence in density implies Cesàro convergence for bounded sequences, and $\langle T^n f, g\rangle$ is uniformly bounded by $\|f\|_2 \|g\|_2$ by Cauchy-Schwarz, we deduce the following.

Proposition 5.8. *Every weak mixing system is ergodic.*

The Bernoulli system from Section 4.2 is an example of a weak mixing system.

5.3 Product characterisation

There is an alternate characterisation of weak mixing systems through product systems. If $X = (X, \mathcal{X}, \mu, T)$ and $Y = (Y, \mathcal{Y}, \mu, S)$ are two measure preserving systems, their direct product $X \times Y = (X \times X, \mathcal{X} \times \mathcal{Y}, \mu \times \nu, T \times S)$ is another measure preserving system. For any $f \in L^2(X)$, $g \in L^2(Y)$, we can construct the function $f \otimes g \in L^2(X \times Y)$, given by $(f \otimes g)(x, y) = f(x)g(y)$. The functions $\left\{f \otimes g : f \in L^2(X), g \in L^2(Y)\right\}$ span linearly a dense subspace of $L^2(X \times Y)$. In many situations, to prove something about $L^2(X \times Y)$, it suffices to check it for functions of the form $f \times g$.

In the following proposition, we show that X is weak mixing if and only if $X \times X$ is ergodic. In fact, we could have defined weak mixing system this way, though the previous definition probably sheds more light on the mixing nature of these systems.

The approach taken in [Fur81, FKO82] uses the product characterisation, while the one taken in [Tao09] avoids it. We discuss both approaches in this essay.

Proposition 5.9 (Product characterisation of weak mixing systems)**.** *Let X be a measure preserving system. The following are equivalent:*

(a) X is weak mixing;

(b) $X \times X$ is weak mixing;

(c) $X \times X$ is ergodic.

Proof. (a) $\implies$ (b): Suppose that X is weak mixing. To show that $X \times X$ is weak mixing, it suffices to show that

$$\operatorname*{D-lim}_{n\to\infty} \langle (T\times T)^n(f_1\otimes f_2), g_1\otimes g_2\rangle_{X\times X} = \mathbf{E}_{X\times X}(f_1\otimes f_2)\mathbf{E}_{X\times X}(g_1\otimes g_2). \tag{5.4}$$

We have,

$$\begin{aligned}
\langle (T\times T)^n(f_1\otimes f_2), g_1\otimes g_2\rangle_{X\times X} &= \int_{X\times X} ((T\times T)^n(f_1\otimes f_2))(g_1\otimes g_2)\ d(\mu\times\mu) \\
&= \int_{X\times X} ((T^n f_1)g_1)\otimes((T^n f_2)g_2)\ d(\mu\times\mu) \\
&= \int_X (T^n f_1)g_1\ d\mu \int_X (T^n f_2)g_2\ d\mu \\
&= \langle T^n f_1, g_1\rangle \langle T^n f_2, g_2\rangle
\end{aligned}$$

Since X is weak mixing, $\langle T^n f_i, g_i\rangle$ converges in density to $\mathbf{E}(f_i)\mathbf{E}(g_i)$ for $i = 1, 2$. Thus, the above quantity converges in density to $\mathbf{E}(f_1)\mathbf{E}(f_2)\mathbf{E}(g_1)\mathbf{E}(g_2) = \mathbf{E}_{X\times X}(f_1\otimes f_2)\mathbf{E}_{X\times X}(g_1\otimes g_2)$. This proves (5.4).

(b) $\implies$ (c): This follows from Proposition 5.8 as every weak mixing system is ergodic.

(c) $\implies$ (a): Suppose that $X \times X$ is ergodic. Then X is also ergodic (or else we would have a nontrivial T-invariant set in $X \times X$ of the form $E \times X$). Let $f, g \in L^2(X)$. By the weak ergodic theorem (Corollary 2.14) on X,

$$\operatorname*{C-lim}_{n\to\infty} \langle T^n f, g\rangle = \mathbf{E}(f)\mathbf{E}(g).$$

And applying the weak ergodic theorem to $X \times X$, we get

$$\begin{aligned}
\operatorname*{C-lim}_{n\to\infty} \langle T^n f, g\rangle^2 &= \operatorname*{C-lim}_{n\to\infty} \langle (T\times T)^n(f\otimes f), (g\otimes g)\rangle_{X\times X} \\
&= \mathbf{E}_{X\times X}(f\otimes f)\mathbf{E}_{X\times X}(g\otimes g) = \mathbf{E}(f)^2\mathbf{E}(g)^2.
\end{aligned}$$

Therefore, $\text{D-lim}_{n\to\infty} \langle T^n f, g\rangle = \mathbf{E}(f)\mathbf{E}(g)$ by Proposition 5.6. Hence X is weak mixing. $\square$

5.4 van der Corput lemma

We need a technical lemma for showing that certain Cesàro limits are zero.

Lemma 5.10 (van der Corput)**.** *Let $v_0, v_1, v_2, \ldots$ be a bounded sequence of vectors in a real Hilbert space. If*

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

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

Lemma 5.11. *Let $v_0, v_1, v_2, \dots$ be sequence of vectors in a real Hilbert space each with norm at most* 1*. Then*

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} v_n \right\|^2 \le \frac{2}{H} \sum_{h=0}^{H-1} \left(1 - \frac{h}{H}\right) \frac{1}{N} \sum_{n=0}^{N-1} \langle v_n, v_{n+h} \rangle + O\left(\frac{H}{N}\right). \tag{5.5}$$

Proof. For any h with $0 \le h < H$, we have

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

Averaging over h, we have

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

By convexity, we have

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

The last step comes from rearranging the terms and absorbing the discrepancies into the $O(H/N)$ error term. □

Proof of Lemma 5.10. By scaling if necessary, we may assume $\|v_i\| \le 1$ for all i. Letting $N \to \infty$ in (5.5) gives

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

Letting $H \to \infty$ and using $\text{C-sup}_{h\to\infty} \text{C-sup}_{n\to\infty} \langle v_n, v_{n+h} \rangle = 0$, we find that $\text{C-lim}_{n\to\infty} v_n = 0$. □

5.5 Weak mixing functions

Previously we defined a weak mixing system. Following Tao [Tao09], we consider the notion of weak mixing functions. Roughly speaking, weak mixing functions are functions f whose shifts $T^n f$ eventually become orthogonal to f, thereby displaying "mixing" behaviour. There is an alternative characterisation of weak mixing systems via the property that every mean zero function is a weak mixing. The notion of of weak mixing functions was not present in the works of Furstenberg [FKO82, Fur77, Fur81].

Definition 5.12 (Weak mixing functions)**.** Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system. A function $f \in L^2(X)$ *weak mixing* if $\text{D-lim}_{n\to\infty} \langle T^n f, f \rangle = 0$.

In Tao [Tao09], the following characterisation was given as the definition of weak mixing systems.

Proposition 5.13. *A measure preserving system $(X, \mathcal{X}, \mu, T)$ is weak mixing if and only if every $f \in L^2(X)$ with mean zero is weak mixing.*

The following proposition shows that not only do weak mixing functions mix with themselves, they also mix with any other function.

Proposition 5.14. *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system, and let $f \in L^2(X)$ be weak mixing. Then for any $g \in L^2(X)$ we have*

$$\operatorname*{D-lim}_{n\to\infty} \langle T^n f, g\rangle = 0, \qquad \textit{and} \qquad \operatorname*{D-lim}_{n\to\infty} \langle f, T^n g\rangle = 0.$$

Proof. Since $\langle f, T^n g\rangle = \langle T^{-n} f, g\rangle$ due to T-invariance, we see that the property of f being weak mixing is preserved upon changing the system by replacing T by T^{-1}. So if we can show that $\text{D-lim}_{n\to\infty} \langle T^n f, g\rangle = 0$ is true for any system, then the other claim also follows.

By Proposition 5.5, it suffices to prove that $\text{C-lim}_{n\to\infty} |\langle T^n f, g\rangle| = 0$. We have

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

By Cauchy-Schwarz, it suffices to show that $\text{C-lim}_{n\to\infty} \langle T^n f, g\rangle\, T^n f = 0$ in $L^2(X)$. By van der Corput lemma (Lemma 5.10), it suffices to prove that

$$\operatorname*{C-lim}_{h\to\infty} \operatorname*{C-sup}_{n\to\infty} \langle T^n f, g\rangle \left\langle T^{n+h} f, g\right\rangle \left\langle T^n f, T^{n+h} f\right\rangle = 0.$$

Both $|\langle T^n f, g\rangle|$ and $\left|\left\langle T^{n+h} f, g\right\rangle\right|$ are bounded, say by C. Also $\left\langle T^n f, T^{n+h} f\right\rangle = \left\langle f, T^h f\right\rangle$, independent of n. Therefore,

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

as f is weak mixing. This completes the proof. □

Proof of Proposition 5.13. If X is weak mixing, then (5.2) shows that every mean zero function is weak mixing. Conversely, if every mean zero function in X is weak mixing, then (5.2) follows from applying Proposition 5.14 to $\langle T^n(f - \mathbf{E}(f)), g\rangle$. □

5.6 Multiple recurrence

In this subsection we prove that every weak mixing system is SZ.

Theorem 5.15. *Every weak mixing system is SZ.*

The following result gives a very precise statement about the mixing behaviour in weak mixing systems. While weak mixing systems are only a priori defined to have mixing behaviour for a pair of functions, the result shows that the mixing behaviour in fact holds for any number of functions.

Proposition 5.16. *Let $(X, \mathcal{X}, \mu, T)$ be a weak mixing system. Let $k \ge 1$. Let $a_1, \dots, a_k \in \mathbb{Z}$ be distinct non-zero integers, and let $f_1, \dots, f_k \in L^\infty(X)$. Then*

$$\operatorname*{C-lim}_{n\to\infty} T^{a_1 n} f_1 \cdots T^{a_k n} f_k = \mathbf{E}(f_1) \cdots \mathbf{E}(f_k)$$

in $L^2(X)$.

Proof. We induct on k. For $k = 1$, note that $(X, \mathcal{X}, \mu, T^{a_1})$ is also a weak mixing system, hence an ergodic system (Proposition 5.8), so $\text{C-lim}_{n\to\infty} T^{a_1 n} f_1 = \mathbf{E}(f_1)$ follows from the mean ergodic theorem (Theorem 2.13).

By considering $f_1 - \mathbf{E}(f_1)$ and using induction on $f_2, \ldots, f_k$, it suffices to show that

$$\operatorname*{C-lim}_{n\to\infty} T^{a_1 n} f_1 \cdots T^{a_k n} f_k = 0$$

in $L^2(X)$ whenever f_1 has mean zero and hence weak mixing. Applying the van der Corput lemma (Lemma 5.10) on the sequence $v_n = T^{a_1 n} f_1 \cdots T^{a_k n} f_k \in L^\infty(X) \subset L^2(X)$, we see that it suffices to prove that $\text{C-sup}_{h\to\infty} \text{C-sup}_{n\to\infty} \langle v_n, v_{n+h}\rangle = 0$. Here

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

where $f_{j,h} = f_j T^{a_j h} f_j$. By Cauchy-Schwarz inequality, it suffices to show that

$$\operatorname*{C-sup}_{h\to\infty} \operatorname*{C-sup}_{n\to\infty} T^{(a_1-a_k)n} f_{1,h} \cdots T^{(a_{k-1}-a_k)n} f_{k-1,h} = 0.$$

Applying the induction hypothesis, we have

$$\operatorname*{C-lim}_{n\to\infty} T^{(a_1-a_k)n} f_{1,h} \cdots T^{(a_{k-1}-a_k)n} f_{k-1,h} = \mathbf{E}(f_{1,h}) \cdots \mathbf{E}(f_{k-1,h}).$$

This quantity converges to 0 in density as $h \to \infty$, since all $\mathbf{E}(f_{j,h})$ are bounded and $\mathbf{E}(f_{1,h}) = \mathbf{E}(f_1 T^{a_1 h} f_1) = \langle f_1, T^{a_1 h} f_1 \rangle$ converges to 0 in density as f_1 is weak mixing. $\square$

Since convergence in $L^2(X)$ implies weak convergence, we have the following corollary.

Corollary 5.17. *Same assumptions as Proposition 5.16. Then*

$$\operatorname*{C-lim}_{n\to\infty} \int_X T^{a_1 n} f_1 \cdots T^{a_k n} f_k \, d\mu = \mathbf{E}(f_1) \cdots \mathbf{E}(f_k).$$

We can bootstrap Corollary 5.17 to obtain the following stronger conclusion, although it will not be needed in what follows.

Corollary 5.18. *Same assumptions as Proposition 5.16. Then*

$$\operatorname*{D-lim}_{n\to\infty} \int_X T^{a_1 n} f_1 \cdots T^{a_k n} f_k \, d\mu = \mathbf{E}(f_1) \cdots \mathbf{E}(f_k).$$

Proof. From Proposition 5.9 we know that $X \times X$ is weak mixing. Applying Corollary 5.17 to $X \times X$ and functions $f_i \otimes f_i$, we have

$$\begin{aligned}&\operatorname*{C-lim}_{n\to\infty} \left(\int_X T^{a_1 n} f_1 \cdots T^{a_k n} f_k \, d\mu \right)^2 \\ &= \operatorname*{C-lim}_{n\to\infty} \int_{X\times X} (T\times T)^{a_1 n} (f_1 \otimes f_1) \cdots (T \times T)^{a_k n} (f_k \otimes f_k) \, d(\mu \times \mu) \\ &= \mathbf{E}_{X\times X}(f_1 \otimes f_1) \cdots \mathbf{E}_{X\times X}(f_k \otimes f_k) \\ &= \mathbf{E}(f_1)^2 \cdots \mathbf{E}(f_k)^2.\end{aligned}$$

Convergence in density then follows from Proposition 5.6. $\square$

Now we can conclude the proof that all weak mixing systems are SZ.

Proof of Theorem 5.15. Let X be a weak mixing system, and let $f \in L^\infty(X)$ be a nonnegative function with positive mean. Applying Corollary 5.17, we get

$$\liminf_{N\to\infty} \frac{1}{N} \int \sum_{n=0}^{N-1} f \, T^n f \, T^{2n} f \cdots T^{kn} f \, d\mu = \mathbf{E}(f)^{k+1} > 0.$$

Therefore X is SZ. $\square$

6 Compact systems

6.1 Compact systems

We saw an example of a compact system in Section 4.3, namely rotations on a unit circle. In this section we look at compact systems in more generality.

In a complete metric space X, a set W is called *precompact* if its closure is compact, and *totally bounded* if for every $\epsilon > 0$, W can be covered by a finite collection of ϵ-balls. It is an exercise in general topology to show that these two notions are equivalent in a complete metric space.

We say that a set of integers S is *syndetic* if there is a positive N such that $S \cap [n, n+N] \neq \emptyset$ for every n. In other words, syndetic sets have bounded gaps.

Definition 6.1. Let $(X, \mathcal{X}, \mu, T)$ be a system. We say that $f \in L^2(X)$ is *almost periodic* its orbit $\{T^n f : n \in \mathbb{Z}\}$ is precompact in $L^2(X)$ in the norm topology. Equivalently, f is almost periodic if for every $\epsilon > 0$, the set $\{n \in \mathbb{Z} : \|f - T^n f\|_2 < \epsilon\}$ is syndetic.

Definition 6.2. The system $(X, \mathcal{X}, \mu, T)$ is *compact* if every $f \in L^2(X, \mathcal{X}, \mu)$ is almost periodic.

Example 6.3. The circle rotation system $X = (\mathbb{R}/\mathbb{Z}, \mathcal{B}, \mu, T_a : x \mapsto x + a)$ from Section 4.3 is compact. From Fourier analysis, $L^2(X)$ is spanned by $e_m \in L^2(X)$ defined by $e_m(x) := e^{2\pi imx}$. Each e_m is an eigenfunction of T_a, as $T_a e_m = e^{-2\pi ima} e_m$. So $T_a^n e_m = e^{-2\pi inma} e_m$. The set of n for which nma is within ϵ of an integer is syndetic for every $\epsilon > 0$, and hence the set of n for which $\left|e^{-2\pi inma} - 1\right| < \epsilon$ is syndetic. It follows that every e_m is almost periodic. Later in Lemma 6.16 we show that the set of almost periodic functions is a closed subspace in $L^2(X)$. Since we have found a complete basis of almost periodic functions in $L^2(X)$, it follows that X is compact.

6.2 Kronecker systems

The circle rotation system in Example 6.3 arises from a group structure. This is an example of a Kronecker system. Recall that every compact topological group has a Haar measure μ which is invariant under translation.

Definition 6.4. An *(abelian) Kronecker system* is a measure preserving system of the form $(G, \mathcal{B}, \mu, T_a)$, where $(G, +)$ is a compact abelian metrisable group, $\mathcal{B}$ is its Borel algebra, μ is its Haar measure, $a \in G$, and $T_a x = x + a$.

It turns out that Kronecker systems are the building blocks of all compact systems. We would like to consider measure preserving systems up to measure zero sets. This can be done by considering the abstraction $(\mathcal{X}/\sim, \mu, T)$ of a system $(X, \mathcal{X}, \mu, T)$, where we drop the underlying space X and consider the equivalence $\sim$ of sets modulo μ, i.e., identifying sets differing by measure zero. We say that two measure preserving systems $(X, \mathcal{X}, \mu, T)$ and $(Y, \mathcal{Y}, \nu, S)$ are *equivalent* if the corresponding abstractions $(\mathcal{X}/\sim, \mu, T)$ and $(\mathcal{Y}/\sim, \nu, S)$ are isomorphic. In general there may not be a concrete map between the underlying spaces X and Y even if they are equivalent in this abstract sense.

It is not true that all compact systems are equivalent to Kronecker systems. For instance, consider the system on 5 points with T acting as a product of a 2-cycle and a 3-cycle. This system has no group structure, but it can be decomposed as a union of a two systems with group structures, namely $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$. This suggests that we should look at systems that are minimal, i.e., ergodic. We state without proof the following classification result by Halmos and von Neumann [HvN42].

Theorem 6.5 (Halmos and von Neumann)**.** *Every ergodic compact system is equivalent to a Kronecker system.*

We will not explicitly use this theorem in the proof of Szemerédi's theorem, although it does shed a lot of light on the structure of a compact system.

6.3 Multiple recurrence

In this subsection we prove the following result.

Theorem 6.6. *Every compact system is SZ.*

The proof is fairly straightforward. The idea is that for an almost periodic $f \in L^\infty(X)$, we have

$$\int_X f\ T^n f\ T^{2n} f \cdots T^{(k-1)n} f\ d\mu \approx \int_X f^k\ d\mu > 0$$

whenever $T^n f \approx f$, and this occurs for a set of n of positive density.

Proposition 6.7. *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system. For an almost periodic $f \in L^\infty(X, \mathcal{X}, \mu)$ with $f \geq 0$ and $\mathbf{E}(f) > 0$, we have*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f\ T^n f\ T^{2n} f \cdots T^{(k-1)n} f\ d\mu > 0. \tag{6.1}$$

Proof. Without loss of generality, we may assume that f is uniformly bounded by 1. Let $\epsilon > 0$. Since f is almost periodic, $\left\{n \in \mathbb{N} : \|f - T^n f\|_2 < \frac{\epsilon}{k2^k}\right\}$ is syndetic. Fix an n with $\|f - T^n f\|_2 < \frac{\epsilon}{k2^k}$. Since T acts isometrically on $L^2(X)$, we have $\left\|T^{jn} f - T^{(j+1)n} f\right\|_2 < \frac{\epsilon}{k2^k}$. By the triangle inequality, we have $\left\|f - T^{jn} f\right\|_2 < \frac{\epsilon}{2^k}$ for $1 \leq j < k$, so that $T^{jn} = f + g_j$ where $\|g_j\|_2 < \frac{\epsilon}{2^k}$. Then

$$\int_X f\ T^n f\ T^{2n} f \cdots T^{(k-1)n} f\ d\mu = \int_X f(f+g_1) \cdots (f+g_{k-1})\ d\mu.$$

When expanded, we get a sum of $\int f^k\ d\mu$ along with terms of the form $\langle g_j, F\rangle$, where F is some function satisfying $\|F\|_\infty \leq 1$, so that $\langle g_j, F\rangle \leq \|g_j\|_2 \|F\|_2 \leq \|g_j\|_2 \leq \frac{\epsilon}{2^k}$. Since there are at most 2^k such terms, we obtain

$$\int_X f\ T^n f\ T^{2n} f \cdots T^{(k-1)n} f\ d\mu > \int_X f^k\ d\mu - \epsilon.$$

Since f is nonnegative and not zero almost everywhere, we have $\int f^k\ d\mu > 0$. Since the above equation holds for n in a syndetic set, we have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f\ T^n f\ T^{2n} f \cdots T^{(k-1)n} f\ d\mu > \frac{1}{d_\epsilon} \left(\int_X f^k\ d\mu - \epsilon \right)$$

where d_ϵ is the maximum gap in the syndetic set $\left\{n \in \mathbb{N} : \|f - T^n f\|_2 < \frac{\epsilon^2}{k2^{2k}}\right\}$. The desired result follows after choosing ϵ sufficiently small (namely smaller than $\int f^k\ d\mu$). $\square$

Remark. It is also possible to give a proof using the Kronecker system characterisation of compact systems. Using the ergodic decomposition theorem (Theorem 2.20), we may assume that the compact system is ergodic, and hence equivalent to a Kronecker system $(G, \mathcal{B}, \mu, T_a)$. We have

$$\frac{1}{N} \sum_{n=0}^{N-1} \int_X f\, T^n f\, T^{2n} f \cdots T^{(k-1)n} f\ d\mu = \frac{1}{N} \sum_{n=0}^{N-1} \int_G f(x) f(x-na) f(x-2na) \cdots f(x-(k-1)na)\ d\mu(x).$$

The sequence $(na)_{n\in\mathbb{N}}$ is equidistributed in X as the system is ergodic. Thus the limit of the above average as $N \to \infty$ must equal to

$$\int_X \int_X f(x)f(x-y)f(x-2y)\cdots f(x-(k-1)y)\ d\mu(x)d\mu(y).$$

By considering the value of the integral when y is near the identity, we see that the integral is positive. In fact, this argument shows that the limit in (6.1) actually exists.

6.4 Hilbert-Schmidt operators

Later we would like to prove some results about the existence of compact factors in systems, which involves finding almost periodic functions. We need some tools from functional analysis, namely that of compact operators. Compact operators always send bounded sets to precompact sets, and so we can use them to construct almost periodic functions. A particular class of compact operators is Hilbert-Schmidt operators.

Definition 6.8. Let H and H' be (separable) Hilbert spaces with orthonormal bases $(e_a)_{a\in A}$ and $(f_b)_{b\in B}$ respectively. A bounded linear operator $\Phi : H \to H'$ is a *Hilbert-Schmidt operator* if its *Hilbert-Schmidt norm* $\|\Phi\|_{HS}$ is finite, where $\|\Phi\|_{HS}$ is defined by

$$\|\Phi\|_{HS}^2 := \sum_{a\in A} \|\Phi e_a\|_{H'}^2 = \sum_{a\in A}\sum_{b\in B} |\langle \Phi e_a, f_b\rangle_{H'}|^2 = \sum_{a\in A}\sum_{b\in B} |\langle e_a, \Phi^* f_b\rangle_{H'}|^2 = \sum_{b\in B} \|\Phi^* f_b\|_{H'}^2$$

where Φ^* is the adjoint of Φ. Note that the Hilbert-Schmidt norm does not depend on the choice of bases, as the first sum above is independent of $(f_b)_{b\in B}$ and the last sum above is independent of $(e_a)_{a\in A}$

There is a more abstract formulation of Hilbert-Schmidt operators as elements of the tensor product $H \otimes H'$, which is the Hilbert space given by the tensor product of the underlying vector spaces and the inner product $\langle v\otimes v', w\otimes w'\rangle_{H\otimes H'} = \langle v, w\rangle_H \langle v', w'\rangle_{H'}$ whenever $v, w \in H$ and $v', w' \in H'$. Every element $K \in H \otimes H'$ gives rise to a Hilbert-Schmidt operator $\Phi_K : H \to H'$ satisfying

$$\left\langle \Phi_K v, v'\right\rangle_{H'} = \left\langle K, v\otimes v'\right\rangle_{H\otimes H'}.$$

The element K is known as the *kernel* of the operator Φ_K, and it is easy to check that $\|K\|_{H\otimes H'} = |\Phi_K|_{HS}$. In fact, using the Riesz representation theorem, we see that every Hilbert-Schmidt operator is Φ_K for some $K \in H \otimes H'$. This makes the space $HS(H \to H')$ of Hilbert-Schmidt operators into a Hilbert space isomorphic to $H \otimes H'$.

Example 6.9. Every linear operator between finite dimensional vector spaces is Hilbert-Schmidt. The Hilber-Schmidt norm of an operator represented by a matrix is equal to the sum of the squares of the magnitudes of the entries of the matrix. On the other hand, the identity operator in an infinite dimensional Hilbert space is never Hilber-Schmidt.

Example 6.10. The operator $\Phi_K : L^2(X) \to L^2(Y)$ defined by $\Phi_K f(y) = \int_X K(x,y)f(x)\ dx$ for some kernel $K \in L^2(X\times Y)$ is Hilbert-Schmidt. Its Hilbert-Schmidt norm is $\|K\|_{L^2(X\times Y)}$.

Definition 6.11. A linear operator between Banach spaces is *compact* if the image of any bounded set is precompact.

Proposition 6.12. *If $\Phi : H \to H'$ is a Hilbert-Schmidt operator between two Hilbert spaces, then it is compact.*

Proof. It suffices to show that the image of the unit ball in H is totally bounded in H'. Let $\epsilon > 0$ be arbitrary. Since the Hilbert-Schmidt norm is finite, we can find an orthonormal sequence of vectors $e_1, \ldots, e_n \in H$ such that $\|\Phi\|_{HS}^2 \geq \sum_{n=1}^{N} \|\Phi e_n\|_{H'}^2 - \epsilon^2$. Let V denote the subspace spanned by $e_1, \ldots, e_n$. Then for any unit vector x orthogonal to V, we have $\|\Phi\|_{HS}^2 \geq \sum_{n=1}^{N} \|\Phi e_n\|_{H'}^2 + \|\Phi x\|_{H'}^2$, so that $\|\Phi x\|_{H'} \leq \epsilon$ for any unit vector $x \in V^{\perp}$. Since any vector within the unit ball in H can be written as $v + x$ where $v \in V$ and $x \in V^{\perp}$, where we know that $\|v\|_H \leq 1$, $\|x\|_H \leq 1$, and $\|\Phi x\|_{H'} \leq \epsilon$, we see that the image of the unit ball in H' under Φ lies within ϵ of the image of the unit ball in the subspace V under Φ, and hence is totally bounded since the the unit ball in the finite dimensional subspace V is compact. ☐

Here is how we use Hilbert-Schmidt operators to find almost periodic functions.

Proposition 6.13. *Let X be a measure preserving system. If $\Phi : L^2(X) \to L^2(X)$ is a Hilbert-Schmidt operator that commutes with T, then Φf is an almost periodic function for any $f \in L^2(X)$.*

Proof. In $L^2(X)$, $\{T^n f : n \in \mathbb{Z}\}$ is a bounded set, so its image $\Phi \{T^n f : n \in \mathbb{Z}\} = \{\Phi T^n f : n \in \mathbb{Z}\} = \{T^n(\Phi f) : n \in \mathbb{Z}\}$ is precompact by Proposition 6.12, and hence Φf is almost periodic. ☐

6.5 Weak mixing and almost periodic components

Let $X = (X, \mathcal{X}, \mu, T)$ be a measure preserving system. Let $AP(X)$ denote the collection of almost periodic functions in $L^2(X)$, and $WM(X)$ the collection of weak mixing functions in $L^2(X)$. We show that there is an orthogonal decomposition of any function into a weak mixing component and an almost periodic component.

Proposition 6.14. *Let X be a measure preserving system. Then we have*

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

as an orthogonal direct sum of Hilbert spaces.

In Section 6.7 we will give a more explicit description of this decomposition using factors and conditional expectations.

Lemma 6.15. *If $f \in WM(X)$ and $g \in AP(X)$, then $\langle f, g\rangle = 0$.*

Proof. It suffices to prove that $\text{C-lim}_{n\to\infty} |\langle T^n f, T^n g\rangle| = 0$ since $\langle T^n f, T^n g\rangle = \langle f, g\rangle$ by T-invariance. Let $\epsilon > 0$ be arbitrary. Since g is almost periodic, there exists $g_1, \ldots, g_k \in L^2(X)$ such that for every n, we have $\|T^n g - g_i\|_2 < \epsilon$ for some i, and thus by the triangle and Cauchy-Schwarz inequalities,

$$|\langle T^n f, T^n g\rangle| \leq |\langle T^n f, g_i\rangle| + \epsilon \|f\|_2 \leq \epsilon \|f\|_2 + \sum_{j=1}^{n} |\langle T^n f, g_j\rangle| .$$

Since f is weak mixing, Proposition 5.14 tells us that $\text{D-lim}_{n\to\infty} |\langle T^n f, g_j\rangle| = 0$ for any j. Thus, $\text{C-sup}_{n\to\infty} |\langle T^n f, T^n g\rangle| \leq \epsilon \|f\|_2$. Since ϵ can be made arbitrarily small, it follows that this limit is zero, and hence $\langle f, g\rangle = 0$. ☐

Lemma 6.16. *The set $AP(X)$ is a closed T-invariant subspace of $L^2(X)$ which is also closed under pointwise operations $f, g \mapsto \max(f, g)$ and $f, g \mapsto \min(f, g)$.*

Proof. Since f and Tf have the same orbit, f is almost periodic if and only if Tf is. This shows that $AP(X)$ is T-invariant.

To see that $AP(X)$ is closed, take a sequence $f_m \in AP(X)$, $m = 1, 2, \ldots$, converging to f in $L^2(X)$. We would like to show that $\{T^n f : n \in \mathbb{Z}\}$ is totally bounded. Let $\epsilon > 0$ be arbitrary.

Choose m so that $\|f_m - f\|_2 < \frac{\epsilon}{2}$. Since f_m is almost periodic, we can find $g_1, \ldots, g_k \in L^2(X)$ so that for every n, we have $\|T^n f_m - g_i\| < \frac{\epsilon}{2}$ for some i, so that $\|T^n f - g_i\|_2 \leq \|T^n(f - f_m)\|_2 + \|T^n f_m - g_i\| < \epsilon$. This shows that the orbit of f is totally bounded, so that f is almost periodic. Therefore $AP(X)$ is closed.

It is easy to see that $AP(X)$ is closed under scalar multiplication. Now we show that $AP(X)$ is closed under addition, pointwise maximum and pointwise minimum. Note that these operations are all uniformly continuous. Let $\phi : L^2(X) \times L^2(X) \to L^2(X)$ be a uniformly continuous function that commutes with T, i.e., $T\phi(f,g) = \phi(Tf, Tg)$. We show that $AP(X)$ is closed under ϕ. Let $f, g \in AP(X)$. We would like to show that the orbit of $\phi(f,g)$ is totally bounded. Let $\epsilon > 0$ be arbitrary. Since ϕ is uniformly continuous, we can find a $\delta > 0$ such that $\|\phi(f_1, g_1) - \phi(f_2, g_2)\|_2 < \epsilon$ whenever $\|f_1 - f_2\|_2 < \delta$ and $\|g_1 - g_2\|_2 < \delta$. Since f and g are almost periodic, we can find $f_1, \ldots, f_k, g_1, \ldots, g_\ell \in L^2(X)$ such that for every n we have $\|T^n f - f_i\| < \frac{\delta}{2}$ and $\|T^n g - g_j\| < \frac{\delta}{2}$ for some i, j, so that $\|T^n\phi(f,g) - \phi(f_i, g_j)\|_2 = \|\phi(T^n f, T^n g) - \phi(f_i, g_j)\|_2 < \epsilon$. Hence the orbit can be covered by ϵ-balls centred at $\phi(f_i, g_j)$, $i = 1, \ldots, k$, $j = 1, \ldots, \ell$. This shows that $\phi(f,g)$ is almost periodic, and hence $AP(X)$ is closed under ϕ. □

Lemma 6.17. *Let $f \in L^2(X)$. Then $f \in WM(X)$ if and only if $\langle f, g\rangle = 0$ for all $g \in AP(X)$.*

Proof. We proved the only if direction in Lemma 6.15. To prove the converse, we'll show that if f is not weak mixing, then there is some $g \in AP(X)$ such that $\langle f, g\rangle \neq 0$.

Consider the rank-one operator $\Phi_f \colon L^2(X) \to L^2(X)$ given by $\Phi_f g = \langle f, g\rangle f$. This operator is Hilbert-Schmidt, since for an orthonormal bases (e_n) of $L^2(X)$, we have

$$\|\Phi_f\|_{HS}^2 = \sum_n \|\Phi_f e_n\|_2^2 = \sum_n |\langle e_n, f\rangle|^2 \|f\|_2^2 = \|f\|_2^4 .$$

Consider the unitary operator U on the Hilbert space $HS(L^2(X) \to L^2(X))$ of Hilbert-Schmidt operators of $L^2(X)$ given by $US = T \circ S \circ T^{-1}$. So $U\Phi_f = \Phi_{Tf}$. By the mean ergodic theorem on Hilbert spaces (Theorem 2.17), there exists some Ψ_f in the U-invariant subspace of $HS(L^2(X) \to L^2(X))$ such that $\frac{1}{N}\sum_{n=0}^{N-1} U^n\Phi_f \to \Psi_f$ in the Hilbert-Schmidt norm as $N \to \infty$. Since Ψ_f is U invariant, $\Psi_f = U\Psi_f = T \circ \Psi_f \circ T^{-1}$, so that Ψ_f commutes with T. Since Ψ_f is Hilbert-Schmidt, by Proposition 6.13, $\Psi_f g \in AP(X)$ for any $g \in L^2(X)$.

We have $U^n\Phi_f f = \Phi_{T^n f} f = \langle f, T^n f\rangle T^n f$, so $\langle U^n\Phi_f f, f\rangle = |\langle f, T^n f\rangle|^2$ Since $\frac{1}{N}\sum_{n=0}^{N-1} U^n\Phi_f$ converges to Ψ_f as $N \to \infty$ in the Hilbert-Schmidt norm, it must also converge in the weak operator topology. Therefore

$$\langle \Psi_f f, f\rangle = \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \langle U^n\Phi_f f, f\rangle = \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} |\langle f, T^n f\rangle|^2 = \operatorname*{C-lim}_{N\to\infty} |\langle f, T^n f\rangle|^2 ,$$

which is nonzero when f is not weak mixing, so that $\langle \Psi_f f, f\rangle \neq 0$ and $\Psi_f f \in AP(X)$. □

Proof of Proposition 6.14. Lemma 6.16 shows that $AP(X)$ is a closed subspace of $L^2(X)$, and Lemma 6.17 shows that $WM(X)$ is the orthogonal complement of $AP(X)$. □

From Proposition 5.13 we see that weak mixing systems are precisely the ones where every $f \in L^2(X)$ orthogonal to all constant functions are weak mixing. The decomposition of $L^2(X)$ gives us the following corollary.

Corollary 6.18. *A measure preserving system X is weak mixing if and only if the only almost periodic functions are the constant a.e. functions.*

It is interesting to compare this result to the fact that a system is ergodic if and only if the only T-invariant functions are the constant a.e. functions.

6.6 Existence of almost periodic functions

In this section, we give an alternate proof of Corollary 6.18, following [FKO82]. We do not use any material in the last subsection. The proof uses the product characterisation of weak mixing systems as given in Section 5.3. The substance of the result is essentially the same as that of Lemma 6.17. However, we do not use weak mixing functions, which were not used in [FKO82].

Proposition 6.19. *If the measure preserving system X is not weak mixing, then there exists an almost periodic function that is not constant a.e.*

Proof. We may assume that X is ergodic, since otherwise any non-constant T-invariant function f would suffice.

From the product characterisation of weak mixing systems (Proposition 5.9), we know that $X \times X$ is not ergodic. Therefore, there exists some non-constant $(T \times T)$-invariant function $K \in L^2(X \times X)$. We may assume that K has mean zero. Consider the Hilbert-Schmidt operator $\Phi_K : L^2(X) \to L^2(X)$ given by

$$\Phi_K f(y) = \int_X K(x,y) f(x) \; d\mu(x).$$

Since K is $(T \times T)$-invariant, Φ commutes with T,

$$\begin{aligned} T\Phi f(y) = \Phi f(T^{-1}y) &= \int_X K(x, T^{-1}y) f(x) \; d\mu(x) = \int_X K(Tx, y) f(x) \; d\mu(x) \\ &= \int_X K(x', y) f(T^{-1}x') \; d\mu(x') = \int_X K(x', y) Tf(x') \; d\mu(x') = \Phi T f(y) \end{aligned}$$

Thus by Proposition 6.13, $\Phi f \in AP(X)$ for all $f \in L^2(X)$. So we are done if we can find some $f \in AP(X)$ such that Φf is not constant a.e.

We claim that there exists some $f \in L^2(X)$ such that Φf is not zero a.e. Indeed, otherwise K would be orthogonal to all of $\{f \otimes g : f, g \in L^2(X)\}$, as

$$\langle K, f \otimes g \rangle_{X \times X} = \int_{X \times X} K(x,y) f(x) g(y) \; d(\mu(x) \times \mu(y)) = \int_Y \Phi f(y) g(y) \; d\mu(y).$$

Note that $x \mapsto \int_X K(x,y) \; d\mu(y)$ is T-invariant, and hence constant a.e. since X is ergodic. Since K was assumed to have mean zero, we see that $\int_X K(x,y) \; d\mu(y) = 0$ for almost every x. Therefore,

$$\int_X \Phi f(y) \; d\mu(y) = \int_{X \times X} K(x,y) f(x) \; d(\mu(x) \times \mu(y)) = \int_X \left(\int_X K(x,y) \; d\mu(y) \right) f(x) = 0.$$

So we have $\Phi f \neq 0$ on some positive measure set, but Φf has mean zero. This shows that $\Phi f \in AP(X)$ is not constant a.e. □

6.7 Existence of compact factors

In this section, we shall prove a structure theorem about a general measure preserving system X, giving the following dichotomy:

> Unless a system is completely pseudorandom, it must contain some structured piece.

To state the result precisely, we introduce the notion of *factors* in a measure preserving system. We will explore this concept in more detail in the next section, but for now it suffices to say that a factor in $(X, \mathcal{X}, \mu, T)$ is a just T-invariant sub-σ-algebra $\mathcal{X}'$. Here T-invariance means that $TE, T^{-1}E \in \mathcal{X}'$ whenever $E \in \mathcal{X}'$. In other words $X' = (X, \mathcal{X}', \mu, T)$ is another measure preserving system. We will also call X' a factor of X. A factor of X is, in some sense, a smaller system contained in X, as $\mathcal{X}'$ is a subset of $\mathcal{X}$.

A factor $\mathcal{X}'$ of $(X, \mathcal{X}, \mu, T)$ is called *trivial* if $\mu(E) \in \{0,1\}$ for all $E \in \mathcal{X}'$. It is *compact* if $(X, \mathcal{X}', \mu, T)$ is a compact measure preserving system.

Theorem 6.20. *Let $X = (X, \mathcal{X}, \mu, T)$ be a measure preserving system. Exactly one of the following is true:*

(i) (Pseudorandomness) X is weak mixing;

(ii) (Structure) X has a nontrivial compact factor.

Let
$$\mathcal{X}_{AP} = \{A \in \mathcal{X} : 1_A \in AP(X)\}.$$
Recall from Lemma 6.16 that $AP(X)$ is a T-invariant closed subspace of $L^2(X)$ that is closed under pointwise maximum and minimum. It follows that $\mathcal{X}_{AP}$ is a T-invariant sub-σ-algebra of $\mathcal{X}$. So $\mathcal{X}_{AP}$ is a factor of X, known as the *Kronecker factor*. We shall see that it is the maximal compact factor of X, characterising all almost periodic functions.

The next result gives a more precise statement about the decomposition of a function into its weak mixing and almost periodic components. Recall that the conditional expectation $\mathbf{E}(\cdot | \mathcal{X}_{AP})$ can be given as the orthogonal projection $L^2(X, \mathcal{X}, \mu) \to L^2(X, \mathcal{X}_{AP}, \mu)$.

Proposition 6.21. *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system. Let $f \in L^2(X, \mathcal{X}, \mu)$.*

(a) $f \in AP(X)$ if and only if f is $\mathcal{X}_{AP}$-measurable, i.e., $AP(X) = L^2(X, \mathcal{X}_{AP}, \mu)$.

(b) $f \in WM(X)$ if and only if $\mathbf{E}(f | \mathcal{X}_{AP}) = 0$ a.e.

(c) We can write $f = f_{AP} + f_{WM}$, where $f_{AP} = \mathbf{E}(f | \mathcal{X}_{AP}) \in AP(X)$ and $f_{WM} = f - f_{AP} \in WM(X)$.

Proof. (a) If f is $\mathcal{X}_{AP}$-measurable, then we can approximate it by simple $\mathcal{X}_{AP}$-measurable functions, which are almost periodic. Since $AP(X)$ is a closed subspace, $f \in AP(X)$.

Conversely, suppose that $f \in AP(X)$. We can use the closure properties of $AP(X)$ to show that for all $a \in \mathbb{R}$, $1_{\{x: f(x) > a\}} \in AP(X)$, so that $\{x : f(x) > a\} \in \mathcal{X}_{AP}$ and hence f is $\mathcal{X}_{AP}$-measurable. Indeed, $1_{\{x: f(x) > a\}}$ is the limit of the sequence of functions (indexed by n) $\min(\max(n(f-a), 0), 1)$ in $L^2(X)$, using dominated convergence.

(b)(c) $\mathbf{E}(\cdot | \mathcal{X}_{AP})$ is the orthogonal projection $L^2(X, \mathcal{X}, \mu) \to AP(X)$, so the results follows from knowing that $WM(X)$ is the orthogonal complement of $AP(X)$. □

Since $\mathcal{X}_{AP}$ is non-trivial if and only if $AP(X)$ contains some non-constant a.e. function, we obtain the following result as a consequence of part (a) of the above proposition.

Corollary 6.22. *In a measure preserving system $(X, \mathcal{X}, \mu, T)$, $\mathcal{X}_{AP}$ is the maximal compact factor of X, and it is non-trivial if and only if there exists some almost periodic function that is not constant a.e.*

Proof of Theorem 6.20. Both (a) and (b) cannot occur at the same time, since weak mixing and almost periodic functions are orthogonal. If X is not weak mixing, then by Corollary 6.18 or Proposition 6.19, there is some non-constant almost periodic function, and so by Corollary 6.22, $\mathcal{X}_{AP}$ is a non-trivial compact factor. □

Remark. We saw in Example 6.3 that almost periodic functions in a circle rotation system can be generated using eigenfunctions of T. This is true in general. The Kronecker factor is generated by eigenfunctions of the system. A system is weak mixing if and only if it has no non-constant eigenfunctions, and it is ergodic if and only if the eigenvalue 1 is simple.

7 Roth's theorem

At this point we are only midway towards our proof of Szemerédi's theorem. Fortunately, we have already developed enough theory to prove Szemerédi's theorem for $k = 3$, which was first established by Roth [Rot53].

Theorem 7.1 (Roth)**.** *Every subset of $\mathbb{Z}$ with positive upper density contains a 3-term arithmetic progression.*

The theorem in fact implies that the subset contains infinitely many 3-term arithmetic progressions, since we can always remove any 3-term arithmetic progressions we've found so far (without changing the upper density) and then find another one within the remaining numbers.

By the correspondence principle, it suffices to prove the following recurrence theorem.

Theorem 7.2 (Roth's theorem)**.** *Every system is SZ of level 3. In other words, let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system. Then for every $f \in L^\infty(X, \mathcal{X}, \mu)$ with $f \geq 0$ and $\mathbf{E}(f) > 0$, we have*

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

The approach is very similar to Proposition 5.16. The intuition is that we should get rid of the weak mixing part of the system and project the problem onto its Kronecker factor, which we know is SZ since it's compact.

Proposition 7.3. *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic measure preserving system. Then for any $f_1, f_2 \in L^\infty(X, \mathcal{X}, \mu)$, we have*

$$\operatorname*{C-lim}_{n\to\infty} \left(T^n f_1\ T^{2n} f_2 - T^n \mathbf{E}\left(f_1 | \mathcal{X}_{AP}\right)\ T^{2n}\mathbf{E}\left(f_2 | \mathcal{X}_{AP}\right)\right) = 0$$

in $L^2(X)$

Remark. In Proposition 5.16 we saw that $\text{C-lim}_{n\to\infty} T^n f_1\ T^{2n} f_2 = \mathbf{E}(f_1)\mathbf{E}(f_2)$ in $L^2(X)$ when the system is assumed to be weak mixing, and in fact we proved this convergence for any number of functions. Proposition 7.3 says that the factor $\mathcal{X}_{AP}$ characterises the Cesàro convergence of $T^n f_1 T^{2n} f_2$, in the sense that the limit remains the same when we project down onto the Kronecker factor $\mathcal{X}_{AP}$. The Kronecker factor is an example of a *characteristic factor.* In order to extend Proposition 7.3 to more than two functions, we will need to consider other characteristic factors, as Kronecker factor would not suffice. This is done in the recent works of Host and Kra [HK05b], which we discuss in Section 12.

Proof of Proposition 7.3. By decomposing each f_i in terms of its weak mixing and almost periodic components via Proposition 6.21, it suffices to prove that

$$\operatorname*{C-lim}_{n\to\infty} T^n f_1\ T^{2n} f_2 = 0$$

in $L^2(X)$ whenever one of f_1 and f_2 is weak mixing.

Let $v_n = T^n f_1\ T^{2n} f_2$. By van der Corput lemma (Lemma 5.10) it suffices to show that $\text{C-sup}_{h\to\infty} \text{C-sup}_{n\to\infty} \langle v_n, v_{n+h}\rangle = 0$. We have

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

Since X is ergodic, by the weak ergodic theorem (Corollary 2.14), we have

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

which converges to zero in density as $h \to \infty$ since one of f_1 and f_2 is assumed to be weak mixing. Therefore $\text{C-sup}_{h\to\infty} \text{C-sup}_{n\to\infty} \langle v_n, v_{n+h}\rangle = 0$. $\square$

Proof of Proposition 7.2. Via ergodic decomposition, we may assume that the system is ergodic. Using Proposition 7.3, we have

$$\begin{aligned}\liminf_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int_X f\,T^nf\,T^{2n}f\,d\mu &= \liminf_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int_X f\,T^n\mathbf{E}\left(f|\,\mathcal{X}_{AP}\right)T^{2n}\mathbf{E}\left(f|\,\mathcal{X}_{AP}\right)\,d\mu\\ &= \liminf_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int_X \mathbf{E}\left(f|\,\mathcal{X}_{AP}\right)T^n\mathbf{E}\left(f|\,\mathcal{X}_{AP}\right)T^{2n}\mathbf{E}\left(f|\,\mathcal{X}_{AP}\right)\,d\mu\end{aligned}$$

which is positive by Proposition 6.7 since $\mathbf{E}\left(f|\,\mathcal{X}_{AP}\right)$ is a nonnegative, positive mean, and almost periodic. □

8 Factors and extensions

8.1 Definitions and examples

In Section 6.7, we introduced the notion of factors, and analysed the Kronecker factor of a system. There we defined a factor of a measure preserving system $(X,\mathcal{X},\mu,T)$ as a T-invariant sub-σ-algebra $\mathcal{X}'$ of $\mathcal{X}$. This gives another measure preserving system $X' = (X,\mathcal{X}',\mu,T)$, along with a map $\pi\colon (X,\mathcal{X},\mu,T)\to(X,\mathcal{X}',\mu,T)$ that is measure preserving and compatible with T. We can extend the notion of factors by considering other maps $\pi\colon X\to Y$ of measure preserving systems.

Definition 8.1 (Factors and extensions)**.** Let $X=(X,\mathcal{X},\mu,T)$ and $Y=(Y,\mathcal{Y},\nu,S)$ be measure preserving systems. An *extension* (also called a *factor map*) $\pi\colon X\to Y$ is a measure preserving map (i.e., if $A\in\mathcal{Y}$, then $\pi^{-1}(A)\in\mathcal{X}$ and $\mu(\pi^{-1}(A))=\nu(A)$) that is shift-compatible, i.e., $\pi\circ T=S\circ\pi$. For a given extension $X\to Y$, we say that Y is a *factor* of X, and that X is an *extension* of Y.

In particular, note that our previous definition of a factor as a sub-σ-algebra is a special case of this more general definition, as $\pi\colon (X,\mathcal{X},\mu,T)\to(X,\mathcal{X}',\mu,T)$ is always a factor map whenever $\mathcal{X}'$ is a T-invariant sub-σ-algebra of $\mathcal{X}$.

Conversely, for any factor map $\pi : (X,\mathcal{X},\mu,T)\to(Y,\mathcal{Y},\nu,S)$, the set $\pi^{-1}(\mathcal{Y}) := \left\{\pi^{-1}(E) : E\in\mathcal{Y}\right\}$ is always a T-invariant sub-σ-algebra of $\mathcal{X}$, and hence a factor of X in our initial interpretation. Furthermore, since π is measure preserving, the pullback map $\pi^{-1}\colon\mathcal{Y}\to\pi^{-1}(\mathcal{Y})$ is one-to-one up to sets of measure zero. Under the abstraction described in Section 6.2 where we discard the underlying spaces and consider equivalence classes of measurable sets up to measure zero, our two definitions of factors, as sub-σ-algebras on one hand, and via extensions on the other hand, are actually equivalent. We will use both interpretations interchangeably when we speak of factors.

Any $\mathcal{Y}$-measurable function on Y can be viewed as a $\pi^{-1}(\mathcal{Y})$-measurable function on X via pullback. Conversely, any $\pi^{-1}(\mathcal{Y})$-measurable function on X is necessarily constant on fibres over Y, so it can be viewed as a $\mathcal{Y}$-measurable function on Y. Thus, whenever we have such a function, we speak of it as both a function on X and a function on Y interchangeably. Hopefully this won't cause too many confusions.

Example 8.2 (Product systems)**.** Let $Y=(Y,\mathcal{Y},\nu,S)$ and $Z=(Z,\mathcal{Z},\rho,R)$ be measure preserving systems. Consider the product system $Y\times Z=(Y\times Z,\mathcal{Y}\times\mathcal{Z},\nu\times\rho,S\times R)$. The projection map $\pi\colon Y\times Z\to Y$ is a factor map. Therefore Y is a factor of $Y\times Z$. This justifies the usage of the term factor.

Example 8.3 (Skew product)**.** Let $(Y,\mathcal{Y},\nu,S)$ be a system, and suppose that $(Z,\mathcal{Z},\rho)$ is a measure space. Suppose that for each $y\in Y$, there is some measure preserving map R_y on

$(Z, \mathcal{Z}, \rho)$, so that $(Z, \mathcal{Z}, \rho, R_y)$ is a measure preserving system. Assume that map $(y, z) \mapsto R_y z : Y \times Z \to Z$ is measurable. Set

$$T(y, z) = (Sy, R_y z).$$

Construct the system $(X, \mathcal{X}, \mu, T) = (Y \times Z, \mathcal{Y} \times \mathcal{Z}, \nu \times \theta, T)$. This is a *skew product*. The projection $Y \times Z \to Y$ gives a factor map $\pi \colon X \to Y$.

By a theorem of Rokhlin [Rok66], if X is ergodic, then every extension $X \to Y$ can be described as a skew product, though we will not need the full strength of this result. It gives us a good way of picturing an extension of systems.

When R_y does not depend on y, we obtain the product from Example 8.2.

In particular, when $Y = Z = \mathbb{R}/\mathbb{Z}$, and $T(y, z) = (y + a, z + y)$ for some fixed α, the system X is a called a *skew torus*. The skew torus is a good guiding example for thinking about results on factors and extensions. For instance, when a is irrational, it turns out that Y forms the Kronecker factor of the skew torus.

In the previous sections we considered properties of weak mixing and compact systems. Our goal for the next few sections is to modify these "absolute" properties of systems into "relative" properties of extensions. We refer to properties of extensions as "conditional" properties due to the use of conditional expectations as explained in the next section. We define what it means for an extension to be weak mixing or compact, and develop these concepts in a way parallel to how we studied weak mixing and compact systems in earlier sections. Indeed, many of the results will be relativised versions of our earlier results, and the proofs are obtained by modifying previous proofs. We show that the SZ property "lifts" via weak mixing and compact extensions, meaning that if $X \to Y$ is such an extension and Y is SZ, then X is SZ as well. We also show that if an extension is not weak mixing, then there is always an intermediate compact factor.

8.2 Conditional expectations

Suppose we have a probability space $(X, \mathcal{X}, \mu)$, and a measurable function $f : X \to \mathbb{R}$. Let $\mathcal{X}'$ be a sub-σ-algebra of $\mathcal{X}$. Then we have the inclusion of Hilbert spaces

$$L^2(X, \mathcal{X}', \mu) \subset L^2(X, \mathcal{X}, \mu).$$

The conditional expectation is the orthogonal projection

$$\mathbf{E}\left(\cdot \middle| \mathcal{X}'\right) : L^2(X, \mathcal{X}, \mu) \to L^2(X, \mathcal{X}', \mu).$$

So in particular, $\mathbf{E}\left(f \middle| \mathcal{X}'\right)$ is the unique function in $L^2(X, \mathcal{X}, \mu)$ such that

$$\langle f, g \rangle = \left\langle \mathbf{E}\left(f \middle| \mathcal{X}'\right), g \right\rangle, \quad \text{for all } g \in L^2(X, \mathcal{X}', \mu),$$

or equivalently,

$$\int_X fg \, d\mu = \int_X \mathbf{E}\left(f \middle| \mathcal{X}'\right) g \, d\mu, \quad \text{for all } g \in L^2(X, \mathcal{X}', \mu). \tag{8.1}$$

We can also extend the definition of conditional expectation to all of $L^1(X, \mathcal{X}, \mu)$, though we omit the details here. From (8.1), we have

$$\mathbf{E}\left(fg \middle| \mathcal{X}'\right) = \mathbf{E}\left(f \middle| \mathcal{X}'\right) g, \quad \text{for all } f \in L^2(X, \mathcal{X}, \mu), g \in L^2(X, \mathcal{X}', \mu), \tag{8.2}$$

so that any $\mathcal{X}'$-measurable functions can be pulled outside of the conditional expectation.

Consider an extension $\pi : (X, \mathcal{X}, \mu, T) \to (Y, \mathcal{Y}, \nu, S)$ of measure preserving systems. The factor Y induces a sub-σ-algebra $\pi^{-1}(\mathcal{Y})$ of $\mathcal{X}$, and therefore we can consider the conditional expectation relative to Y defined by

$$\mathbf{E}\left(f \middle| Y\right) := \mathbf{E}\left(f \middle| \pi^{-1}(\mathcal{Y})\right).$$

The function $\mathbf{E}(f|Y)$ is $\pi^{-1}(\mathcal{Y})$-measurable function on X, and hence it can also be viewed as a $\mathcal{Y}$-measurable function on Y, as remarked in the earlier section.

Here is another way to construct conditional expectations with respect to some extension $\pi : (X, \mathcal{X}, \mu, T) \to (Y, \mathcal{Y}, \nu, S)$. The pullback gives a map of Hilbert spaces $\pi^* \colon L^2(Y, \mathcal{Y}, \nu) \to L^2(X, \mathcal{X}, \mu)$, and its adjoint map is the conditional expectation $\mathbf{E}(\cdot|Y) = \pi_* \colon L^2(X, \mathcal{X}, \mu) \to L^2(Y, \mathcal{X}, \mu)$. In other words, $\mathbf{E}(\cdot|Y)$ is characterised by

$$\langle f, \pi^* g\rangle_X = \langle \mathbf{E}(f|Y), g\rangle_Y . \tag{8.3}$$

or every $f \in L^2(X, \mathcal{X}, \mu)$ and $g \in L^2(Y, \mathcal{Y}, \nu)$.

It is easy to check that the shift maps are also compatible with conditional expectation, in the sense that

$$\mathbf{E}(Tf|Y) = S\mathbf{E}(f|Y).$$

We sometimes write $S\mathbf{E}(f|Y)$ as $T\mathbf{E}(f|Y)$, viewing $\mathbf{E}(f|Y)$ as a function on X. This identity is a generalisation of $\mathbf{E}(Tf) = \mathbf{E}(f)$ in the absolute case.

Example 8.4 (Conditional expectation for skew products)**.** Let $X \to Y$ be the skew product from Example 8.3. Then for any $f \in L^2(X)$, written as $f(y, z)$ for $y \in Y$ and $z \in Z$, we have $\mathbf{E}(f|Y) \in L^2(Y)$ given by

$$\mathbf{E}(f|Y)(y) = \int_Z f(y, z)\ d\rho(z).$$

8.3 Hilbert modules

Previously when we had worked with measure preserving systems $(X, \mathcal{X}, \mu, T)$, the Hilbert space $L^2(X, \mathcal{X}, \mu)$ played a important role. We now need a relative version of this Hilbert space. The idea is that instead of working with a vector space, with the constants being $\mathbb{R}$, we work with a module over $L^\infty(Y)$, with elements $L^\infty(Y)$ behaving as if there were the constants in a Hilbert space.

Let $Y = (Y, \mathcal{Y}, \nu, S)$ be a factor of $X = (X, \mathcal{X}, \mu, T)$. Define the *Hilbert module* $L^2(X, \mathcal{X}, \mu| Y, \mathcal{Y}, \mu)$ (usually written as $L^2(X|Y)$) over the *commutative von Neumann algebra* $L^\infty(Y, \mathcal{Y}, \nu)$ to be the space of all $f \in L^2(X, \mathcal{X}, \mu)$ such that the conditional norm

$$\|f\|_{L^2(X|Y)} := \mathbf{E}\left(|f|^2 \middle| Y\right)^{1/2}$$

lies in $L^\infty(Y)$ (note that $\|f\|_{L^2(X|Y)} \in L^2(Y)$ for any $f \in L^2(X)$). We can make $L^2(X|Y)$ a module over $L^\infty(Y)$, since by (8.2) we have

$$\mathbf{E}(cf + dg|Y) = c\mathbf{E}(f|Y) + d\mathbf{E}(g|Y), \quad \text{for all } f, g \in L^2(X|Y),\ c, d \in L^\infty(Y).$$

We also have inner products in $L^2(X|Y)$ defined by

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

This inner product initially only lies in $L^1(Y)$, but it actually lies in $L^\infty(Y)$ due to the following relativised version of Cauchy-Schwarz inequality.

Theorem 8.5 (Conditional Cauchy-Schwarz inequality)**.** *Let $X \to Y$ be an extension. Then for any $f, g \in L^2(X|Y)$ we have*

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

almost everywhere.

Proof. We relativise the standard proof of Cauchy-Schwarz, with $L^2(Y)$ taking the role of constants. Let $c = \|g\|_{L^2(X|Y)}$ and $d = \|f\|_{L^2(X|Y)}$. Then almost everywhere we have

$$\begin{aligned}
0 &\leq \langle cf - dg, cf - dg \rangle_{X|Y} \\
&= c^2 \langle f, f \rangle_{X|Y} + d^2 \langle g, g \rangle_{X|Y} - 2cd \langle f, g \rangle_{X|Y} \\
&= 2 \|f\|^2_{L^2(X|Y)} \|g\|^2_{L^2(X|Y)} - 2 \langle f, g \rangle_{X|Y} \|f\|_{L^2(X|Y)} \|g\|_{L^2(X|Y)} \\
&= 2 \|f\|_{L^2(X|Y)} \|g\|_{L^2(X|Y)} \left(\|f\|_{L^2(X|Y)} \|g\|_{L^2(X|Y)} - \langle f, g \rangle_{X|Y} \right).
\end{aligned}$$

Note that $\langle f, g \rangle_{X|Y} = 0$ a.e. on the set where $\|f\|_{L^2(X|Y)} \|g\|_{L^2(X|Y)} = 0$. It follows that $\|f\|_{L^2(X|Y)} \|g\|_{L^2(X|Y)} \geq \langle f, g \rangle_{X|Y}$ a.e., and since the left-hand side is nonnegative, we can change the right hand side to an absolute value, giving us the desired inequality. $\square$

8.4 Disintegration measures

In this section we present a somewhat different viewpoint of extensions by introducing disintegration measures, following [FKO82, Fur77]. Although we do not strictly need to use disintegration in our proof of Szemerédi's theorem, this viewpoint gives us a somewhat more concrete way of thinking about extensions and conditional expectations, motivated by the example of the skew product. Disintegration measures also allow us to give an alternative formulation of weak mixing extensions, leading to an alternate approach to one of the steps of the proof of Szemerédi's theorem.

Recall the skew product from Example 8.3. In a skew product $X = (Y \times Z, \mathcal{Y} \times \mathcal{Z}, \nu \times \rho, T)$, where $T(y, z) = (Sy, R_y z)$, we can write down the conditional expectation relative to Y for a function on X quite explicitly as an integral, as we did in Example 8.4. It turns out that, under mild regularity hypotheses, the conditional expectation can always be expressed this way.

In the skew product example, the conditional expectation $\mathbf{E}(f|Y)$, evaluated at $y \in Y$, can be found by taking the mean of f over the fibre of y under $X \to Y$. The fibre of y is another copy of Z, and we can integrate f using the measure ρ on Z with the coordinate y fixed. Equivalently,

$$\mathbf{E}(f|Y)(y) = \int_X f \, d(\delta_y \times \rho)$$

where δ_y is the Dirac delta measure supported on the point $y \in Y$.

In general, given an extension $\pi \colon X \to Y$, we show that for each $y \in Y$, we can assign some *disintegration measure* μ_y on X, so that conditional expectations can be computed by integrating with respect to μ_y. In other words, we would like

$$\mathbf{E}(f|Y)(y) = \int_X f \, d\mu_y.$$

By comparing with the skew product, we see that μ_y plays the role of the measure $\delta_y \times \rho$ on the fibre of y.

To state result about the existence of disintegration measures, we need to state some hypotheses regarding regularity. A measurable space $(X, \mathcal{X})$ is said to be *regular* if there exists a compact metrisable topology on X for which $\mathcal{X}$ is the Borel σ-algebra. In particular, every topological measure preserving system is regular, and as is every system that arises in the proof of Szemerédi's theorem. In fact, every separable measure space is equivalent up to zero measure sets to a regular measure space ([Fur81, Prop. 5.3]).

Definition 8.6 (Disintegration measure)**.** Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \mu)$ be probability spaces, and let $\pi \colon X \to Y$ be a morphism of probability spaces (i.e., a measure preserving map). Let $\mathcal{M}(X)$ denote the set of measures on X. Then a *disintegration measure* for $\pi \colon X \to Y$ is an assignment $\mu_y \in \mathcal{M}(X)$ to each $y \in Y$ so that for any $A \in \mathcal{X}$, the map $y \mapsto \mu_y(A) : Y \to [0,1]$ is measurable, and $f \in L^1(X, \mathcal{X}, \mu)$, we have $f \in L^1(X, \mathcal{X}, \mu_y)$ for a.e. $y \in Y$, and

$$\mathbf{E}\left(f \middle| Y \right)(y) = \int_X f \, d\mu_y$$

for a.e. $y \in Y$.

Theorem 8.7 (Disintegration). *Let $(X, \mathcal{X}, \mu) \to (Y, \mathcal{Y}, \mu)$ be a morphism of probability spaces, and suppose that $(X, \mathcal{X})$ is regular, then there exists a disintegration measure μ_y, which is unique for a.e. $y \in Y$.*

The existence of μ_y can be deduced using Riesz representation theorem. For proof of the result see [FKO82, Tao09].

Like in the skew product case, each μ_y is supported on the fibre $\pi^{-1}(y)$ a.e. Indeed, using (8.3) we see that for any $A \in \mathcal{Y}$,

$$\left\langle \int_X 1_{\pi^{-1}(A)} \, d\mu_y, 1_A \right\rangle_Y = \left\langle \mathbf{E}\left(1_{\pi^{-1}(A)} \middle| Y \right), 1_A \right\rangle_Y = \left\langle 1_{\pi^{-1}(A)}, 1_{\pi^{-1}(A)} \right\rangle_X = \mu\left(\pi^{-1}(A)\right) = \nu(A).$$

Note that a factor map between measure preserving systems is always a morphism of probability spaces, so we can speak of disintegration measures for extensions of systems. The following proposition shows that the disintegration measures μ_y must be compatible with the shift maps.

Proposition 8.8. *Let $(X, \mathcal{X}, \mu, T)$ be an extension of $(Y, \mathcal{Y}, \nu, S)$, and let μ_y be a disintegration measure. Then we have $\mu_{Sy} = T\mu_y$ a.e. $y \in Y$.*

Proof. For every $f \in L^1(X, \mathcal{X}, \mu)$, we have for a.e. $y \in Y$,

$$\int_X f \, d\mu_{Sy} = \mathbf{E}\left(f \middle| Y \right)(Sy) = \mathbf{E}\left(T^{-1} f \middle| Y \right)(y) = \int_X T^{-1} f \, d\mu_y = \int_X f \, d(T\mu_y).$$

It follows that $\mu_{Sy} = T\mu_y$ a.e. $y \in Y$. □

9 Weak mixing extensions

In this section, we relativise the concepts in Section 5 related to weak mixing from systems to extensions. In particular, we show that the SZ property lifts through weak mixing extensions.

9.1 Ergodic and weak mixing extensions

In the absolute setting (i.e., systems), we work with functions in $L^2(X)$, and $\mathbb{R}$ plays the role of constants, whereas in the relative/conditional setting (i.e., extensions), we work with functions in $L^2\left(X \middle| Y\right)$, and $L^\infty(Y)$ play the role of constants.

The first property that we relativise is that of ergodicity. Recall that a system X is ergodic if every T-invariant function f (i.e., $Tf = f$) is constant a.e.

Definition 9.1 (Ergodic extension). An extension $X \to Y$ of measure preserving systems is called *ergodic* if every T-invariant function in $f \in L^2\left(X \middle| Y\right)$ lies in $L^\infty(Y)$.

It is easy to see that if $X \to Y$ is an ergodic extension and Y is an ergodic system, then X must be ergodic as well.

Proposition 9.2 (Conditional weak ergodic theorem). *An extension $X \to Y$ is ergodic if and only if*

$$\underset{n\to\infty}{\text{C-lim}} \Big(\langle T^n f, g \rangle_{X|Y} - S^n \mathbf{E}\left(f \middle| Y \right) \mathbf{E}\left(g \middle| Y \right) \Big) = 0 \tag{9.1}$$

in $L^2(Y)$, for every $f, g \in L^2\left(X \middle| Y\right)$.

Proof. First we show that (9.1) implies that $X \to Y$ is an ergodic extension. Suppose $f \in L^2(X|Y)$ satisfies $Tf = f$. Then $S\mathbf{E}(f|Y) = \mathbf{E}(Tf|Y) = \mathbf{E}(f|Y)$ as well. Then the sequence in (9.1) is independent of n, and thus $\langle f, g\rangle_{X|Y} - \mathbf{E}(f|Y)\mathbf{E}(g|Y) = 0$ a.e. for every $g \in L^2(X|Y)$. That is, $\langle f - \mathbf{E}(f|Y), g\rangle_{X|Y} = 0$ a.e. Hence $f = \mathbf{E}(f|Y) \in L^\infty(Y)$. This shows that $X \to Y$ is an ergodic extension.

Now suppose that $X \to Y$ is an ergodic extension. Let

$$f_N = \frac{1}{N}\sum_{n=0}^{N-1} T^n(f - \mathbf{E}(f|Y)).$$

Applying the mean ergodic theorem in $L^2(X)$, we see that there exists some T-invariant $\widetilde{f} \in L^2(X)$ such that $f_N \to \widetilde{f}$ in $L^2(X)$. Since $\|\cdot\|_{L^2(X)} = \left\|\|\cdot\|_{L^2(X|Y)}\right\|_{L^2(Y)}$, we see that $\|f_N\|_{L^2(X|Y)} \to \|\widetilde{f}\|_{L^2(X|Y)}$ in $L^2(Y)$. Since $\left\|\|f_N\|_{L^2(X|Y)}\right\|_{L^\infty(Y)} \le 2\left\|\|f\|_{L^2(X|Y)}\right\|_{L^\infty(Y)}$ uniformly for N, it follows that $\left\|\|\widetilde{f}\|_{L^2(X|Y)}\right\|_{L^\infty(Y)} \le 2\left\|\|f\|_{L^2(X|Y)}\right\|_{L^\infty(Y)}$, so that $\widetilde{f} \in L^2(X|Y)$.

Since $X \to Y$ is an ergodic extension and $T\widetilde{f} = \widetilde{f}$, it follows that $\widetilde{f} = \mathbf{E}\left(\widetilde{f}\,\middle|\,Y\right)$ a.e. Note that $\mathbf{E}(f_N|Y) = 0$ a.e. for every N, so $\mathbf{E}\left(\widetilde{f}\,\middle|\,Y\right) = 0$ a.e., and hence $\widetilde{f} = 0$ a.e. Thus $f_N \to 0$ in $L^2(X)$, so $\langle f_N, g\rangle_{X|Y} \to 0$ in $L^2(Y)$ for every $g \in L^2(X|Y)$, which is equivalent to (9.1). $\square$

Definition 9.3 (Weak mixing extension)**.** An extension $X \to Y$ of measure preserving systems is called a *weak mixing extension* if

$$\operatorname*{D-lim}_{n\to\infty}\left(\langle T^n f, g\rangle_{X|Y} - S^n\mathbf{E}(f|Y)\mathbf{E}(g|Y)\right) = 0 \tag{9.2}$$

in $L^2(Y)$, for every $f, g \in L^2(X|Y)$.

Previously we saw that weak mixing systems are always ergodic. The same is true for extensions, as convergence in density implies Cesàro convergence for bounded sequences.

Proposition 9.4. *If $X \to Y$ is a weak mixing extension, then it is an ergodic extension.*

Example 9.5. Let Y and Z be measure preserving systems, and consider the product system $X = Y \times Z$. The extension $X \to Y$ induced by projection is ergodic (resp. weak mixing) if and only if Z is ergodic (resp. weak mixing).

9.2 Conditionally weak mixing functions

Definition 9.6 (Conditionally weak mixing function)**.** Let $X \to Y$ be an extension of measure preserving systems. A function $f \in L^2(X|Y)$ is *conditionally weak mixing* relative to Y if $\operatorname{D-lim}_{n\to\infty}\langle T^n f, f\rangle_{X|Y} = 0$ in $L^2(Y)$.

If $f \in L^2(X|Y)$ is conditionally weak mixing relative to Y, then f, viewed as a function in $L^2(X)$, is weak mixing in X. Indeed, this statement follows from taking the Cesàro limit of the following inequality.

$$|\langle T^n f, f\rangle| = \left|\int_Y \langle T^n f, f\rangle_{X|Y}\, d\nu\right| \le \left\|\langle T^n f, f\rangle_{X|Y}\right\|_{L^2(Y)}.$$

In Proposition 5.13 we saw that a system is weak mixing if and only if every mean zero function is a weak mixing function. A similar statement holds for extensions.

Proposition 9.7. *An extension $X \to Y$ of measure preserving systems is weak mixing if and only if every $f \in L^2(X|Y)$ with conditional mean zero (i.e., $\mathbf{E}(f|Y) = 0$ a.e.) is conditionally weak mixing.*

Here is the relative version of Proposition 5.14.

Proposition 9.8. *Let $X \to Y$ be an extension of measure preserving systems. Let $f \in L^2(X|Y)$ be conditionally weak mixing and $g \in L^2(X|Y)$. Then*

$$\operatorname*{D-lim}_{n\to\infty} \langle T^n f, g\rangle_{X|Y} = 0 \quad \textit{and} \quad \operatorname*{D-lim}_{n\to\infty} \langle f, T^n g\rangle_{X|Y} = 0$$

in $L^2(Y)$.

Proof. As in the proof of Proposition 5.14, it suffices to prove the first limit.

From Proposition 5.5, it suffices to prove that $\text{C-lim}_{n\to\infty} \left\| \langle T^n f, g\rangle_{X|Y} \right\|^2_{L^2(Y)} = 0$. We have

$$\begin{aligned}
\frac{1}{N}\sum_{n=0}^{N-1} \left\| \langle T^n f, g\rangle_{X|Y} \right\|^2_{L^2(Y)} &= \frac{1}{N}\sum_{n=0}^{N-1} \int_Y \left| \langle T^n f, g\rangle_{X|Y} \right|^2 \, d\nu \\
&= \int_Y \left\langle \frac{1}{N}\sum_{n=0}^{N-1} \langle T^n f, g\rangle_{X|Y} T^n f, g \right\rangle_{X|Y} d\nu \\
&= \left\langle \frac{1}{N}\sum_{n=0}^{N-1} \langle T^n f, g\rangle_{X|Y} T^n f, g \right\rangle_{X}. \qquad (9.3)
\end{aligned}$$

By Cauchy-Schwarz inequality, to prove that the above quantity converges to 0 as $N \to \infty$, it suffices to show that

$$\operatorname*{C-lim}_{n\to\infty} \langle T^n f, g\rangle_{X|Y} T^n f = 0$$

in $L^2(X)$. Using van der Corput lemma (Lemma 5.10), it suffices to show that

$$\operatorname*{C-lim}_{h\to\infty} \operatorname*{C-sup}_{n\to\infty} \left\langle \langle T^n f, g\rangle_{X|Y} T^n f, \left\langle T^{n+h} f, g\right\rangle_{X|Y} T^{n+h} f \right\rangle = 0. \qquad (9.4)$$

We have

$$\begin{aligned}
&\left| \left\langle \langle T^n f, g\rangle_{X|Y} T^n f, \left\langle T^{n+h} f, g\right\rangle_{X|Y} T^{n+h} f \right\rangle \right| \\
&\quad = \left| \int_Y \langle T^n f, g\rangle_{X|Y} \left\langle T^{n+h} f, g\right\rangle_{X|Y} \left\langle T^n f, T^{n+h} f\right\rangle_{X|Y} \, d\nu \right| \\
&\quad \le \left\| \langle T^n f, g\rangle^2_{X|Y} \left\langle T^{n+h} f, g\right\rangle_{X|Y} \right\|_{L^2(Y)} \left\| \left\langle T^n f, T^{n+h} f\right\rangle_{X|Y} \right\|_{L^2(Y)}.
\end{aligned}$$

By conditional Cauchy-Schwarz inequality (Proposition 8.5) $\left\| \langle T^n f, g\rangle_{X|Y} \left\langle T^{n+h} f, g\right\rangle_{X|Y} \right\|_{L^2(Y)}$ is bounded by $\left\| \|f\|_{L^2(X|Y)} \right\|^2_{L^\infty(Y)} \left\| \|g\|_{L^2(X|Y)} \right\|^2_{L^\infty(Y)}$. By T-invariance, we have $\left\| \left\langle T^n f, T^{n+h} f\right\rangle_{X|Y} \right\|_{L^2(Y)} = \left\| \left\langle f, T^h f\right\rangle_{X|Y} \right\|_{L^2(Y)}$. So (9.4) follows from $\text{D-lim}_{h\to\infty} \left\langle f, T^h f\right\rangle_{X|Y} = 0$ as f is conditionally weak mixing. □

Proof of Proposition 9.7. If $X \to Y$ is weak mixing, then taking $g = f$ any conditional mean zero function in (9.2) shows that f is conditionally weak mixing.

Conversely, suppose that every conditional mean zero function in $L^2(X|Y)$ is conditionally weak mixing. By Proposition 9.8 we have $\text{D-lim}_{n\to\infty} \langle T^n(f - \mathbf{E}(f|Y)), g\rangle_{X|Y} = 0$ for every $f, g \in L^2(X|Y)$, which expands into (9.2). □

9.3 Fibre product characterisation

We saw in Proposition 5.9 that weak mixing systems X can be characterised by $X \times X$ being ergodic. In this section, we explain how weak mixing extensions can be similarly characterised in terms of relative products, also known as fibre products. As in Section 5.3, this section is not strictly necessary for our proof of Szemerédi's theorem, although it does provide an alternative to one of the steps.

Definition 9.9 (Fibre product)**.** Let $Y = (Y, \mathcal{Y}, \nu, S)$ be a measure preserving system, and let $\pi \colon X = (X, \mathcal{X}, \mu, T) \to Y$ and $\pi' \colon X' = (X', \mathcal{X}', \mu', T') \to Y$ be two extensions, and suppose that we have disintegration measures μ_y and μ'_y for π and π' respectively. Then the *fibre product* of X and X' relative to Y is defined to be $X \times_Y X' = (X \times_Y X', \mathcal{X} \times_Y \mathcal{X}', \mu \times_Y \mu', T \times_Y T')$, where the underlying space $X \times_Y X'$ is the set theoretic fibre product, defined as

$$X \times_Y X' = \big\{(x, x') \in X \times X' : \pi(x) = \pi'(x') \in Y\big\} = \bigcup_{y \in Y} \pi^{-1}(y) \times \pi'^{-1}(y) \subset X \times X'.$$

The σ-algebra $\mathcal{X} \times_Y \mathcal{X}'$ is the restriction of $\mathcal{X} \times \mathcal{X}'$ from $X \times X'$ to $X \times_Y X'$. The measure $\mu \times_Y \mu'$ is given by its disintegration as $(\mu \times_Y \mu')_y = \mu_y \times \mu'_y$ for $y \in Y$, which is supported on $\pi^{-1}(y) \times \pi'^{-1}(y)$. And the action $T \times T'$ is given by $T \times T'(x, x') = (Tx, T'x')$.

Remark. This is basically the definition of fibre products given in [FKO82]. In [Fur81], the fibre product $X \times_Y X'$ is defined as $(X \times X', \mathcal{X} \times \mathcal{X}', \mu \times_Y \mu', T \times T')$, where the underlying space is the usual product set as opposed to the set-theoretic fibre product $X \times_Y X'$. Since the measure $\mu \times_Y \mu'$ is supported on the subspace $X \times_Y X' \subset X \times X'$, these two constructions are essentially equivalent, as we do not care about sets of measure zero. Sometimes we'll extend the underlying space in the system fibre product to $X \times X$ in order to simplify notation, as we have

$$\int_{X \times_Y X} f \; d(\mu \times_Y \mu') = \int_{X \times X} f \; d(\mu \times_Y \mu') = \int_Y \left(\int_{X \times X'} f \; d(\mu_y \times \mu'_y) \right) d\nu(y)$$

for any for any measurable $f \colon X \times X' \to \mathbb{R}$.

We have the following diagram of extensions of systems.

$$\begin{tikzcd} X \times_Y X' \arrow[r] \arrow[d] & X' \arrow[d, "\pi'"] \\ X \arrow[r, "\pi"] & Y \end{tikzcd}$$

It can be shown that the diagram commutes, thereby giving a natural extension $X \times_Y X' \to Y$. When Y is trivial (e.g., a single point), then $X \times_Y X'$ is just the usual product $X \times X'$.

The fibre product is somewhat more intuitive when we work with skew products (Example 8.3).

Example 9.10 (Fibre product of skew products)**.** Let $Y = (Y, \mathcal{Y}, \nu, S)$ be a measure preserving system. Let X denote the skew product $(Y \times Z, \mathcal{Y} \times \mathcal{Z}, \nu \times \rho, S \times R_y)$ and X' the skew product $(Y \times Z', \mathcal{Y} \times \mathcal{Z}', \nu \times \rho', S \times R'_y)$. Then the fibre product of X and X' relative to Y is given by

$$X \times_Y X' = (Y \times Z \times Z', \mathcal{Y} \times \mathcal{Z} \times \mathcal{Z}', \nu \times \rho \times \rho', S \times R_y \times R'_y).$$

Proposition 9.11. *Let $X \times_Y X'$ be a fibre product of measure preserving systems. Let $f \in L^2(X)$ and $f' \in L^2(X')$. Let $f \otimes f' \in L^2(X \times_Y X')$ denote the function given by $f \otimes f'(x, x') = f(x) f'(x')$. Then*

$$\mathbf{E}\left(f \otimes f' \middle| Y\right) = \mathbf{E}\left(f \middle| Y\right) \mathbf{E}\left(f' \middle| Y\right)$$

almost everywhere. Note that the first conditional expectation is taken with respect to $X \times_Y X' \to Y$, the second with respect to $X \to Y$ and the third with respect to $X' \to Y$.

Proof. Using the disintegration measures, we find that

$$\begin{aligned}\mathbf{E}\left(f \otimes f' \middle| Y\right)(y) &= \int_{X \times_Y X'} f \otimes f' d(\mu \times_Y \mu')_y = \int_{X \times X'} f \otimes f' d(\mu_y \times \mu'_y) \\ &= \int_X f d\mu_y \int_{X'} f' \, d\mu'_y = \mathbf{E}\left(f \middle| Y\right)(y) \mathbf{E}\left(f' \middle| Y\right)(y).\end{aligned}$$

for a.e. $y \in Y$. $\square$

Proposition 9.12. *Let $X \to Y$ be an extension of measure preserving systems, with fibre product $X \times_Y X$. The following are equivalent.*

(a) $X \to Y$ is a weak mixing extension;

(b) $X \times_Y X \to Y$ is a weak mixing extension;

(c) $X \times_Y X \to Y$ is an ergodic extension.

Proof. (a) $\implies$ (b): Since the expression inside the limit (9.2) is bilinear in f and g, it suffices to verify this limit for spanning set of a dense subspace of functions. Thus, to check that $X \times_Y X \to Y$ is a weak mixing extension, it suffices to check that for every $f_1 \otimes f_2, g_1 \otimes g_2 \in L^2\left(X \times_Y X \middle| Y\right)$, where $f_1, f_2, g_1, g_2 \in L^2\left(X \middle| Y\right)$, we have

$$\operatorname*{D-lim}_{n\to\infty} \Big(\langle (T \times T)^n (f_1 \otimes f_2), g_1 \otimes g_2 \rangle_{X|Y} - S^n \mathbf{E}\left(f_1 \otimes f_2 \middle| Y\right) \mathbf{E}\left(g_1 \otimes g_2 \middle| Y\right) \Big) = 0 \tag{9.5}$$

in $L^2(Y)$. Using Proposition 9.11, we have $\langle (T \times T)^n (f_1 \otimes f_2), g_1 \otimes g_2 \rangle_{X|Y} = \langle T^n f_1, g_1 \rangle_{X|Y} \langle T^n f_2, g_2 \rangle_{X|Y}$, $\mathbf{E}\left(f_1 \otimes f_2 \middle| Y\right) = \mathbf{E}\left(f_1 \middle| Y\right) \mathbf{E}\left(f_2 \middle| Y\right)$, and $\mathbf{E}\left(g_1 \otimes g_2 \middle| Y\right) = \mathbf{E}\left(g_1 \middle| Y\right) \mathbf{E}\left(g_2 \middle| Y\right)$. Thus (9.5) follows from (9.2) applied to f_1, g_1, and f_2, g_2.

(b) $\implies$ (c): This follows from Proposition 9.4.

(c) $\implies$ (a): Note that (9.2) can be rewritten as

$$\operatorname*{D-lim}_{n\to\infty} \langle T^n (f - \mathbf{E}\left(f \middle| Y\right)), g \rangle_{X|Y} = 0,$$

in $L^2(Y)$ for every $f, g \in L^2\left(X \middle| Y\right)$. Thus to show that $X \to Y$ is weak mixing, it suffices to show that for every $f, g \in L^2\left(X \middle| Y\right)$ with f conditional mean zero, we have $\operatorname{D-lim}_{n\to\infty} \langle T^n f, g \rangle_{X|Y} = 0$ in $L^2(Y)$. By Proposition 5.5, it suffices to show that $\operatorname{C-lim}_{n\to\infty} \left\| \langle T^n f, g \rangle_{X|Y} \right\|^2_{L^2(Y)} = 0$. We have

$$\begin{aligned}\operatorname*{C-lim}_{n\to\infty} \left\| \langle T^n f, g \rangle_{X|Y} \right\|^2_{L^2(Y)} &= \operatorname*{C-lim}_{n\to\infty} \int_Y \langle T^n f, g \rangle^2_{X|Y} \, d\nu \\ &= \operatorname*{C-lim}_{n\to\infty} \int_Y \langle (T \times T)^n f \otimes f, g \otimes g \rangle_{X \times_Y X|Y} \, d\nu,\end{aligned}$$

which equals to zero by the conditional weak ergodic theorem (Proposition 9.2) since $X \times_Y X \to Y$ is ergodic and $\mathbf{E}\left(f \otimes f \middle| Y\right) = E\left(f \middle| Y\right)^2 = 0$. $\square$

9.4 Multiple recurrence

In this section we prove that the SZ property lifts through weak mixing extensions.

Theorem 9.13. *Let $X \to Y$ be a weak mixing extension. If Y is SZ, then X is SZ.*

Here is a relativised version of Proposition 5.16, whose proof can be obtained by modifying the proof of Proposition 5.16 similar to how we relativised the previous results. We omit the details this time.

Proposition 9.14. *Let $X \to Y$ be a weak mixing extension. Let $k \geq 1$. Let $a_1, \ldots, a_k \in \mathbb{Z}$ be distinct non-zero integers, and let $f_1, \ldots, f_k \in L^\infty(X)$. Then*

$$\operatorname*{C-lim}_{n\to\infty} (T^{a_1 n} f_1 \cdots T^{a_k n} f_k - T^{a_1 n}\mathbf{E}(f_1|Y) \cdots T^{a_k n}\mathbf{E}(f_k|Y)\ d\nu) = 0$$

in $L^2(X)$.

Corollary 9.15. *Same assumptions as Proposition 9.14. Then*

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

Proof of Theorem 9.13. Let $f \in L^\infty(X)$ be a nonnegative function with positive mean. Then $\mathbf{E}(f|Y) \in L^\infty(Y)$ is also nonnegative with the same positive mean. Since Y is assumed to be SZ, we have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_Y \mathbf{E}(f|Y)\ S^n\mathbf{E}(f|Y)\ S^{2n}\mathbf{E}(f|Y) \cdots S^{kn}\mathbf{E}(f|Y)\ d\nu > 0.$$

Then it follows from Corollary 9.15 that

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

Therefore X is SZ. $\square$

10 Compact extensions

In this section we relativise the ideas from Section 6 in order to analyse compact extensions. We prove that the SZ property lifts via compact extensions, and also that every extension that is not weak mixing contains an intermediate non-trivial compact extension.

10.1 Definition and examples

To define the conditional versions of almost periodic functions, we need a conditional analogue of precompact sets.

Definition 10.1. A subset E of $L^2(X|Y)$ is said to be *conditionally precompact* if for every $\epsilon > 0$, we can find $f_1, \ldots, f_d \in L^2(X|Y)$ so that $\min_{1\leq i\leq d} \|f - f_i\|_{L^2(X|Y)} < \epsilon$ a.e.-y for every $f \in E$.

In Tao [Tao09], a different (but equivalent) definition for conditionally precompactness is used. Again motivated by the non-relative case, where one can show that a subset of $L^2(X)$ (or any Hilbert space in general) is compact if and only if it lies within the ϵ-neighbourhood of some bounded finite-dimensional *zonotope* $\{c_1 f_1 + \cdots + c_d f_d : c_1, \cdots, c_d \in \mathbb{R}, |c_1|, \ldots, |c_d| \leq 1\}$. Relativising, we can define a *finitely generated module zonotope* to be a subset of $L^2(X|Y)$ of the form $\left\{c_1 f_1 + \cdots + c_d f_d : c_1, \ldots, c_d \in L^\infty(Y), \|c_1\|_{L^\infty(Y)}, \ldots, \|c_d\|_{L^\infty(Y)} \leq 1\right\}$. The definition given in Tao [Tao09] is that a subset E of $L^2(X|Y)$ is conditionally precompact if for every ϵ there exists some finitely generated module zonotope Z of $L^2(X|Y)$ such that E lies within the ϵ-neighbourhood of Z with respect to the norm $\left\| \|\cdot\|_{L^2(X|Y)} \right\|_{L^\infty(Y)}$.

These two definition are equivalent. If E can be approximated by $f_1, \ldots, f_d \in L^2(X|Y)$ as in the first definition, then these functions generate the desired module zonotope in the second definition. Conversely, if we can approximate E by a module zonotope generated by

$f_1, \ldots, f_d \in L^2(X|Y)$, then we can take a collection of functions of the form $b_1 f_1 + \cdots + b_d f_d$ where the b_i's are constants of the form $\frac{k}{M}$ for some large M and integer k with $|k| \leq M$. By choosing the appropriate b_i's at each point in Y, we can pointwise approximate any function lying inside the polytope, thereby getting the desired set of functions as in the first definition.

Definition 10.2. A function $f \in L^2(X|Y)$ is *conditionally almost periodic* if its orbit $\{T^n f : n \in \mathbb{Z}\}$ is conditionally precompact in $L^2(X|Y)$, and it is *conditionally almost periodic in measure* if for every $\epsilon > 0$ there exists a set E in Y with $\mu(E) > 1 - \epsilon$ such that $f 1_E$ is conditionally almost periodic.

Definition 10.3. An extension $X \to Y$ of measure preserving systems is said to be *compact* if every function in $L^2(X|Y)$ is conditionally almost periodic in measure.

Example 10.4. Consider the skew torus from Example 8.3, given by $(\mathbb{R}/\mathbb{Z})^2$ with the shift map $T(y,z) = (y + a, z + y)$. So $T^n(y,z) = (y + na, z + ny + \binom{n}{2} a)$. Consider the skew torus X as an extension of $Y = (\mathbb{R}/\mathbb{Z}, \mathcal{B}, \mu, y \mapsto y + a)$. Let $f(y,z) = e^{2\pi i z}$. So

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

which lies in the zonotope $\left\{ cf : c \in L^\infty(Y), \|c\|_{L^\infty(Y)} \leq 1 \right\}$ and hence f is conditionally almost periodic, whereas it is not almost periodic in the unconditional sense. In fact, $X \to Y$ is a compact extension. This is an analogue of the circle rotation example from Example 6.3.

10.2 Multiple recurrence

In this section we prove that the SZ property lifts through compact extensions. While the proof that every compact system is SZ (Proposition 6.6) was very straight-forward, the proof for the conditional case is more technically involved. We provide two proofs. The first proof assumes van der Waerden's theorem. The second proof, as given in [FKO82], does not use van der Waerden's theorem.

Theorem 10.5. *Let $X \to Y$ be a compact extension. If Y is SZ, then so is X.*

First proof. (Assuming van der Waerden's theorem) Fix $k \geq 1$, and let $f \in L^\infty(X)$ be a nonnegative function with positive mean. We wish to show that

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f \, T^n f \cdots T^{(k-1)n} f \, d\mu > 0. \tag{10.1}$$

Since $X \to Y$ is a compact extension, we know that f is conditionally almost periodic in measure. As we can replace f by a lower bound, we can just assume that f is conditionally almost periodic. Furthermore, we can normalise so that $\|f\|_{L^\infty(X)} = 1$.

Let $\epsilon > 0$ be a small number, and K a large integer, both to be decided later. Choose $\delta > 0$ so that the subset $A = \{y \in Y : \mathbf{E}(f|Y)(y) > \delta\}$ of Y satisfies $\nu(A) > 0$. Then since the system Y is SZ, we know that

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

In other words, there is some constant $c > 0$ such that

$$\nu(A \cap T^n A \cap \cdots \cap T^{(K-1)n} A) > c \tag{10.2}$$

for a set of n of positive lower density.

For now, fix n satisfying the above property. We find that $\mathbf{E}(T^{an}f|Y)(y) \geq \delta$ for all $y \in A \cap T^n A \cap \cdots \cap T^{(K-1)n}A$ and $0 \leq a < K$.

Since f is conditionally almost periodic, we can find $f_1, \ldots, f_d \in L^2(X|Y)$ so that for $0 \leq a < K$, we have $\min_{1 \leq i \leq d} \|T^{an}f - f_i\|_{L^2(X|Y)} < \epsilon$ almost everywhere in Y. Then, for almost every y, we have some map $i_y : \{0, 1, \ldots, K-1\} \to \{1, \ldots, d\}$ so that $\left\|T^{an}f - f_{i_y(a)}\right\|_{L^2(X|Y)}(y) < \epsilon$. Viewing i_y as a colouring, by van der Waerden's theorem, if we choose K large enough, we have some monochromatic k-term arithmetic progression $a_y, a_y + r_y, \ldots, a_y + (k-1)r_y$ in $\{0, 1, \ldots, K-1\}$, with respect to the colouring i_y. The quantities a_y and r_y can be chosen as measurable functions in y. Since there are only finitely many possibilities for a_y and r_y, by pigeonhole principle we can find some subset B_n of $A \cap T^n A \cap \cdots \cap T^{(K-1)n}A$ of positive measure so that a_y and r_y are constant on $y \in B_n$ (they could still depend on n, but that does not affect the argument), and furthermore there is some $\sigma > 0$ independent of n such that $\nu(B_n) > \sigma$ as long as (10.2) holds. This means that there is some f_i so that

$$\left\|T^{(a+jr)n}f - f_i\right\|_{L^2(X|Y)}(y) < \epsilon, \qquad \forall y \in B_n,\ j = 0, 1, \ldots, k-1.$$

So by triangle inequality,

$$\left\|T^{(a+jr)n}f - T^{an}f\right\|_{L^2(X|Y)}(y) < 2\epsilon, \qquad \forall y \in B_n,\ j = 0, 1, \ldots, k-1.$$

Using $\|f\|_{L^\infty(X)} = 1$, we have

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

It follows that

$$\mathbf{E}\left(T^{an}f\ T^{(a+r)n}f \cdots T^{(a+(k-1)r)n}f|Y\right)(y) \geq \mathbf{E}\left((T^{an}f)^k\middle| Y\right)(y) - O_k(\epsilon).$$

We have $\mathbf{E}(T^{an}f|Y)(y) \geq \delta$ for $y \in B_n$. So we can find some $c(\delta, k) > 0$ so that

$$\mathbf{E}\left((T^{an}f)^k\middle| Y\right)(y) \geq c(k, \delta) > 0, \qquad \forall y \in B_n.$$

Then

$$\begin{aligned}
\mathbf{E}\left(T^{an}f\ T^{(a+r)n}f \cdots T^{(a+(k-1)r)n}f\middle| Y\right)(y) &\geq \mathbf{E}\left((T^{an}f)^k\middle| Y\right)(y) - O_k(\epsilon) \\
&\geq c(k, \delta) - O_k(\epsilon) \\
&\geq \frac{1}{2}c(k, \delta), \qquad \forall y \in B_n
\end{aligned}$$

by choosing ϵ small enough. Integrating over y, we find that

$$\int_X T^{an}f\ T^{(a+r)n}f \cdots T^{(a+(k-1)r)n}f\ d\mu \geq \frac{1}{2}c(k, \delta)\nu(B_n) > \frac{1}{2}c(k, \delta)\sigma.$$

Since μ is T-invariant, we have

$$\int_X f\ T^{rn}f \cdots T^{(k-1)rn}f\ d\mu \geq \frac{1}{2}c(k, \delta)\sigma.$$

Recall that although r depends on n, it ranges between 1 and $K-1$. The above inequality is true for a set of n of positive lower density, and hence

$$\int_X f\ T^n f \cdots T^{(k-1)n}f\ d\mu \geq \frac{1}{2}c(k, \delta)\sigma.$$

is true for a set of n of positive lower density. Then (10.1) follows, showing that X is SZ. $\square$

Second proof (sketch). (Without assuming van der Waerden's theorem) Fix $k \geq 1$, and let $f \in L^2(X)$ be a nonnegative function with positive mean. The idea once again is to show that for some subset of n of positive lower density, the functions $T^n f$, $T^{2n} f$, $\dots$, $T^{(k-1)n} f$, are all "close" to f in some sense. In the previous proof where we assumed van der Waerden's theorem, we obtained proximity between the various functions $T^{jn} f$ by considering some ϵ-net of the shifts $T^n f$ using the almost periodicity condition, effectively partitioning the functions $\{T^n f : n \in \mathbb{N}\}$ into nearby clusters so that we can use van der Waerden's theorem to extract an arithmetic progression from one of the clusters. However, without van der Waerden's theorem, it will be difficult to extract an arithmetic progression via a partitioning of $\{T^n f : n \in \mathbb{Z}\}$ induced through the almost periodicity condition. Instead, we consider vectors of the form $(f, T^n f, T^{2n} f, \dots, T^{(k-1)n} f)$ for $n \in \mathbb{Z}$, so that the arithmetic progression arises naturally from the setup.

As in the previous proof, we begin by considering some positive measure subset A of Y where the conditional expectation of f is bounded below by some $\delta > 0$.

We consider k-tuples $\bigoplus_{i=0}^{k-1} L^2(X)$, and the conditional norm on it given by $\|(f_0, \dots, f_{k-1})\|_{L^2(X|Y)} = \max_{0 \leq i < k} \|f_i\|_{L^2(X|Y)}$. Then, since f is conditionally almost periodic, the set

$$L = \left\{ (f, T^n f, \dots, T^{(k-1)n} f) : n \in \mathbb{Z} \right\}$$

is uniformly totally bounded over almost all $y \in Y$ with respect to the norm just defined. Let $\epsilon > 0$. For each finite set F of integers, let A_F denote the subset of A above which $L_F = \left\{ (f, T^m f, \dots, T^{(k-1)m} f) : m \in F \right\}$ has all the components in all the vectors at least δ, and such that L_F forms a maximal ϵ-separated set in L over every $y \in A_F$. Then the A_F's cover A, since maximal ϵ-separated sets exist above almost every $y \in Y$ due to totally boundedness. Then some A_F has positive measure. Replace A by A_F. By reducing A furthermore if necessary, we can make it so that for distinct $m, m' \in F$, and each $0 \leq i < k$, the conditional distance between $T^{im} f$ and $T^{im'} f$ do not vary too much over A.

So $L_F = \left\{ (f, T^m f, \dots, T^{(k-1)m} f) : m \in F \right\}$ if maximally ϵ-separated over A. Whenever $A_n = A \cap T^n A \cap T^{2n} A \cap \dots \cap T^{(k-1)n} A$ satisfies $\mu(A_n) > c$, by shift invariance on each coordinate, we see that $L_{F+n} = \left\{ (f, T^{n+m} f, \dots, T^{k(n+m)} f) : m \in F \right\}$ is also maximally ϵ-separated, but over the A_n. Since Y is SZ, such n form a set of positive lower density when $c > 0$ is sufficiently small. We know that L_{F+n} is maximally ϵ-separated and we made sure that the distances between different vectors do not vary too much over A_n, and since $(f, f, \dots, f) \in L$, we see that there exists some $m_n \in F$ such that $(f, f, \dots, f)$ is ϵ-close to $(f, T^{n+m_n} f, \dots, T^{(k-1)(n+m_n)} f)$ over all of A_n. Then

$$\begin{aligned} \int_X f\ T^{n+n_m} f\ T^{2(n+n_m)} f\ \cdots\ T^{(k-1)(n+n_m)} f\ d\mu &\geq \int_{A_n} f\ T^{n+n_m} f\ T^{2(n+n_m)} f\ \cdots\ T^{(k-1)(n+n_m)} f\ d\mu \\ &\approx_\epsilon \int_{A_n} f \cdot f \cdot f \cdot \dots \cdot f\ d\mu \\ &\geq c\delta^k. \end{aligned}$$

All the details can be made rigourous. Since $n_m \in F$ is bounded, this gives the desired lower bound, showing that

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

thereby showing that X is SZ. □

10.3 Conditionally Hilbert-Schmidt operators

In Section 6.4 we defined Hilbert-Schmidt operators and showed to how use it to produce almost periodic functions. Now we consider a conditional version of Hilbert-Schmidt operators.

Definition 10.6 (Sub-orthonormal sets)**.** Let $X \to Y$ be an extension of systems. A *sub-orthonormal set* in $L^2(X|Y)$ is any at most countable sequence $(e_a)_{a\in A}$ in $L^2(X|Y)$ such that $\langle e_a, e_b\rangle_{X|Y} = 0$ a.e. for all $a \neq b \in A$ and $\langle e_a, e_a\rangle_{X|Y} \leq 1$ a.e. for all $a \in A$.

Definition 10.7. Let $X \to Y$ be an extension of systems. A $L^\infty(Y)$-linear module homomorphism $\Phi : L^2(X|Y) \to L^2(X|Y)$ is said to be a *conditionally Hilbert-Schmidt operator* if there is some constant $C > 0$ such that

$$\sum_{a\in A}\sum_{b\in B}\left|\langle \Phi e_a, f_b\rangle_{X|Y}\right|^2 \leq C^2 \text{ a.e.}$$

whenever $\{e_a\}_{a\in A}$ and $\{f_b\}_{b\in B}$ are sub-orthonormal sets. The minimum value of C is called the *(uniform) conditional Hilbert-Schmidt norm* of Φ, denoted $\left\|\|\Phi\|_{HS(X|Y)}\right\|_{L^\infty(Y)}$.

In the absolute case, we saw that Hilbert-Schmidt operators are compact, meaning that the image of a unit ball is precompact. This is also true for the conditional case.

Proposition 10.8 (Conditional Hilbert-Schmidt operators are compact)**.** *Let $X \to Y$ be an extension. Let $\Phi : L^2(X|Y) \to L^2(X|Y)$ be a conditionally Hilbert-Schmidt operator. Then the image of the the unit ball in $L^2(X|Y)$ is conditionally precompact, where the unit ball in $L^2(X|Y)$ is given by the $\left\|\|\cdot\|_{L^2(X|Y)}\right\|_{L^\infty(Y)}$ norm.*

The proposition can be proved by relativising the proof of Proposition 6.12. We omit the details, which may be found in [Tao09].

Proposition 10.9. *Let $X \to Y$ be an extension. Let $\Phi : L^2(X|Y) \to L^2(X|Y)$ be a conditional Hilbert-Schmidt operator that commutes with T. Then Φf is a conditionally almost periodic function for any $f \in L^2(X|Y)$.*

Proof. The set $\{T^n f : n \in \mathbb{Z}\}$ is bounded in the $\left\|\|\cdot\|_{L^2(X|Y)}\right\|_{L^\infty(Y)}$ norm, so its image under Φ, $\Phi\{T^n f : n \in \mathbb{Z}\} = \{\Phi T^n f : n \in \mathbb{Z}\} = \{T^n(\Phi f) : n \in \mathbb{Z}\}$ is conditionally precompact by Proposition 10.8. This shows that Φf is conditionally almost periodic. □

10.4 Weak mixing and almost periodic components

Let $X \to Y$ be an extension of systems. Let $WM(X|Y)$ be the set of conditionally weak mixing functions in $L^2(X|Y)$ and $AP(X|Y)$ the set of conditionally almost periodic in measure functions in $L^2(X|Y)$. In this subsection we prove an orthogonal decomposition of $L^2(X|Y)$, relativising Section 6.5.

Proposition 10.10. *Let $X \to Y$ be an extension of systems. Then as $L^\infty(X)$-modules we have*

$$L^2(X|Y) = WM(X|Y) \oplus AP(X|Y)$$

such that whenever $f \in WM(X|Y)$ and $g \in AP(X|Y)$ we have $\langle f, g\rangle_{X|Y} = 0$ a.e.

Lemma 10.11. *Let $X \to Y$ be an extension of systems. If $f \in WM(X|Y)$ and $g \in AP(X|Y)$, then $\langle f, g\rangle_{X|Y} = 0$ a.e.*

Proof. We have

$$\left\|\langle T^n f, T^n g\rangle_{X|Y}\right\|_{L^2(Y)} = \left\|T^{-n}\langle T^n f, T^n g\rangle_{X|Y}\right\|_{L^2(Y)} = \left\|\langle f, g\rangle_{X|Y}\right\|_{L^2(Y)}.$$

Therefore, to show that $\langle f, g\rangle_{X|Y} = 0$ a.e, it suffices to show that

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

Let $\epsilon > 0$ be arbitrary. Since g is conditionally almost periodic in measure, we can find $g_1, \ldots, g_d \in L^2(X|Y)$ such that $\min_{1 \le i \le d} \|T^n g - g_i\|_{L^2(X|Y)} < \epsilon$ except possibly on a set of measure ϵ. From this we deduce that

$$\left\| \langle T^n f, T^n g \rangle_{X|Y} \right\|_{L^2(Y)} \le \sum_{i=1}^{d} \| \langle T^n f, g_i \rangle \|_{L^2(Y)} + O(\epsilon).$$

By Proposition 9.8, the Cesàro limit of the above expression is at most $O(\epsilon)$. Since ϵ was arbitrary, the result follows. $\square$

We have the following relative version of Lemma 6.16, whose proof we omit.

Lemma 10.12. *Let $(X, \mathcal{X}, \mu, T)$ be an extension of $(Y, \mathcal{Y}, \nu, S)$. Then $AP(X|Y)$ is a T-invariant $L^\infty(Y)$-module (i.e., it is closed under addition, scalar multiplication by $L^\infty(Y)$, and T) which is also closed under pointwise operations $f, g \mapsto \max(f, g)$ and $f, g \mapsto \min(f, g)$.*

Lemma 10.13. *Let $f \in L^2(X|Y)$. Then $f \in WM(X|Y)$ if and only if $\langle f, g \rangle_{X|Y} = 0$ a.e. for every $g \in AP(X|Y)$.*

Proof. If $f \in WM(X|Y)$, then the result follows from Lemma 10.11. So assume $f \notin WM(X|Y)$. It suffices to show that in this case there exists some $g \in AP(X|Y)$ such that $\langle f, g \rangle_X \neq 0$.

From the calculations in (9.3) we have

$$\frac{1}{N} \sum_{n=0}^{N-1} \left\| \langle T^n f, g \rangle_{X|Y} \right\|_{L^2(Y)}^2 = \left\langle \frac{1}{N} \sum_{n=0}^{N-1} \langle T^n f, g \rangle_{X|Y} T^n f, g \right\rangle_X \tag{10.3}$$

for all $f, g \in L^2(X|Y)$. Define operators $\Phi_f, \Phi_{f,N}, \Psi_f \colon L^2(X|Y) \to L^2(X|Y)$ as follows. Let

$$\Phi_f g = \langle f, g \rangle_{X|Y} f.$$

Note that $\Phi_{Tf} = T \Phi_f T^{-1}$. Let their Cesàro sums be

$$\Phi_{f,N} = \frac{1}{N} \sum_{n=0}^{N-1} \Phi_{T^n f}.$$

Since f is not conditionally weak mixing, there is some $\delta > 0$ and some sequence $N_i \to \infty$ such that $\frac{1}{N_i} \sum_{n=0}^{N_i - 1} \left\| \langle T^n f, f \rangle_{X|Y} \right\|_{L^2(Y)}^2 \ge \delta$ for all i. Let Ψ_f be any limit point of Φ_{f,N_i} in the weak operator topology (by sequential compactness as we show momentarily that $\Phi_{f,N}$ has uniformly bounded norm). Then setting $g = f$ in (10.3) yields $\langle \Phi_{f,N} f, f \rangle_X \ge \delta$, so $\langle \Psi_f f, f \rangle_X \ge \delta$ by weak convergence. It remains to show that $\Psi_f f \in AP(X|Y)$.

For any two sub-orthonormal basis $(g_a)_{a \in A}$ and $(g_n)_{b \in B}$ of $L^2(X|Y)$,

$$\begin{aligned}
\sum_{a \in A} \sum_{b \in B} \left| \langle \Phi_f g_a, h_b \rangle_{X|Y} \right|^2 &= \sum_{a \in A} \sum_{b \in B} \left| \langle f, g_a \rangle_{X|Y} \langle f, h_b \rangle_{X|Y} \right|^2 \\
&= \left(\sum_{a \in A} \left| \langle f, g_a \rangle_{X|Y} \right|^2 \right) \left(\sum_{b \in B} \left| \langle f, h_b \rangle_{X|Y} \right|^2 \right) \\
&\le \|f\|_{L^2(X|Y)}^2 \text{ a.e.}
\end{aligned}$$

Thus Φ_f is a conditionally Hilbert-Schmidt operator with conditional Hilbert-Schmidt norm at most $\left\| \|f\|_{L^2(X|Y)}^2 \right\|_{L^\infty Y}$, which we may assume to be at most 1 through normalisation. By the triangle inequality, each $\Phi_{f,N}$ also has conditional Hilbert-Schmidt norm at most 1, thus the conditional Hilbert-Schmidt norm of Ψ_f is also at most 1. In particular, this shows that Ψ_f is a conditionally Hilbert-Schmidt operator.

We claim that Ψ_f commutes with T. Indeed, we have the telescoping identity

$$T\Phi_{f,N}T^{-1} - \Phi_{f,N} = \Phi_{Tf,N} - \Phi_{f,N} = \frac{1}{N}\left(\Phi_{T^N f} - \Phi_f\right)$$

which converges to zero even in the conditional Hilbert-Schmidt norm, and hence also in the strong and weak operator topologies. Taking a weak limit of $\Phi_{f,N}$ gives $T\Psi_f T^{-1} = \Psi_f$, and hence Ψ_f commutes with T. We conclude by Proposition 10.9 that $\Psi_f f$ is conditionally almost periodic, as claimed. □

To finish the proof of the decomposition $L^2(X|Y) = WM(X|Y) \oplus AP(X|Y)$ in Proposition 10.10, it remains to show that every function $f \in L^2(X|Y)$ can be decomposed as a sum of an element of $WM(X|Y)$ and an element of $AP(X|Y)$. We defer this claim to Section 10.6. For now, we state the following corollary.

Corollary 10.14. *An extension $X \to Y$ is weak mixing if and only if $AP(X|Y) = L^\infty(Y)$.*

10.5 Existence of conditionally almost periodic functions

The purpose of this section is similar to that of Section 6.6. We sketch an alternate proof, using the fibre product characterisation of weak mixing extensions, that if $X \to Y$ is not weak mixing, then there exists some conditionally almost periodic function that is not in $L^\infty(Y)$ (which play the role of constants in the Hilbert module).

Proposition 10.15. *If $\pi\colon (X, \mathcal{X}, \mu, T) \to (Y, \mathcal{Y}, \nu, S)$ is an extension that is not weak mixing, then there exists some $f \in AP(X|Y)$ which is not in $L^\infty(Y)$.*

Proof. (Sketch) By Proposition 9.12, $X \times_Y X \to Y$ is not an ergodic extension. So there exists some T-invariant $K \in L^2(X \times_Y X|Y)$ which does not lie in $L^\infty(Y)$. It can be shown that the map $\Phi_K\colon L^2(X|Y) \to L^2(X|Y)$ given by

$$\Phi_K f(x) = \int_X K(x, x') f(x') d\mu_{\pi(x)}(x')$$

is a conditionally Hilbert-Schmidt operator. There exists some $f \in L^2(X|Y)$ so that $\Phi_K f$ does not lie in $L^\infty(Y)$. Since Φ_K commutes with T, we have $\Phi_K f \in AP(X|Y) \setminus L^\infty(Y)$ by Proposition 10.9. □

10.6 Existence of compact extensions

In this subsection we relativise the dichotomy of structure and randomness from Section 6.7.

Theorem 10.16. *Let $X \to Y$ be an extension. Exactly one of the following is true*

(a) $X \to Y$ is a weak mixing extension;

(b) There exists a tower of extensions $X \to X^* \to Y$ such that $X^* \to Y$ is a non-trivial compact extension.

Let

$$\mathcal{X}_{AP(X|Y)} = \{A \in \mathcal{X} : 1_A \in AP(X|Y)\}.$$

This factor is the relative version of the Kronecker factor. It characterises the set of almost periodic in measure functions.

Proposition 10.17. *Let $(X, \mathcal{X}, \mu, T) \to (Y, \mathcal{Y}, \nu, S)$ be an extension of systems. Let $f \in L^2(X|Y)$.*

(a) $f \in AP(X|Y)$ if and only if f is $\mathcal{X}_{AP(X|Y)}$-measurable, i.e., $AP(X|Y) = L^2\left(X, \mathcal{X}_{AP(X|Y)}, \mu \middle| Y, \mathcal{Y}, \nu\right)$.

(b) $f \in WM(X|Y)$ if and only if $\mathbf{E}\left(f \middle| \mathcal{X}_{AP(X|Y)}\right) = 0$ a.e.

(c) We can write $f = f_{AP} + f_{WM}$, where $f_{AP} = \mathbf{E}\left(f \middle| \mathcal{X}_{AP(X|Y)}\right) \in AP(X|Y)$ and $f_{WM} = f - f_{AP} \in WM(X|Y)$.

We omit the proof, which can be formed by relativising Proposition 6.21. The decomposition $L^2(X|Y) = WM(X|Y) \oplus AP(X|Y)$ in Proposition 10.10 also follows.

Corollary 10.18. *Let $X \to Y$ be an extension of systems. Then $\mathcal{X}_{AP(X|Y)}$ is the largest factor of X such that $\mathcal{X}_{AP(X|Y)}$ is a compact extension of Y. Furthermore, this compact extension is nontrivial if and only if $AP(X|Y) \neq L^\infty(Y)$.*

Theorem 10.16 then follows from Corollaries 10.18 and 10.14 (or Proposition 10.15).

11 Tower of extensions

We have already shown that the SZ property lifts through weak mixing and compact extensions. We can build a tower of factors of X: $K_1 \leftarrow K_2 \leftarrow K_3 \leftarrow \cdots$, where K_1 is the Kronecker factor, and at each step we take K_j to be the factor of X induced by $\mathcal{X}_{AP(X|K_{j-1})}$, and so each $K_{j+1} \to K_j$ is a compact extension. This shows that the SZ property can be lifted to each K_j. If at some point we reach a factor K_j such that $X \to K_j$ is a weak mixing extension, then no more compact extensions are possible, but we would also be done, since SZ lifts through weak extensions as well. However, we might never reach this stopping point. If this is the case, we can construct the factor K_ω generated by the union of the σ-algebras of all $\{K_j : j \in \mathbb{N}\}$, and then keep growing the tower $K_\omega \leftarrow K_{\omega+1} \leftarrow \cdots$ enumerated by ordinals.

In this section we study this tower of extensions and show that the SZ property can indeed be lifted all the way to the top.

11.1 Furstenberg-Zimmer structure theorem

We give a structural result of Furstenberg [Fur77] and Zimmer [Zim76] that says that every measure preserving system can be written as a tower of compact and weak mixing extensions.

Theorem 11.1 (Furstenberg-Zimmer)**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system. Then there exists an ordinal α and a factor $\pi_\beta\colon : X \to Y_\beta = (Y_\beta, \mathcal{Y}_\beta, \nu_\beta, S_\beta)$ for every $\beta \leq \alpha$ with the following properties:*

(1) Y_0 is trivial (e.g., a one point system).

(2) For every successor ordinal $\beta + 1 \leq \alpha$, $Y_{\beta+1} \to Y_\beta$ is a compact extension.

(3) For every limit ordinal $\beta \leq \alpha$, Y_β is a limit of Y_γ for $\gamma < \beta$, in the sense that $\pi_\beta^{-1}\mathcal{Y}_\beta$ is generated by $\bigcup_{\gamma<\beta} \pi_\gamma^{-1}\mathcal{Y}_\gamma$.

(4) $X \to Y_\alpha$ is a weak mixing extension.

Proof. Let Σ denote the family of systems of factors of X such that for each $\sigma \in \Sigma$ there is an ordinal α^σ and a factor $\pi_\beta^\sigma\colon : X \to Y_\beta^\sigma = (Y_\beta^\sigma, \mathcal{Y}_\beta^\sigma, \nu_\beta^\sigma, S_\beta^\sigma)$ for each $\beta \leq \alpha^\sigma$ such that

(1') Y_0^σ is trivial.

(2') For every successor ordinal $\beta^\sigma + 1 \leq \alpha^\sigma$, $Y_{\beta+1}^\sigma \to Y_\beta^\sigma$ is a *non-trivial* compact extension.

(3') For every limit ordinal $\beta \leq \alpha^\sigma$, Y_β^σ is a limit of Y_γ^σ for $\gamma < \beta$.

Consider the partial order on Σ given by: $\sigma_1 \leq \sigma_2$ if $\alpha^{\sigma_1} \leq \alpha^{\sigma_2}$ and $Y_\beta^{\sigma_1} = Y_\beta^{\sigma_2}$ whenever $\beta \leq \alpha^{\sigma_1}$ (i.e., the tower σ_2 is an extension of the tower σ_1). We use Zorn's lemma to prove that Σ has a maximal element. Consider a totally ordered subset $T \subset \Sigma$. Let α denote the minimum ordinal greater than or equal to all of $\{\alpha^\sigma : \sigma \in T\}$. If $\alpha = \alpha^\sigma$ for some $\sigma \in T$, then σ is an upper bound to T. Otherwise, α is a limit ordinal. Construct $\sigma' \in \Sigma$ by $\alpha^{\sigma'} = \alpha$, $Y_\beta^{\sigma'} = Y_\beta^\sigma$ for $\beta < \alpha^\sigma < \alpha$, and set $Y_\alpha^{\sigma'}$ to be the limit of $Y_\beta^{\sigma'}$ for $\beta < \alpha$. Then σ' is an upper bound to T. Therefore, every totally order set has an upper bound, and by Zorn's lemma we deduce that Σ has a maximal element σ.

We claim that the maximal element $\sigma \in \Sigma$ gives the desired tower of extensions $\{Y_\beta : \beta \leq \alpha\}$. The conditions (1), (2), (3) follow from (1'), (2'), (3') respectively. If X is not a weak mixing extension of Y_α, then by Theorem 10.16 we can extend the tower one step further with a nontrivial compact extension of Y_α, thereby contradicting the maximality of σ. So (4) is satisfied as well. $\square$

11.2 Limit of SZ systems

In the Furstenberg-Zimmer tower, in addition to compact and weak mixing extensions, there are also limits. In this subsection we show that SZ lifts through limits.

Proposition 11.2 (SZ lifts through limits)**.** *Let $(Y_\beta)_{\beta\in B}$ be a totally ordered chain of factors of a measure preserving system X, and suppose that X is the limit of $(Y_\beta)_{\beta\in B}$. If each Y_β is SZ, then X is SZ as well.*

Lemma 11.3. *Let $(X, \mathcal{X}, \mu, T)$ be an extension of $(Y, \mathcal{Y}, \nu, S)$. Suppose that Y is SZ. Let k be a nonnegative integer. Let $E \in \mathcal{X}$ and $F = \left\{y \in Y : \mathbf{E}\left(1_E \middle| Y\right) > 1 - \frac{1}{2k}\right\}$. If $\nu(F) > 0$, then*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n-0}^{N-1} \mu\left(E \cap T^n E \cap \cdots \cap T^{(k-1)n} E\right) > 0. \tag{11.1}$$

Proof. Since Y is SZ and $\nu(F) > 0$, there is some $c > 0$ such that $\nu(F \cap T^n F \cap \cdots \cap T^{(k-1)n} F) > c$ for all n in a set of positive lower density. Observe that

$$\mathbf{E}\left(1_{E\cap T^n E\cap\cdots T^{(k-1)n}E}\middle| Y\right)(y) \geq 1 - \frac{k}{2k} = \frac{1}{2} \qquad \text{for } y \in F \cap T^n F \cap \cdots \cap T^{(k-1)n} F.$$

It follows that $\mu\left(E \cap T^n E \cap \cdots T^{(k-1)n} E\right) \geq \frac{c}{2}$ for all n in a set of positive lower density, and the lemma follows. $\square$

Proof of Proposition 11.2. Let $E \in \mathcal{X}$ with $\mu(E) > 0$. Since X is a limit of $(Y_\beta)_{\beta\in B}$, for any $\epsilon > 0$, there is some Y_β such that $\left\|1_E - \mathbf{E}\left(1_E\middle| Y_\beta\right)\right\|_{L^2(X)} \leq \epsilon$. We claim that if $\epsilon > 0$ if sufficiently small, then

$$\nu_\beta\left(\left\{y \in Y_\beta : \mathbf{E}\left(1_E\middle| Y_\beta\right) > 1 - \frac{1}{2k}\right\}\right) > 0. \tag{11.2}$$

Otherwise, we have $\mathbf{E}\left(1_E\middle| Y_\beta\right) \leq 1 - \frac{1}{2k}$ a.e., so that $\left|1_E - \mathbf{E}\left(1_E\middle| Y_\beta\right)\right|^2 \geq \frac{1}{4k^2}$ in E, and hence $\left\|1_E - \mathbf{E}\left(1_E\middle| Y_\beta\right)\right\|_{L^2(X)} \geq \frac{\sqrt{\mu(E)}}{2k}$. Thus choosing $\epsilon < \frac{\sqrt{\mu(E)}}{2k}$ guarantees (11.2). Since Y_β is SZ, we may apply Lemma 11.3 to obtain (11.1). It follows that X is SZ. $\square$

11.3 Conclusion of proof

From Theorem 11.1 we know that there exists a tower of extensions

$$X \to Y_\alpha \to \cdots \to Y_3 \to Y_2 \to Y_1 \to Y_0$$

indexed by ordinals, where Y_0 is the trivial system and each step is either a weak extension, a compact extension, or a limit. Also we have proven the following results about lifting the SZ property via extensions.

- If $X \to Y$ is a weak mixing extension, and Y is SZ, then X is SZ. (Theorem 9.13)
- If $X \to Y$ is a compact extension, and Y is SZ, then X is SZ. (Theorem 10.5)
- If X is the limit of a chain of factor (Y_β), and each Y_β is SZ, then X is SZ. (Proposition 11.2).

Putting everything together using transfinite induction, we see that the SZ property can be lifted all the way to the top of the tower, so X is SZ. This concludes the proof of Theorem 4.2, and hence that of Szemerédi's theorem.

12 Recent advances

In this final section, we highlight some recent advances on certain ergodic elements that arise in the proof of Szemerédi's theorem.

12.1 Nonconventional ergodic averages

Furstenberg's proof of Szemerédi's theorem established

$$\liminf_{N\to\infty} \frac{1}{N} \int_X f\, T^n f\, T^{2n} f \cdots T^{(k-1)n} f \, d\mu > 0.$$

However it leaves open the question of whether the limit actually exists. Host and Kra [HK05b] recently answered this question in the affirmative.

Theorem 12.1 (Host-Kra)**.** *Let X be a measure preserving system. Let $f_1, \dots, f_k \in L^\infty(X)$. Then*

$$\operatorname*{C-lim}_{n\to\infty} T^n f_1\, T^{2n} f_2 \cdots T^{kn} f_k \tag{12.1}$$

exists in $L^2(X)$.

Such averages are called *non-conventional ergodic averages*, as they differ from the $k = 1$ case when the result follows from the mean ergodic theorem (Theorem 2.13). When $k = 2$ and X is ergodic, the convergence was established by Furstenberg [Fur77]. We discussed this case while proving Roth's theorem in Section 7.

When X is weak mixing, we saw in Proposition 5.16 that the Cesàro limit equals to the constant $\mathbf{E}(f_1) \cdots \mathbf{E}(f_k)$. For general systems this limit might not be constant.

Example 12.2. Consider the circle rotation system on $\mathbb{R}/\mathbb{Z}$ with shift map $Tx = x + a$. Let $f_1(x) = e^{4\pi i x}$ and $f_2 = e^{-2\pi i x}$. Then $T^n f_1\, T^{2n} f_2 = \overline{f_2}$. So the limit in Theorem 12.1 is $\overline{f_2}$, which is nonconstant.

When X is an ergodic compact system, and hence equivalent to some Kronecker system with $Tx \mapsto x + a$, the limit in Theorem 12.1 is

$$x \mapsto \operatorname*{C-lim}_{n\to\infty} f_1(x - na) f_2(x - 2na) \cdots f_k(x - kna) = \int_X f_1(x - y) f_2(x - 2y) \cdots f_k(x - ky)\, d\mu(y)$$

since $(na)_{n\in\mathbb{N}}$ is equidistributed in X.

Theorem 12.1 has been generalised to polynomial averages.

Theorem 12.3 (Host-Kra [HK05a], Leibman [Lei05a]). *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system. Let $f_1, \ldots, f_k \in L^\infty(X)$. Then for any polynomials $p_1, \ldots, p_k : \mathbb{Z} \to \mathbb{Z}$, the limit*

$$\operatorname*{C-lim}_{n\to\infty} T^{p_1(n)} f_1 \; T^{p_2(n)} f_2 \cdots T^{p_k(n)} f_k$$

exists in $L^2(X)$.

There is also a version with multiple commuting transformations, but with stronger hypotheses of ergodicity. The general case remains open.

Theorem 12.4 (Host-Kra [HK05a], Leibman [Lei05a]). *Let $T_1, \ldots, T_k$ be commuting invertible ergodic measure preserving transformations of a probability space $(X, \mathcal{X}, \mu)$ such that $T_i \circ T_j^{-1}$ is ergodic whenever $i \neq j$. If $f_1, \ldots, f_k \in L^\infty(X)$, then*

$$\operatorname*{C-lim}_{n\to\infty} T_1^n f_1 \; T_2^n f_2 \cdots T_k^n f_k$$

exists in $L^2(X)$.

In the remainder of this section, we describe the ideas used in analysing these nonconventional ergodic averages. For more on the subject, see the expository articles [Hos, Kra06, Kra07] or the original papers.

12.2 Characteristic factors

A key ingredient in Furstenberg's proof is the idea of using factors to characterise certain behaviour, such as almost periodicity. It turns out that to understand nonconventional ergodic averages, we also need to consider factors that characterise these averages. We have actually already seen examples of this technique. From the mean ergodic theorem (Theorem 2.13) we saw that

$$\operatorname*{C-lim}_{n\to\infty} T^n f = \mathbf{E}\left(f \middle| \mathcal{X}^T\right)$$

in $L^2(X)$, where $\mathcal{X}^T$ is the sub-σ-algebra of all T-invariant sets. So the factor $\mathcal{X}^T$ characterises the average of $T^n f$. In Proposition 7.3, we saw that in ergodic systems,

$$\operatorname*{C-lim}_{n\to\infty} \left(T^n f_1 \; T^{2n} f_2 - T^n \mathbf{E}\left(f_1 \middle| \mathcal{X}_{AP}\right) \; T^{2n} \mathbf{E}\left(f_1 \middle| \mathcal{X}_{AP}\right)\right) = 0$$

in $L^2(X)$. So the Kronecker factor $\mathcal{X}_{AP}$ characterises the average of $T^n f_1 \; T^{2n} f_2$, in the sense that to evaluate this average, it suffices to work in $\mathcal{X}_{AP}$.

In general, the *characteristic factor* for $T^n f_1 \; T^{2n} f_2 \cdots \; T^{kn} f_k$ is a factor Y of X such that

$$\operatorname*{C-lim}_{n\to\infty} \Big(T^n f_1 \; T^{2n} f_2 \cdots \; T^{kn} f_k - T^n \mathbf{E}\left(f_1 \middle| Y\right) \; T^{2n} \mathbf{E}\left(f_2 \middle| Y\right) \cdots \; T^{kn} \mathbf{E}\left(f_k \middle| Y\right)\Big) = 0.$$

The characteristic factor allows us to reduce the problem to the factor Y. The Kronecker factor is an example of a characteristic factor for $k = 2$ when X is ergodic, and it has the advantage of having an algebraic structure, so that it can be understood easily. Unfortunately, taking repeated extensions of the Kronecker factor as we did in the beginning of Section 11 does not suffice in giving characteristic factors for higher k. We would like to have characteristic factors with useful algebraic/geometric structure in order to prove the convergence of the limit. This was the approach taken by Host and Kra [HK05b]. They identified a tower $Z_0 \leftarrow Z_1 \leftarrow Z_2 \leftarrow \cdots$ of factors of X such that Z_{k-1} is a characteristic factor for $T^n f_1 \; T^{2n} f_2 \cdots \; T^{kn} f_k$. Furthermore each factor Z_k has a nice description in terms of nilpotent Lie groups, much like how the Kronecker factor has an abelian group structure. We will describe these factors in the next subsection.

12.3 Nilsystems

Previously we saw that the limiting behaviour of the running average of $T^n f_1\, T^{2n} f_2$ is controlled by the Kronecker factor, which can be modelled by an abelian group. Host and Kra [HK05b] identified certain characteristic factors for the averages in Theorem 12.1 that exhibit the structure of a nilpotent Lie group, as we describe in this subsection.

Let G be a group. For $g, h \in G$, let $[g, h] = g^{-1}h^{-1}gh$ denote the commutator of g and h. For $A, B \subset G$, let $[A, B]$ the subgroup of G generated by $\{[a, b] : a \in A, b \in B\}$. Define the *lower central series*

$$G = G_0 = G_1 \rhd G_2 \rhd G_3 \rhd \cdots$$

where $G_0, G_1 := G$, $G_{i+1} := [G_i, G]$ for $i \geq 1$. We say that G is *k-step nilpotent* if G_{k+1} is trivial.

A *k-step nilmanifold* is a space of the form G/Γ, where G is a k-step nilpotent Lie group and $\Gamma \subset G$ is a discrete co-compact subgroup. Note that Γ need not be normal. The group G acts naturally on X by left translation. Let $a \in G$, and $T_a \colon X \to X$ given by $T_a(x\Gamma) = (ax)\Gamma$. We assign a translation-invariant Haar measure μ on the Borel σ-algebra $\mathcal{G}/\Gamma$ of X. This makes $(G/\Gamma, \mathcal{G}/\Gamma, \mu, T_a)$ a measure preserving system, called a *k-step nilsystem.*

A system X is a *k-step pro-nilsystem* if it is the inverse limit of an increase sequence of factors in X with each factor being a k-step nilsystem.

Host and Kra constructed characteristic factors for the averages in (12.1) using k-step pro-nilsystems.

Theorem 12.5 (Host-Kra [HK05b])**.** *There is a characteristic factor for (12.1) equivalent to a $(k-1)$-step pro-nilsystem.*

This result reduces the problem of convergence of (12.1) to nilsystems, which have been analysed by Lesigne [Les91] and Leibman [Lei05b] (see [Kra07]).

Example 12.6. Let

$$G = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 0 \end{pmatrix}; \qquad \Gamma = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 0 \end{pmatrix}.$$

Then G/Γ is a 2-step nilmanifold. As a topological space, it is isomorphic to the 2-torus by sending $(x, y) \in (\mathbb{R}/\mathbb{Z})^2$ to the coset of

$$\begin{pmatrix} 1 & 0 & y \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix}.$$

Let

$$a := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix}.$$

The translation T_a corresponds the action on the 2-torus given by $(x, y) \mapsto (x + \alpha, y + x)$, which is the same as that of a skew torus in Example 8.3. Thus the skew torus is a 2-step nilsystem.

References

[BL96] V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc. **9** (1996), no. 3, 725–753. MR 1325795 (96j:11013)

[ET36] P. Erdős and P. Turán, *On some sequences of integers,*, J. London Math. Soc. **11** (1936), 261–264.

[EW11] M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London Ltd., London, 2011.

[FK78] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291 (1979). MR 531279 (82c:28032)

[FK91] ______, *A density version of the Hales-Jewett theorem*, J. Anal. Math. **57** (1991), 64–119. MR 1191743 (94f:28020)

[FKO82] H. Furstenberg, Y. Katznelson, and D. Ornstein, *The ergodic theoretical proof of Szemerédi's theorem*, Bull. Amer. Math. Soc. (N.S.) **7** (1982), no. 3, 527–552.

[Fur77] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256. MR 0498471 (58 #16583)

[Fur81] ______, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, Princeton, N.J., 1981, M. B. Porter Lectures. MR 603625 (82j:28010)

[Gow01] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588. MR 1844079 (2002k:11014)

[Gow07] ______, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. (2) **166** (2007), no. 3, 897–946. MR 2373376 (2009d:05250)

[GRS90] R. L. Graham, B. L. Rothschild, and J. H. Spencer, *Ramsey theory*, second ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., New York, 1990, A Wiley-Interscience Publication. MR 1044995 (90m:05003)

[GT08] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), no. 2, 481–547. MR 2415379 (2009e:11181)

[HK05a] B. Host and B. Kra, *Convergence of polynomial ergodic averages*, Israel J. Math. **149** (2005), 1–19, Probability in mathematics. MR 2191208 (2007c:37004)

[HK05b] ______, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397–488. MR 2150389 (2007b:37004)

[Hos] B. Host, *Convergence of multiple ergodic averages*, Proceedings of school on "Information and Randomness," Chili (to appear).

[HvN42] P. R. Halmos and J. von Neumann, *Operator methods in classical mechanics. II*, Ann. of Math. (2) **43** (1942), 332–350. MR 0006617 (4,14e)

[Kra06] B. Kra, *From combinatorics to ergodic theory and back again*, International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, pp. 57–76. MR 2275670 (2007m:37014)

[Kra07] ______, *Ergodic methods in additive combinatorics*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 103–143. MR 2359470 (2008k:37018)

[Lei05a] A. Leibman, *Convergence of multiple ergodic averages along polynomials of several variables*, Israel J. Math. **146** (2005), 303–315. MR 2151605 (2006c:28016)

[Lei05b] ———, *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold*, Ergodic Theory Dynam. Systems **25** (2005), no. 1, 201–213. MR 2122919 (2006j:37004)

[Les91] E. Lesigne, *Sur une nil-variété, les parties minimales associées à une translation sont uniquement ergodiques*, Ergodic Theory Dynam. Systems **11** (1991), no. 2, 379–391. MR 1116647 (92j:58060)

[NRS06] B. Nagle, V. Rödl, and M. Schacht, *The counting lemma for regular k-uniform hypergraphs*, Random Structures Algorithms **28** (2006), no. 2, 113–179. MR 2198495 (2007d:05084)

[Phe01] R. R. Phelps, *Lectures on Choquet's theorem*, second ed., Lecture Notes in Mathematics, vol. 1757, Springer-Verlag, Berlin, 2001. MR 1835574 (2002k:46001)

[Rok66] V. A. Rokhlin, *Selected topics from the metric theory of dynamical system*, Amer. Math. Soc. Transi. (2) **49** (1966), 171–209.

[Rot53] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109. MR 0051853 (14,536g)

[RS04] V. Rödl and J. Skokan, *Regularity lemma for k-uniform hypergraphs*, Random Structures Algorithms **25** (2004), no. 1, 1–42. MR 2069663 (2005d:05144)

[RS06] ———, *Applications of the regularity lemma for uniform hypergraphs*, Random Structures Algorithms **28** (2006), no. 2, 180–194. MR 2198496 (2006j:05099)

[RS07a] V. Rödl and M. Schacht, *Regular partitions of hypergraphs: counting lemmas*, Combin. Probab. Comput. **16** (2007), no. 6, 887–901. MR 2351689 (2008j:05238)

[RS07b] ———, *Regular partitions of hypergraphs: regularity lemmas*, Combin. Probab. Comput. **16** (2007), no. 6, 833–885. MR 2351688 (2008h:05083)

[Sze69] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104. MR 0245555 (39 #6861)

[Sze75] ———, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245, Collection of articles in memory of Juriĭ Vladimirovič Linnik. MR 0369312 (51 #5547)

[Tao06a] T. Tao, *A quantitative ergodic theory proof of Szemerédi's theorem*, Electron. J. Combin. **13** (2006), no. 1, Research Paper 99, 49 pp. (electronic). MR 2274314 (2007i:37016)

[Tao06b] ———, *A variant of the hypergraph removal lemma*, J. Combin. Theory Ser. A **113** (2006), no. 7, 1257–1280. MR 2259060 (2007k:05098)

[Tao07a] ———, *A correspondence principle between (hyper)graph theory and probability theory, and the (hyper)graph removal lemma*, J. Anal. Math. **103** (2007), 1–45. MR 2373263 (2009e:60022)

[Tao07b] ———, *The ergodic and combinatorial approaches to Szemerédi's theorem*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 145–193. MR 2359471 (2008i:11017)

[Tao09] ———, *Poincaré's legacies, pages from year two of a mathematical blog. Part I*, American Mathematical Society, Providence, RI, 2009. MR 2523047 (2010h:00003)

[Var59] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. **34** (1959), 358–360. MR 0106865 (21 #5595)

[vdW27] B. L. van der Waerden, *Beweis einer baudetschen vermutung*, Nieuw. Arch. Wisk. **15** (1927), 212–216.

[Wal82] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York, 1982. MR 648108 (84e:28017)

[Zim76] R. J. Zimmer, *Extensions of ergodic group actions*, Illinois J. Math. **20** (1976), no. 3, 373–409. MR 0409770 (53 #13522)