

# SZEMERÉDI'S THEOREM: THE COMPACT CASE

## 1. ALMOST PERIODICITY AND COMPACT SYSTEMS

In the last set of notes we discussed weak-mixing systems. We begin this set of notes by introducing a very different class of m.p.s., the *compact* systems. We begin by defining almost periodic functions.

**Definition 1.1** (Almost Periodic functions). Let  $(X, \mu, T)$  be an m.p.s. and suppose that  $f \in L^2(X)$ . Then we say that  $f$  is almost periodic (AP) if the closure of the forward orbit  $(U^n f)_{n \geq 0}$  is compact in  $L^2(X)$  in the topology induced by the norm  $\|\cdot\|_2$ .

It is not hard to see that an equivalent definition is the following: for any  $\varepsilon > 0$ , the set of  $n$  for which  $\|f - U^n f\|_2 \leq \varepsilon$  is syndetic.

**Definition 1.2** (Compact systems). Let  $(X, \mu, T)$  be a m.p.s. Then we say that this system is compact if every function in  $L^2(X)$  is AP.

The archetypal example of a compact system is a circle rotation. Before establishing this, however, let us prove some basic properties of AP functions.

**Lemma 1.3** (AP functions form a closed subalgebra of  $L^2(X)$ ). *Write  $\text{AP}(X)$  for the collection of almost periodic functions in  $L^2(X)$ . Then  $\text{AP}(X)$  is  $U$ -invariant, topologically closed, and also closed under addition and the lattice operations  $\max$  and  $\min$ . Write  $\text{AP}_\infty(X) := \text{AP}(X) \cap L^\infty(X)$ . Then  $\text{AP}_\infty(X)$  is also closed under multiplication.*

*Proof.* It is clear that  $\text{AP}(X)$  is  $U$ -invariant since the forward orbit of  $Uf$  is contained in that of  $f$ .

A subset  $S$  of a complete metric space, such as  $L^2(X)$ , has compact closure if and only if it is covered by finitely many balls of radius  $\varepsilon$ , which we may take to be centred on points in  $S$  if we so desire.

Suppose that  $f_m$ ,  $m = 1, 2, \dots$  are in  $\text{AP}(X)$  and that  $f_m \rightarrow f$  in  $L^2$ . Let  $\varepsilon > 0$ . Then there is some  $m$  such that  $\|f_m - f\|_2 \leq \varepsilon/3$ . For this  $m$ , there are  $r_1, \dots, r_k$  such that, for any  $n \geq 0$ , there is an  $i$  for which  $\|U^n f_m - U^{r_i} f_m\|_2 \leq \varepsilon/3$ . Since  $U$  is an isometry, we have  $\|U^n f - U^n f_m\|_2 \leq \varepsilon/3$  and  $\|U^{r_i} f - U^{r_i} f_m\|_2 \leq \varepsilon/3$ . By the triangle inequality it follows that  $\|U^n f - U^{r_i} f\|_2 \leq \varepsilon$ . Since  $\varepsilon$  was arbitrary, this means that  $f \in \text{AP}(X)$ , and so  $\text{AP}(X)$  is indeed topologically closed.

Now suppose that  $f, g \in \text{AP}(X)$ . Take functions  $f_i, g_j$  with the property that for any  $n$  there are  $i, j$  such that  $\|T^n f - f_i\|_2, \|T^n g - g_j\|_2 \leq \varepsilon/2$ . Consideration of the collection of functions  $f_i + g_j$  then shows that  $f + g \in \text{AP}(X)$ . A moment's reflection confirms that this observation hinges on the fact that the map  $+$  :  $L^2(X) \times L^2(X) \rightarrow L^2(X)$  is uniformly continuous. The same is true of the maps  $\max$  and  $\min$  (this is an exercise on the third example sheet), and of the multiplication map from  $L^\infty(X) \times L^\infty(X)$  to itself, which is in fact uniformly continuous in the  $L^\infty$ -norm.  $\square$

**Lemma 1.4** (Circle rotations are compact). *Let  $X = \mathbb{R}/\mathbb{Z}$  and let  $\mu$  be Lebesgue measure. Let  $\alpha \in \mathbb{R}$  and write  $T : X \rightarrow X$  for the map  $x \mapsto x + \alpha \pmod{1}$ . Then the system  $(X, \mu, T)$  is compact.*

*Proof.* By Lemma 1.3 we need only exhibit a collection of AP functions which are dense in  $L^2(X)$ . We may take the collection of exponentials  $e_m(\theta) := e^{2\pi i m \theta}$ . It is clear that  $e_m$  is an AP function when  $\alpha \in \mathbb{Q}$ , so suppose that  $\alpha$  is irrational. Then we have

$$\|U^n e_m - U^{n'} e_m\|_2 = |e^{2\pi i m \alpha (n - n')} - 1| \leq 10 \|m \alpha (n - n')\|_{\mathbb{R}/\mathbb{Z}}.$$

Now the points  $(m \alpha n')_{n' \in \mathbb{N}}$  are dense in  $\mathbb{R}/\mathbb{Z}$ , and so for any  $n$  there is some  $n' \leq N_0(\varepsilon, m)$  such that  $\|U^n e_m - U^{n'} e_m\|_2 \leq \varepsilon$ . This proves the lemma.  $\square$

## 2. THE SZ PROPERTY FOR COMPACT SYSTEMS

In this section we establish the SZ property for compact systems; this is a triviality compared to the proof for weak-mixing systems.

**Proposition 2.1.** *Suppose that  $(X, \mu, T)$  is a compact system. Then it satisfies the SZ property at level  $k$ , for any  $k \geq 1$ .*

*Proof.* Suppose that  $f \in L^\infty(X)$  is a function with  $f \geq 0$  and  $f$  not zero a.e. Without loss of generality we may suppose that  $\|f\|_\infty = 1$ . Then there is a syndetic set of  $n \geq 1$  with  $\|f - U^n f\|_2 \leq \varepsilon/k2^k$ . Fix such an  $n$ . Since  $U$  is an isometry this implies that  $\|U^{jn} f - U^{(j+1)n} f\|_2 \leq \varepsilon/k2^k$ , and hence by the triangle inequality that  $\|f - U^{jn} f\|_2 \leq \varepsilon/2^k$  for  $j = 1, \dots, k-1$ . Write  $g_j := f - U^{jn} f$ ; then we have

$$\int f \cdot U^n f \dots U^{(k-1)n} f \, d\mu = \int f(f - g_1) \dots (f - g_{k-1}) \, d\mu.$$

This may be split as  $\int f^k \, d\mu$  plus a sum of at most  $2^k$  other terms, each of the form  $\langle g_j, F \rangle$  for some function  $F$  with  $\|F\|_\infty \leq 1$ . Each such term is bounded by  $\varepsilon$ . Therefore

$$\int f \cdot U^n f \dots U^{(k-1)n} f \, d\mu \geq \int f^k \, d\mu - \varepsilon.$$

Since this is true for a syndetic set of  $n$ , we obtain

$$\mathbb{E}_{0 \leq n < N} \int f \cdot U^n f \dots U^{(k-1)n} f \, d\mu \gg_\varepsilon \left( \int f^k \, d\mu - \varepsilon \right).$$

Now  $\int f^k \, d\mu$  is manifestly positive; the result therefore follows upon choosing  $\varepsilon$  sufficiently small.  $\square$

## 3. STRUCTURE AND RANDOMNESS

I would like to encourage you to think of a weak-mixing system as behaving somehow “randomly”. By contrast compact systems are highly “structured” or rigid. It is not the case that every system is either weak-mixing or compact (the 2-dimensional skew torus is an example, and we will discuss it in detail in the next set of notes). Our aim in this section is to show that a system which fails to be weakly mixing does at least have a small piece of compactness.

**Theorem 3.1.** *Suppose that  $(X, \mu, T)$  is not weakly mixing. Then  $\text{AP}(X)$  contains nonconstant functions.*

We shall give two proofs of this theorem. The first uses the Bochner-Herglotz spectral theorem, and gives quite precise information: the system has a nontrivial *eigenfunction*, that is to say function  $f$  such that  $Uf = \lambda f$  for some  $\lambda \in \mathbb{C}$  (which must have modulus one). The second proof is somewhat softer but, because of that, is easier to generalise to the “relative” setting that must be considered in a full proof of Szemerédi’s theorem.

*First proof.* We will show that the Koopman operator  $U$  has a nonconstant eigenfunction  $f$ . If the eigenvalue is  $\lambda$  then we have  $U^n f = \lambda^n f$ , and so the closure of the forward orbit  $(U^n f)_{n \geq 0}$  is homeomorphic to a closed subset of the unit circle, and is hence compact.

Suppose as a hypothesis for contradiction that  $U$  has no nonconstant eigenfunctions. By the Bochner-Herglotz spectral theorem, associated to each  $f \in L^2(X)$  is a regular Borel probability measure  $\mu_f$  on  $\mathbb{R}/\mathbb{Z}$  defined by

$$\langle U^n f, f \rangle = \int e^{-2\pi i n \theta} d\mu_f(\theta).$$

We claim that  $\mu_f(\{\theta\}) = 0$  for  $\theta \neq 0$ ; the argument is almost identical to the one we used when proving the Furstenberg-Sárközy theorem.

Suppose not, and assume that  $f \in L^2(X)$  and  $\theta \in \mathbb{R}/\mathbb{Z}$ ,  $\theta \neq 0$ , are such that  $\mu_f(\{\theta\}) \neq 0$ . By the Hilbert space version of the von Neumann ergodic theorem applied to the operator  $\tilde{U} := e^{-2\pi i \theta} U$  we have

$$\mathbb{E}_{0 \leq n < N} \tilde{U}^n f \rightarrow g$$

in  $L^2$ , for some function  $g \in L^2(X)$ . It is clear that  $\tilde{U}g = g$  and hence that  $Ug = e^{-2\pi i \theta} g$ . Now

$$\begin{aligned} \langle g, f \rangle &= \lim_{N \rightarrow \infty} \mathbb{E}_{0 \leq n < N} \langle \tilde{U}^n f, f \rangle \\ &= \lim_{N \rightarrow \infty} \mathbb{E}_{0 \leq n < N} e^{2\pi i n \theta} \langle U^n f, f \rangle \\ &= \lim_{N \rightarrow \infty} \mathbb{E}_{0 \leq n < N} e^{2\pi i n \theta} \int_0^1 e^{-2\pi i n \phi} d\mu_f(\phi) \\ &= \int_0^1 \lim_{N \rightarrow \infty} \mathbb{E}_{0 \leq n < N} e^{2\pi i n (\theta - \phi)} d\mu_f(\phi) \\ &= \mu_f(\{\theta\}), \end{aligned}$$

the interchange in the order of integration and limit being justified by the dominated convergence theorem. But this is nonzero, and therefore  $g \neq 0$  in  $L^2(X)$ . Since  $U$  has no nonconstant eigenfunctions, this is a contradiction.

Much the same argument works for  $\theta = 0$  if we assume that  $f$  is orthogonal to the constants. In that case we have  $\tilde{U} = U$ , and so each  $\tilde{U}^n f$  and hence  $g$  is orthogonal to the constants. Assuming that  $\mu_f(\{0\}) > 0$ , an identical computation to the one we just gave shows that  $g$  does not vanish almost everywhere. This is again a contradiction.

Thus if  $f \in L^2(X)$  is orthogonal to the constants then the measure  $\mu_f$  is *nonatomic*. In this case we have

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathbb{E}_{0 \leq n < N} |\langle U^n f, f \rangle|^2 &= \lim_{N \rightarrow \infty} \mathbb{E}_{0 \leq n < N} \int_0^1 \int_0^1 e^{-2\pi i(\theta - \phi)n} d\mu_f(\theta) d\mu_f(\phi) \\ &= \lim_{N \rightarrow \infty} \int_0^1 \int_0^1 \mathbb{E}_{0 \leq n < N} e^{-2\pi i(\theta - \phi)n} d\mu_f(\theta) d\mu_f(\phi) \\ &= \int_0^1 \int_0^1 1_{\theta = \phi} d\mu_f(\theta) d\mu_f(\phi), \end{aligned}$$

the justification for the interchange of limit and integral coming from another application of dominated convergence. Since  $\mu_f$  is nonatomic, this is zero. By Cauchy-Schwarz we have

$$\lim_{N \rightarrow \infty} \mathbb{E}_{0 \leq n < N} |\langle U^n f, f \rangle| = 0,$$

Suppose now that  $f \in L^2(X)$  is arbitrary. Applying the preceding estimate to the function  $f - \int f d\mu$ , we obtain

$$\lim_{N \rightarrow \infty} \mathbb{E}_{0 \leq n < N} |\langle T^n f, f \rangle - (\int f d\mu)^2| = 0.$$

Suppose now that  $f, g \in L^2(X)$  are arbitrary. Recalling<sup>1</sup> the depolarisation identity

$$\begin{aligned} \langle T^n f, g \rangle &= \frac{1}{4} (\langle T^n(f + g), f + g \rangle - \langle T^n(f - g), f - g \rangle \\ &\quad + i \langle T^n(f + ig), f + ig \rangle - i \langle T^n(f - ig), f - ig \rangle), \end{aligned}$$

it follows that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{0 \leq n < N} |\langle T^n f, g \rangle - \int f d\mu \int g d\mu| = 0,$$

which is one of the properties we know to be equivalent to weak-mixing.  $\square$

*Second proof of Theorem 3.1.* If  $T$  is not ergodic then the result is trivial. Suppose, then, that  $T$  is ergodic but not weakly-mixing. By the equivalences we established in the last set of notes,  $T \times T$  is not ergodic on  $X \times X$ . Let  $F(x, x') \in L^2(X \times X)$  be a non-constant  $T \times T$ -invariant function; we will use  $F$  to fashion an AP function on  $X$ . The marginal function  $G(x') = \int F(x, x') d\mu(x)$  is  $T$ -invariant, since by change of variables we have

$$\int F(x, Tx') d\mu(x) = \int F(Tx, Tx') d\mu(x) = \int F(x, x') d\mu(x).$$

As  $T$  is being assumed ergodic, this implies that  $G$  is constant a.e.; replacing  $F$  by  $F - G$ , we may thus assume that  $\int F(x, x') d\mu(x) = 0$  a.e.

Now  $F$  does not vanish a.e., and so there is some  $\phi \in L^2(X)$  such that

$$f(x) := \int F(x, x') \phi(x') d\mu(x') \neq 0$$

for a set of  $x$  of positive measure. Indeed if there were not then  $F(x, x')$  would certainly be orthogonal to all functions  $\phi(x')\psi(x)$ , and the closed linear span of such functions

<sup>1</sup>Note that everything we have done is valid for  $\mathbb{C}$ -valued functions.

is all of  $L^2(X \times X)$ . The function  $f$  is not constant since, by our assumption on  $F$ ,  $\int f d\mu = 0$ .

We claim that  $f \in \text{AP}(X)$ . By change of variables and the  $T \times T$ -invariance of  $F$  we have

$$\begin{aligned} U^n f(x) &= \int F(T^n x, x') \phi(x') d\mu(x') \\ &= \int F(T^n x, T^n x') \phi(T^n x') d\mu(x') \\ &= \int F(x, x') U^n \phi(x') d\mu(x'), \end{aligned}$$

and so  $U^n f$  lies in the image of some closed ball in  $L^2(X)$  under the operator  $S : L^2(X) \rightarrow L^2(X)$  defined by

$$S\psi(x) := \int F(x, x') \psi(x') d\mu(x').$$

But it is well known that integral operators such as  $S$  are compact; indeed they are Hilbert-Schmidt operators. See N. Young's *An introduction to Hilbert Space*, Chapter 8, for further details. It follows immediately from the definition of a compact operator that the forward orbit  $(U^n f)_{n \geq 0}$  has compact closure.  $\square$

#### 4. FACTORS

In this section we make some further, at times rather nebulous, remarks on factors. We assume that the transformation  $T$  is ergodic; for the purposes of establishing the SZ property this assumption is quite anodyne, as something called the *ergodic decomposition* allows us to decompose an arbitrary measure-preserving transformation into an average of ergodic ones. (Interpreting ergodicity in terms of the geometry of  $\mathcal{M}(X)$ , as we did in the middle of the course, this follows from *Choquet theory*, which allows one to write any point inside a convex body as a convex combination of its extreme points.

We have established that if  $(X, \mu, T)$  is not weakly-mixing then it contains nonconstant AP functions. In our first proof of this statement we actually established rather more, namely that there is a nontrivial eigenfunction  $f \in L^2(X)$ . Now the modulus  $|f(x)|$  is  $T$ -invariant and hence, since  $T$  is ergodic, is constant a.e. Rescaling if necessary, we may assume that  $|f(x)| = 1$  a.e., in which case  $f$  provides a measurable map from  $X$  to the unit circle in the complex plane and hence, upon composing with the logarithm, to  $\mathbb{R}/\mathbb{Z}$ . If the eigenvalue corresponding to  $f$  is  $e^{2\pi i\alpha}$  then the following diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ \downarrow f & & \downarrow f \\ \mathbb{R}/\mathbb{Z} & \xrightarrow{R_\alpha} & \mathbb{R}/\mathbb{Z} \end{array}$$

Note that the circle rotation system  $(\mathbb{R}/\mathbb{Z}, m, R_\alpha)$  is compact; we say that  $X$  has a *compact factor*.

\*\*In fact the second proof that AP contains a nonconstant function may also be used to furnish a compact factor  $(Y, \nu, S)$  of  $(X, \mu, T)$ . It is important to have this proof to hand when it comes to dealing with the failure of *relative weak-mixing*, as it generalises much more immediately than the argument using spectral theory.

The basic idea is to use the function  $F$  to define a  $T$ -invariant “distance” (which may be zero for distinct points) via

$$d(x_1, x_2) := \int |F(x_1, x') - F(x_2, x')| dx';$$

the factor space  $Y$  may then be constructed as the quotient of  $X$  by  $d$ .\*\*