SZEMERÉDI'S THEOREM: A STUDY OF THE SKEW TORUS

We have proved that the SZ property holds for weak-mixing systems and for compact systems. In this section we study the two-dimensional skew torus $(\tilde{X}, \tilde{\mu}, \tilde{T})$, viewed as a circle extension of the circle. Here $T : \tilde{X} \to \tilde{X}$ is the map defined by $T(x, \theta) = (x + \alpha, \theta + x)$, and we shall assume that $\alpha \notin \mathbb{Q}$ so that this is an ergodic system. As we shall see shortly, such a system is neither compact nor weakly mixing; it is in fact "relatively compact" over the base system (X, μ, T) which is a circle rotation by α.

Theorem 0.1. *The skew torus has the SZ property at level k, for all k.*

An important ingredient in the proof of this statement is the fact that the circle rotation (X, μ, T) satisfies the SZ property at all levels, which follows from the fact that such rotations are compact and the main result of the previous set of notes. The relevance of this is unsurprising, since a function $f \in L^\infty(X)$ falsifying the SZ property could easily be lifted to a function $\tilde{f} \in L^\infty(X)$ without the property.

More surprisingly, we use van der Waerden's theorem. The use of this provides what is probably the simplest argument, but it is possible to proceed without. The paper of Furstenberg, Katznelson and Weiss may be consulted for details.

1. The relative compactness property

Write m_x for Lebesgue measure in the circle fibre above the point $x \in X$. We say that a function $f \in L^2(\tilde{X})$ is AP relative to X if, for every $\varepsilon > 0$, there is a collecton $\{g_1, \dots, g_k\} \subseteq L^2(\tilde{X})$ with the following property. For every $n \geqslant 0$ and for a.e. $x \in X$ there is some $i = i(n, x)$ such that $\|\tilde{U}^n \tilde{f} - g_i\|_{L^2(m_x)} \leqslant \varepsilon$. Write $\mathrm{AP}(\tilde{X}|X)$ for the collection of all $f \in L^2(\tilde{X})$ which are AP relative to X.

Proposition 1.1. $\mathrm{AP}(\tilde{X}|X)$ *is dense in* $L^2(\tilde{X})$.

Proof. It is easy to see that $\mathrm{AP}(\tilde{X}|X)$ is closed under addition and scalar multiplication. It therefore suffices to check that the functions $f_{r,s}(x, \theta) := e^{2\pi i(rx + s\theta)}$ are all almost periodic relative to X.

Now a short computation confirms that $\tilde{T}^n(x, \theta) = (x + n\alpha, \theta + nx + \frac{1}{2}\alpha n(n-1))$, and so

$$U^n f_{r,s}(x, \theta) = e^{2\pi i(rn\alpha + \frac{1}{2}s\alpha n(n-1))} e^{2\pi i snx} f_{r,s}(x, \theta).$$

It is easy to see that we can take the functions g_j to be simply $g_j := e^{2\pi i j/L} f_{r,s}$, $j = 1, \dots, L$, where $L = \lfloor 10/\varepsilon \rfloor$ (say). □

2. Proof of the SZ property: reduction to AP functions

In this section we establish a simple technical lemma, which allows us to reduce the establishment of the SZ property for arbitrary nonnegative $f \in L^\infty(X)$ to the case when

f is bounded and almost periodic relative to X. It is clear that such a reduction is a consequence of the following lemma.

Lemma 2.1. *Suppose that $f \in L^\infty(\tilde{X})$ has $\|f\|_\infty \leqslant 1$, $f \geqslant 0$ for all x and $\int f d\mu > 0$. Then there is a function $f' \in \mathrm{AP}(\tilde{X}|X)$ with the same properties, and such that $f(x) \geqslant cf'(x)$ for all x and for some absolute constant $c > 0$.*

Proof. If f and $\tilde{f}$ are close in L^2 and if $\tilde{f}$ is almost periodic over X then f is in some sense almost periodic "in most vertical fibres". The idea of the proof is to make proper sense of this notion, and then to strip out the "bad" fibres from f.

We begin by restricting to fibres on which f is reasonably large. There is a measurable set $A \subseteq X$ with $\mu(A) > 0$ and a $\delta > 0$ such that $\int f dm_x \geqslant \delta$ for $x \in A$. Let $\varepsilon_1 > \varepsilon_2 > \dots$ be a sequence such that $\sum_i \varepsilon_i < \frac{1}{10}\mu(A)$. For each i there is some $f_i \in \mathrm{AP}(\tilde{X}|X)$ such that $\|f - f_i\|_2 \leqslant \varepsilon_i$. Let $g_{i,1}, \dots, g_{i,K_i}$ be the collection of functions associated to f_i and the parameter ε_i in the definition of relative almost-periodicity. Then for every $n \geqslant 0$, for a.e. $x \in X$ and for each i there is some j, $1 \leqslant j \leqslant K_i$, such that $\|\tilde{U}^n f_i - g_{i,j}\|_{L^2(m_x)} \leqslant \varepsilon_i$. Adding the zero function $g_{i,0} := 0$ to the collection, it follows that for any fixed set $S \subseteq X$, for any $n \geqslant 0$, a.e. $x \in X$ and for each i there is some j, $0 \leqslant j \leqslant K_i$, such that

$$\|\tilde{U}^n(f_i 1_S) - g_{i,j}\|_{L^2(dm_x)} \leqslant \varepsilon_i. \tag{2.1}$$

Now the closeness of f and f_i easily implies that there is a set E_i, $\mu(E_i) \leqslant \varepsilon_i$, such that $\|f - f_i\|_{L^2(m_x)} \leqslant \sqrt{\varepsilon_i}$ for $x \notin E_i$. Define $S := A \setminus \bigcup_{i=1}^\infty E_i$. By construction we have $\mu(S) \geqslant \frac{9}{10}\mu(A)$ and

$$\|f 1_S - f_i 1_S\|_{L^2(dm_x)} \leqslant \sqrt{\varepsilon_i}$$

for a.e. $x \in X$. Since $\tilde{T}$ maps m_x to $m_{\tilde{T}x}$ it follows that

$$\|\tilde{U}^n(f 1_S) - \tilde{U}^n(f_i 1_S)\|_{L^2(dm_x)} \leqslant \sqrt{\varepsilon_i}$$

for a.e $x \in X$. Comparing with (2.1), we see that for every $n \geqslant 0$, for a.e. $x \in X$ and for each i there is some j, $0 \leqslant j \leqslant K_i$, such that

$$\|\tilde{U}^n(f 1_S) - g_{i,j}\|_{L^2(m_x)} \leqslant \varepsilon_i + \sqrt{\varepsilon_i}.$$

It follows that $f' := f 1_S$ lies in $\mathrm{AP}(\tilde{X}|X)$. It is clear that this function has all of the stated properties. □

3. Proof of the SZ property: an application of van der Waerden's theorem

Let $W(r, k)$ be the least integer such that, if $n \geqslant W(r, k)$ and $\{1, \dots, n\}$ is r-coloured, there is a monochromatic progression of length k.

Proof of Theorem 0.1. By the main result of the previous section it suffices to establish the SZ property for functions $f \in \mathrm{AP}(\tilde{X}|X)$ with $\|f\|_\infty \leqslant 1$. We are assuming, of course, that $f \geqslant 0$ and that $\int f dm_x > 0$. There is a measurable set $A \subseteq X$ with $\mu(A) > 0$ and a $\delta > 0$ such that $\int f dm_x \geqslant \delta$ for $x \in A$ (in fact, this was built into the construction of the function output by Lemma 2.1). Let $\varepsilon = \varepsilon(\delta, k)$ be a small positive quantity to be specified later, and let $\{g_1, \dots, g_K\}$ be a collection of functions associated to f by the definition of relative almost-periodicity, with parameter ε. Let $L = W(K, k)$. Since the SZ property at level L holds for X, we see that for a proportion at least $\eta > 0$ of $n < N$ we have

$$\mu(A \cap T^{-n}A \cap \cdots \cap T^{-(L-1)n}A) \geqslant \delta' > 0.$$

Write S_N for the set of such n, and for each such n set $E_n := A \cap T^{-n}A \cap \cdots \cap T^{-(L-1)n}A$.

Look at the functions $U^{mn}f$, $m = 0, \dots, L-1$. For each m, each $n \in S_N$ and for a.e. $x \in X$ it follows from the defining property of relative almost-periodicity that there is at least one $i = i(m, n, x)$ such that $\|U^{mn}f - g_i\|_{L^2(m_x)} \leqslant \varepsilon$. Choose a specific measurable[1] map $(m, n, x) \mapsto i(m, n, x)$. For fixed n, x this gives a K-colouring of $\{0, \dots, L-1\}$, and by van der Waerden's theorem there is a monochromatic k-term progression. Writing $\{a, a+d, \dots, a+(k-1)d\}$ for this progression, we see that

$$\|U^{(a+jd)n}f - U^{an}f\|_{L^2(m_x)} \leqslant 2\varepsilon \tag{3.1}$$

for all $j \leqslant k-1$, and for almost every $x \in X$. Note carefully that a and d may depend on both $n \in S_N$ and $x \in X$. Fix n; by the pigeonhole principle we may find a set $E'_n \subseteq E_n$ with $\mu(E'_n) \geqslant L^{-2}\mu(E_n)$ and some fixed a, d which work for every $x \in E'_n$ (they may depend on n).

Integrate over the vertical fibre above x. Using the inequality

$$\left|\int F_1 \dots F_k d\mu\right| \leqslant \|F_j\|_1 \max_{i \neq j} \|F_i\|_\infty \leqslant \|F_j\|_2 \max_{i \neq j} \|F_i\|_\infty$$

together with (3.1) we have

$$\int_{\mathbb{R}/\mathbb{Z}} U^{an}f \cdot U^{(a+d)n}f \cdots U^{(a+(k-1)d)n}f dm_x = \int_{\mathbb{R}/\mathbb{Z}} (U^{an}f)^k dm_x - O_k(\varepsilon).$$

Integrating over $x \in E'_n$ and using non-negativity, we have

$$\int_{\tilde{X}} U^{an}f \cdot U^{(a+d)n}f \cdots U^{(a+(k-1)d)n}f d\tilde{\mu} \geqslant \int_X 1_{E'_n}(x) \int_{\mathbb{R}/\mathbb{Z}} (U^{an}f)^k dm_x - O_k(\varepsilon)\mu(E'_n).$$

Since $E'_n \subseteq E_n$ and $0 \leqslant a \leqslant L-1$, we have $\int_{\mathbb{R}/\mathbb{Z}} U^{an}f dm_x \geqslant \delta$ and hence $\int_{\mathbb{R}/\mathbb{Z}} (U^{an}f)^k dm_x \geqslant \delta^k$ for all $x \in E'_n$. The above is therefore at least $\mu(E'_n)(\delta^k - O_k(\varepsilon))$. Since $\tilde{\mu}$ is $\tilde{T}$-invariant we may make the change of variables $x' := \tilde{T}^{an}x$ and obtain

$$\int_{\tilde{X}} f \cdot U^{dn}f \cdots U^{(k-1)dn}f d\tilde{\mu} \geqslant \mu(E'_n)(\delta^k - O_k(\varepsilon)) \geqslant L^{-2}\delta'(\delta^k - O_k(\varepsilon)). \tag{3.2}$$

Averaging this over $n \in S_N$ and using positivity, we obtain

$$\mathbb{E}_{0 \leqslant n < N}\mathbb{E}_{0 \leqslant d \leqslant L-1} \int_{\tilde{X}} f \cdot U^{dn}f \cdots U^{(k-1)nd}f d\tilde{\mu} \geqslant \eta L^{-3}\delta'(\delta^k - O_k(\varepsilon)).$$

Here, the dependence of d on n in (3.2) has been suppressed by averaging over the whole range $0 \leqslant d \leqslant L-1$ and using non-negativity.

Since no $n' < LN$ has more than L representations as dn with $d < L$ and $n < N$ this implies that

$$\mathbb{E}_{0 \leqslant n' < NL} \int_{\tilde{X}} f \cdot U^{n'}f \cdots U^{(k-1)n'}f d\tilde{\mu} \geqslant \eta L^{-4}\delta'(\delta^k - O_k(\varepsilon)).$$

[1] We leave it as an exercise to the reader to confirm that this map *can* be chosen to be neasurable.

Rescaling by a further factor of L we have

$$\mathbb{E}_{0\leqslant n<N}\int_{\tilde{X}} f\cdot U^n f\cdots U^{(k-1)n} f\,d\tilde{\mu} \geqslant \eta L^{-4}\delta'(\delta^k - O_k(\varepsilon)).$$

Choosing $\varepsilon = \varepsilon(\delta, k)$ sufficiently small we see that the liminf of this as $N \to \infty$ is greater than zero, as required. □

4. References

Furstenberg, Katzelson and Ornstein avoid the use of van der Waerden's theorem, and I recommend trying to work through their argument for the skew torus. The use of van der Waerden's theorem seems to have first occured in the paper of Furstenberg and Katznelson on the multidimensional Szemerédi theorem, which appeared in *J. d'Analyse*, vol 35 (1978). The same material is discussed in Furstenberg's book. I found Tao's blog quite helpful in preparing these notes, though his point of view is somewhat different (involving the use of "Hilbert modules").