SZEMERÉDI'S THEOREM: THE WEAK-MIXING CASE

In this section we being our discussion of the Furstenberg-Katznelson-Ornstein proof of Szemerédi's theorem. Let us begin by recalling the statement of Szemerédi's theorem.

Theorem 0.1 (Szemerédi)**.** *Let $\alpha > 0$ be a real number, and let $k \geqslant 2$ be an integer. Then there is an $N_0(\alpha, k)$ with the following property. If $N > N_0(\alpha, k)$ and if $A \subseteq [N]$ is a set with cardinality at least αN, then A contains an arithmetic progression of length k.*

1. The SZ property for measure-preserving systems

In the last set of notes we deduced Sárközy's theorem from a statement about measure-preserving systems. In this section we formulate a recurrence property of m.p.s. which corresponds to Szemerédi's theorem in the same way. Here, and throughout this discussion, we write $U = U_T$ for the Koopman operator associated to the transformation T.

Definition 1.1 (The SZ property)**.** We say that a m.p.s. (X, μ, T) has the SZ-property at level k if, for any function $f \in L^\infty(X)$ with $f \geqslant 0$ and f not equal to zero almost everywhere we have

$$\liminf_{N \to \infty} \mathbb{E}_{0 \leqslant n < N} \int f \cdot U^n f \cdots U^{(k-1)n} f d\mu > 0.$$

Note that we ask for more than just *one* value of $n \geqslant 1$ for which $\int f \cdot U^n f \cdots U^{(k-1)n} f d\mu > 0$, a condition which has a more obvious analogy with the condition $\langle f, U^{n^2} f\rangle > 0$ that we used in proving the Furstenberg-Sárközy theorem.

Here is the promised recurrence result.

Theorem 1.2 (Ergodic-theoretical formulation of Szemerédi's theorem)**.** *Every m.p.s. has the* SZ *property at level k.*

It turns out that this result is *equivalent* to Szemerédi's theorem (in the sense that the deduction of each form the other is reasonably straightforward). We will only show that Theorem 1.2 *implies* Szemerédi's theorem, leaving the other direction to the third example sheet. The argument is very similar to that we used to deduce Sárközy's theorem from the measure-theoretic version: take a putative sequence of counterexamples to Szemerédi's theorem, and use them to fashion a measure-preserving system which violates Theorem 1.2.

Deduction of Szemerédi's theorem from Theorem 1.2. Suppose the theorem is false; then for every N there is some set $A_N \subseteq [N]$ with $|A_N| \geqslant \delta N$ but not containing any nontrivial k-term arithmetic progression.

Consider the (infinite) set

$$A := \bigcup_{N=1}^{\infty} (N^2 + A_N).$$

The construction now proceeds exactly as in the proof of Sárközy's theorem: construct the space $\Omega = \{0,1\}^{\mathbb{Z}}$, then the space $X = \overline{(T^n \omega_A)_{n\in\mathbb{Z}}}$. Let S be the intersection of X with the cylinder set $\{\omega \in \Omega : \omega(0) = 1\}$. Define measures μ_N exactly as before, and let μ be some weak limit of them. Then $\mu(S) \geqslant \delta$, as before. Once again, the weak limit μ is T-invariant.

Now apply Theorem 1.2 with $f = 1_S$ to obtain

$$\liminf_{N\to\infty} \mathbb{E}_{0\leqslant n<N}\mu(S \cap T^{-n}S \cap \cdots \cap T^{-(k-1)n}S) > 0.$$

As a consequence of this, there is certainly some $n > 0$ such that $\mu(S \cap T^{-n}S \cap \cdots \cap T^{-(k-1)n}S) > 0$. For large enough N this implies (since $\mu_N \to \mu$ weakly) that $\mu_N(S \cap T^{-n}S \cap \cdots \cap T^{-(k-1)n}S) > 0$, which immediately implies that the finite set A_N contains a non-trivial k-term arithmetic progression. □

If one had to explain the proof of Theorem 1.2 in a short paragraph, one might say the following. The are two very different types of m.p.s, called weakly-mixing and compact systems. It is not too hard to establish the SZ property for either type, and we will do this in the next two sections. The beef of the argument is then to show how an arbitrary m.p.s. can, in a certain sense, be decomposed into pieces which have a behaviour resembling one of these two extremes.

2. Weak-mixing systems and their basic properties

In this and subsequent sections we will write $o_{a,b,c,\ldots;N\to\infty}(1)$ for a quantity which tends to zero as $N \to \infty$. The speed of convergence to zero is allowed to depend on the parameters $a, b, c, \ldots$.

Definition 2.1 (Weak-mixing systems)**.** Suppose that (X, μ, T) is a m.p.s. We say that this system is weak-mixing if, for all measurable sets $A, B \subseteq X$, we have

$$\mathbb{E}_{0\leqslant n<N}|\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| \to 0$$

as $N \to \infty$.

To give a little intuition, let us remark that weak-mixing implies ergodicity. Indeed if $T^{-1}A = A$ then the weak-mixing condition clearly implies that $\mu(A \cap B) = \mu(A)\mu(B)$ for all measurable B. Taking $B = A$, we see that $\mu(A) = 0$ or 1. Weak-mixing is a much stronger assumption than ergodicity, however. The most basic example of an ergodic system, rotation on $\mathbb{R}/\mathbb{Z}$ by some irrational number α, fails to be weak-mixing as can be seen by taking $A = B = I$ for a suitable interval I.

It is traditional, and I think rather pleasant, to supply a list of properties that are equivalent to weak-mixing. The equivalence of the properties in the following proposition is relatively easy to establish. In the next set of notes we will establish that weak mixing is equivalent all eigenfunctions of the Koopman operator U_T acting on $L^2(X)$ being constant a.e., a fact which lies somewhat deeper.

Proposition 2.2 (Equivalent notions of weak-mixing). *Suppose that (X,μ,T) is a m.p.s. Then the following are equivalent.*

(i) *T is weak-mixing;*
(ii) *$\mathbb{E}_{0\leqslant n<N}|\langle f,U^n g\rangle - \int f d\mu \int g d\mu| \to 0$ for all $f,g\in L^2(X)$;*
(iii) *$T\times T$ is weakly-mixing;*
(iv) *$T\times S$ is ergodic on $X\times Y$ for any ergodic m.p.s. (Y,ν,S);*
(v) *$T\times T$ is ergodic.*

Proof. Statement (ii) obviously implies (i), and conversely (i) implies (ii) by approximating f and g by simple-measurable functions.

To see that (i) implies (iii), suppose that $\varepsilon>0$. Let A,B,C,D be four measurable subsets of X. Then (i) is easily seen to imply that $|\mu(T^{-n}A\cap B)-\mu(A)\mu(B)|\leqslant\varepsilon$ except for $o_{\varepsilon;N\to\infty}(1)N$ values of $n\leqslant N$, and also $|\mu(T^{-n}C\cap D)-\mu(C)\mu(D)|\leqslant\varepsilon$ except for $o_{\varepsilon;N\to\infty}(1)N$ values of $n\leqslant N$. These two facts therefore hold simultaneously for $o_{\varepsilon;N\to\infty}(1)N$ values of $n\leqslant N$. For all such n we have

$$\begin{aligned}&|(\mu\times\mu)((T\times T)^{-n}(A\times C)\cap(B\times D))-(\mu\times\mu)(A\times C)(\mu\times\mu)(B\times D)|\\&=|\mu(T^{-n}A\cap C)\mu(T^{-n}B\cap D)-\mu(A)\mu(B)\mu(C)\mu(D)|\\&\leqslant|\mu(T^{-n}A\cap C)-\mu(A)\mu(C)|\mu(T^{-n}B\cap D)+|\mu(T^{-n}B\cap D)-\mu(B)\mu(D)|\mu(A)\mu(C)\\&\leqslant 2\varepsilon.\end{aligned}$$

It follows that

$$\mathbb{E}_{0\leqslant n<N}|(\mu\times\mu)((T\times T)^{-n}(A\times C)\cap(B\times D)-\mu(A)\mu(B)\mu(C)\mu(D)|\leqslant 2\varepsilon+o_{\varepsilon;N\to\infty}(1).$$

Since $\varepsilon>0$ was arbitrary, this quantity tends to zero with N. This verifies that $T\times T$ is weak mixing when applied to product sets $A\times C, B\times D$. However we may approximate arbitrary measurable subsets of $X\times X$ by finite sums of such product sets (I have put this on the third example sheet), and so $T\times T$ is indeed weakly-mixing.

To see that (iii) implies (iv), we show that for all μ-measurable A,B and ν-measurable C,D we have

$$\mathbb{E}_{0\leqslant n<N}(\mu\times\nu)((T\times S)^{-n}(A\times C)\cap(B\times D))\to\mu(A)\mu(B)\mu(C)\mu(D). \tag{2.1}$$

Approximating measurable subsets of $X\times Y$ by finite sums of product sets, it then follows that

$$\mathbb{E}_{0\leqslant n<N}(\mu\times\nu)((T\times S)^{-n}E\cap F)\to(\mu\times\nu)(E)(\mu\times\nu)(F)$$

for all measurable E,F. Supposing that E is $T\times S$-invariant and taking $E=F$, we conclude that $(\mu\times\nu)(E)=0$ or 1, and so $T\times S$ is indeed ergodic.

It remains to establish (2.1). The left-hand side minus the right is simply

$$\mathbb{E}_{0\leqslant n<N}\big(\mu(T^{-n}A\cap C)\nu(S^{-n}B\cap D)-\mu(A)\mu(C)\nu(B)\nu(D)),$$

which equals

$$\mathbb{E}_{0\leqslant n<N}(\mu(T^{-n}A\cap C)-\mu(A)\mu(C))\nu(S^{-n}B\cap D)+\mu(A)\mu(C)(\nu(S^{-n}B\cap D)-\nu(B)\nu(D)).$$

This is bounded by

$$\mathbb{E}_{0\leqslant n<N}|\mu(T^{-n}A\cap C)-\mu(A)\mu(C)|+|\mathbb{E}_{0\leqslant n<N}(\nu(S^{-n}B\cap D)-\nu(B)\nu(D))|$$

(note carefully the position of the modulus signs). Now if $T \times T$ is weak-mixing then it is easy to see that T is too, and so the first term tends to zero. The second term tends to zero because S is ergodic; in fact one needs nothing more than the *weak* L^2-ergodic theorem, which was the one with the softest proof.

To see that (iv) implies (v), first take S to be a one-point system to conclude that T itself is ergodic. Then, taking $S = T$, we see that $T \times T$ is ergodic.

Finally we must show that (v) implies (i), to which end it suffices (by Cauchy-Schwarz) to show that

$$\mathbb{E}_{0 \leqslant n < N} |\mu(T^{-n}A \cap B) - \mu(A)\mu(B)|^2 \rightarrow 0.$$

Expanding out the square, this may (with a little effort) be rewritten as

$$\mathbb{E}_{0 \leqslant n < N}(\mu \times \mu)((T \times T)^{-n}(A \times A) \cap (B \times B)) - \\ 2\mu(A)\mu(B)\mathbb{E}_{0 \leqslant n < N}(\mu \times \mu)((T \times T)^{-n}(A \times X) \cap (B \times X)) + \mu(A)^2\mu(B)^2.$$

Since $T \times T$ is ergodic, the weak L^2-ergodic theorem may again be pressed into service, allowing us to conclude that this tends to 0, as required. □

Remark for congoscenti. The list of equivalent properties characterising weak-mixing is somewhat reminiscent of the list of equivalent properties characterising a pseudorandom graph, or a pseudorandom set $A \subseteq \mathbb{Z}/N\mathbb{Z}$ (that is, a set with no large Fourier coefficients). The correspondence between these two concepts is not quite as close as might appear at first sight. The notion of weak-mixing definitely takes advantage of the infinitary nature of a typical m.p.s., and in particular the fact that the space $L^2(X)$ is "not much bigger" than $\mathbb{Z}$ (since this Hilbert space is separable). In some nebulous sense a set $A \subseteq \mathbb{Z}/N\mathbb{Z}$ might be thought of as weakly mixing if all sets in the algebra generated by A (that is, shifts $U^h 1_A$ with $h = O(1)$ and derivatives $\Delta_h 1_A$) are pseudorandom. We have been deliberately vague about what is meant by this as there does not seem to be a pleasant discretisation of the notion.

3. Weak-mixing and the SZ property

Theorem 3.1. *Suppose that (X, μ, T) is weak-mixing. Then it has the* SZ *property at level k.*

Before embarking on the main part of the proof, we isolate two lemmas. We shall set $\Delta_h f := fU^h f$.

Lemma 3.2. *Suppose that (X, μ, T) is weak-mixing and that $f \in L^\infty(X)$ is a function with $\int f d\mu = 0$. Then $|\int \Delta_h f d\mu| \leqslant \varepsilon$ for all except $o_\varepsilon(H)$ values of h, $0 \leqslant h < H$.*

Proof. From Proposition 2.2 (2) with $f = g$ we immediately have

$$\mathbb{E}_{0 \leqslant h < H} \left| \int \Delta_h f d\mu \right| = o_{H \rightarrow \infty}(1).$$

This immediately implies the result. □

Lemma 3.3 (van der Corput inequality)**.** *Let $(a_n)_{n \in \mathbb{Z}}$ be a sequence of real numbers with $|a_n| \leqslant 1$, and let H, N be integers with $1 \leqslant H \leqslant N$. Then*

$$|\mathbb{E}_{0 \leqslant n < N} a_n|^2 \leqslant 2\mathbb{E}_{0 \leqslant h < H}(1 - \frac{|h|}{H})\mathbb{E}_{0 \leqslant n < N} a_n a_{n+h} + O(H/N).$$

Proof. For each h with $0 \leqslant h < H$ we have

$$\mathbb{E}_{0 \leqslant n < N} a_{n+h} = \mathbb{E}_{0 \leqslant n < N} a_n + O(H/N).$$

Taking averages over $0 \leqslant h < H$ we have

$$\mathbb{E}_{0 \leqslant n < N} a_n = \mathbb{E}_{0 \leqslant n < N} \mathbb{E}_{0 \leqslant h < H} a_{n+h} + O(H/N).$$

By the Cauchy-Schwarz inequality we thus obtain

$$\begin{aligned} |\mathbb{E}_{0 \leqslant n < N} a_n|^2 &\leqslant \mathbb{E}_{0 \leqslant n < N} |\mathbb{E}_{0 \leqslant h < H} a_{n+h}|^2 + O(H/N) \\ &= \mathbb{E}_{0 \leqslant n < N} \mathbb{E}_{0 \leqslant h, h' < H} a_{n+h} a_{n+h'} + O(H/N) \\ &= 2\mathbb{E}_{0 \leqslant n < N} \mathbb{E}_{0 \leqslant h < H} (1 - \frac{|h|}{H}) a_n a_{n+h} + O(H/N), \end{aligned}$$

as required.

Proof of Theorem 3.1. We shall prove, by induction on k, that if $f_0, \dots, f_k$ are in the unit ball of $L^\infty(X)$ and if $\varepsilon > 0$ then

$$|\int f_0 \cdot U^n f_1 \cdots U^{kn} f_k d\mu - \int f_0 d\mu \cdots \int f_k d\mu| \leqslant \varepsilon$$

for all except $o_{\varepsilon,k;N\to\infty}(1)N$ values of $n < N$; this is easily seen to imply the SZ property at level $k+1$, and the base case $k = 1$ follows immediately from the basic properties of weak-mixing. We begin by reducing to the case where at least one of the integrals $\int f_i d\mu$, $i = 1, \dots, k$, is zero. Writing

$$L_n(g_0, \dots, g_k) := \int g_0 \cdot U^n g_1 \cdots U^{kn} g_k \, d\mu,$$

$c_i := \int f_i \, d\mu$ and $f_i' := f_i - c_i$ we have

$$L_n(f_0, \dots, f_k) = c_0 \dots c_k + L_n(f_0', c_1, \dots, c_k) + (2^{k+1} - 2) \text{ terms of the form } L_n(\dots).$$

Here at least one of the unspecified functions in $L_n(\dots)$ is an f_i', $i = 1, \dots, k$, and the integral of each f_i' is zero. Since $L_n(f_0', c_1, \dots, c_k) = 0$, it suffices to show that each expression $L_n(\cdots)$ in the above is bounded by $\varepsilon/2^{k+1}$ for $o_{\varepsilon,k}(N)$ values of $n < N$.

Suppose, then, that $\int f_i \, d\mu = 0$ for some $i = 1, \dots, k$ and examine the expression

$$\int |\mathbb{E}_{0 \leqslant n < N} f_0 \cdot U^n f_1 \cdots U^{kn} f_k|^2 \, d\mu. \tag{3.1}$$

Since $\|f_0\|_\infty \leqslant 1$, this is immediately seen to be at most

$$\int |\mathbb{E}_{0 \leqslant n < N} U^n f_1 \cdots U^{kn} f_k|^2 \, d\mu.$$

Let H, $1 \leqslant H \leqslant N$, be a quantity to be specified later. By van der Corput's lemma and a switch in the order of integration and summation over $n < N$, (3.1) is bounded by

$$2\mathbb{E}_{0 \leqslant h < H} (1 - \frac{|h|}{H}) \mathbb{E}_{0 \leqslant n < N} \int U^n \Delta_h f_1 \cdots U^{kn} \Delta_{kh} f_k \, d\mu + O(H/N)$$

where, recall, we have written $\Delta_h f := f U^h f$. Making the change of variables $x' := T^n x$, this may be rewritten as

$$2\mathbb{E}_{0 \leqslant h < H} (1 - \frac{|h|}{H}) \mathbb{E}_{0 \leqslant n < N} \int \Delta_h f_1 \cdot U^n \Delta_{2h} f_2 \cdots U^{(k-1)n} \Delta_{kh} f_k \, d\mu + O(H/N). \tag{3.2}$$

Now for each fixed h we have, by the induction hypothesis,

$$|\int \Delta_h f_1 \cdot U^n \Delta_{2h} f_2 \cdots U^{(k-1)n} \Delta_{kh} f_k \, d\mu - \int \Delta_h f_1 d\mu \cdots \int \Delta_{kh} f_k d\mu| \leqslant \varepsilon$$

for all but $o_{\varepsilon,k,h;N\to\infty}(1)N$ values of $n < N$. Taking averages over $n < N$, we see that

$$\begin{aligned}&\mathbb{E}_{0\leqslant n<N} \int \Delta_h f_1 \cdot U^n \Delta_{2h} f_2 \cdots U^{(k-1)n} \Delta_{kh} f d\mu \\ &\qquad = \int \Delta_h f_1 d\mu \dots \int \Delta_{kh} f_k d\mu + O(\varepsilon) + o_{\varepsilon,k,h;N\to\infty}(1).\end{aligned}$$

Taking averages over $0 \leqslant h < H$, we obtain that the quantity in (3.2) is at most

$$2\mathbb{E}_{0\leqslant h<H} |\int \Delta_h f_1 d\mu \dots \int \Delta_{kh} f_k d\mu| + O(\varepsilon) + o_{\varepsilon,k,H;N\to\infty}(1) + O(H/N). \tag{3.3}$$

Now we are assuming that $\int f_i d\mu = 0$ for some i, and so by Lemma 3.2 we have $|\int \Delta_h f_i d\mu| \leqslant \varepsilon$ for all but $o_{\varepsilon;H\to\infty}(1)H$ values of h, $0 \leqslant h < H$. It follows that

$$\mathbb{E}_{0\leqslant h<H} |\int \Delta_h f_1 d\mu \dots \int \Delta_{kh} f_k d\mu| = O(\varepsilon) + o_{\varepsilon;H\to\infty}(1). \tag{3.4}$$

Recall that (3.1), the quantity we are interested in, was bounded by (3.2). This in turn we have bounded using (3.3) and (3.4). Putting all this together we see that

$$\int |\mathbb{E}_{0\leqslant n<N} f_0 U^n f_1 \dots U^{kn} f_k|^2 \, d\mu = O(\varepsilon) + o_{\varepsilon;H\to\infty}(1) + o_{\varepsilon,k,H,N\to\infty}(1) + O(H/N).$$

For fixed ε this can be made less than 4ε by choosing first $H > H_0(\varepsilon)$ and then $N > N_0(\varepsilon, k, H)$. Since $\varepsilon > 0$ was arbitrary, it follows that

$$\int |\mathbb{E}_{0\leqslant n<N} f_0 U^n f_1 \dots U^{kn} f_k|^2 \, d\mu = o_{k;N\to\infty}(1),$$

and hence by Cauchy-Schwarz that

$$\mathbb{E}_{0\leqslant n<N} \int f_0 U^n f_1 \dots U^{kn} f_k d\mu = o_{k;N\to\infty}(1). \tag{3.5}$$

Now we may apply precisely the same analysis to the product system $(X\times X, \mu\times\mu, T\times T)$ which, by Proposition 2.2, is weak-mixing. Working with the functions $f_i \otimes f_i$, the analogue of (3.5) is

$$\mathbb{E}_{0\leqslant n<N} |\int f_0 U^n f_1 \dots U^{kn} f_k|^2 \, d\mu = o_{k;N\to\infty}(1).$$

This immediately implies that $|\int f_0 U^n f_1 \dots U^{kn} f_k d\mu| \leqslant \varepsilon$ except for $o_{\varepsilon,k;N\to\infty}(1)N$ values of $n < N$, which is the inductive hypothesis we set out to establish. □