6. Proof of Szemerédi's theorem

In this section we sketch the proof of Szemerédi's theorem and formulate the main steps. Recall that we already showed how the combinatorial statement Theorem 1.9 is equivalent to the multiple recurrence theorem Theorem 3.3. As was the case with other multiple recurrence theorems, we actually prove a stronger statement involving averages.

Theorem 6.1. *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic system and let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then for every $k \in \mathbb{N}$,*

$$\liminf_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} \mu(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-kn}A) > 0. \tag{6.1}$$

It turns out that the liminf in (6.1) is an actual limit, but this was not proved until 2005 [13], almost 30 years after Theorem 6.1 was first established. Observe that in Theorem 6.1 we assume that the system is ergodic; but as was explained in the previous section this is just for convenience and one can still obtain Theorem 3.3.

6.1. Kronecker factor.

Let X be a compact metrizable abelian group, with the Borel σ-algebra and Haar measure and let $\alpha \in X$. Then the map $T : X \to X$ given by $Tx = x + \alpha$ is a measure preserving transformation. This transformation is ergodic precisely when α generates a dense subgroup of X. Systems of this form are called ***Kronecker systems***.

Theorem 6.2. *An ergodic system $(X, \mathcal{B}, \mu, T)$ is (isomorphic[4] to) a Kronecker system if and only if every $f \in L^2$ is compact (in the sense of Definition 5.21).*

Recall that a measure preserving system $\mathbf{Y} = (Y, \mathcal{B}, \nu, S)$ is a factor of another system $\mathbf{X} = (X, \mathcal{A}, \mu, T)$ if there exists a measurable map $\phi : X \to Y$ pushing μ to ν (i.e. satisfying $\mu(\phi^{-1}B) = \nu(B)$ for every $B \in \mathcal{B}$) and intertwining T and S, in the sense that $S \circ \phi = \phi \circ T$. In this situation one can embed $L^2(Y)$ into $L^2(X)$ by taking $f \in L^2(Y)$ to $f \circ \pi \in L^2(X)$; noting that this map is an isometric operator. We will then assume simply that $L^2(Y) \subset L^2(X)$. Let H_c denote the set of all compact functions in $L^2(X)$. If $\mathbf{Y}$ is a Kronecker system, then every function in $L^2(Y)$ is compact, and hence $L^2(Y) \subset H_c$. Conversely, if $\mathbf{Y}$ is a factor of $\mathbf{X}$ with $L^2(Y) \subset H_c$, then $\mathbf{Y}$ is a Kronecker system.

It turns out that every ergodic measure preserving system $\mathbf{X} = (X, \mathcal{A}, \mu, T)$ has a factor $\mathbf{Y} = (Y, \mathcal{B}, \nu, S)$ such that $L^2(Y) = H_c$; such factor is called the ***Kronecker factor*** of $\mathbf{X}$.

Theorem 6.3 (Kronecker factor)**.** *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be an ergodic system and let H_c denote the space of all compact functions in $L^2(X)$. Then there exists a factor map $\pi : \mathbf{X} \to \mathbf{Y}$ such that*

- *$\mathbf{Y}$ is a Kronecker system.*
- *$L^2(Y) = H_c$.*
- *$\mathbf{Y}$ is the maximal Kronecker factor of $\mathbf{X}$ in the sense that if $\mathbf{Z}$ is a Kronecker system and there is a factor map $\phi : \mathbf{X} \to \mathbf{Z}$, then there exists a factor map $\psi : \mathbf{Y} \to \mathbf{Z}$ such that $\phi = \psi \circ \pi$.*
- *$\mathbf{Y}$ is unique up to isomorphism, and can be described as $\mathbf{Y} = (X, \mathcal{D}, \mu, T)$, where $\mathcal{D} = \{D \in \mathcal{B} : 1_D \in H_c\}$.*

It is possible that the Kronecker factor of a system is trivial (i.e., isomorphic to the identity transformation). This occurs precisely when H_c consists only of constant functions, which in view of Theorem 5.24 is equivalent to the system being weak mixing.

Remark 6.4. *In view of Theorem 6.3, one can re-interpret the Jacobs-de Leeuw-Glicksberg in terms of conditional expectations. Indeed, given $f \in L^2(X)$ and writing $f = f_c + f_{wm}$ with $f_c \in H_c$ and $f_{wm} \in H_{wm}$, we know that f_c is the orthogonal projection of f onto the space H_c, which now we know is the same space as the space $L^2(X, \mathcal{D}, \mu)$ (where $\mathcal{D}$ is the σ-algebra described in the last item of Theorem 6.3) of L^2 functions that are measurable with respect to $\mathcal{D}$. It follows that $f_c = \mathbb{E}[f \mid \mathcal{D}]$. This is often denoted by $\mathbb{E}[f \mid \mathbf{Y}]$, and we can then re-formulate Theorem 5.24 as stating that for any ergodic system $\mathbf{X}$ with Kronecker factor $\mathbf{Y}$ and any $f \in L^2(X)$, the difference $f - \mathbb{E}[f \mid \mathbf{Y}]$ is a weak mixing function.*

[4] Two systems $\mathbf{X}$ and $\mathbf{Y}$ are isomorphic if there exists a bijective factor map $\pi : \mathbf{X} \to \mathbf{Y}$ whose inverse is also a factor map.

6.2. **Special cases of multiple recurrence.** We start by proving Theorem 6.1 when the system is weak mixing.

Theorem 6.5. *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be a weak mixing system, let $k \in \mathbb{N}$ and let $f_1, \ldots, f_k \in L^\infty(X)$. Then*

$$UC\text{-}\lim_n \prod_{i=1}^k T^{ni} f_i = \prod_{i=1}^k \int_X f_i \, \mathsf{d}\mu \qquad \text{in } L^2(X). \tag{6.2}$$

Replacing each f_i with 1_A, a direct corollary of Theorem 6.5 is that Theorem 6.1 holds whenever the system is weak mixing.

The main idea in the proof of Theorem 6.5 is to use the van der Corput trick.

Proof of Theorem 6.5. We proceed by induction on k. The case $k = 1$ follows from the mean ergodic theorem (Theorem 2.17). Assume now that $k > 1$ and the result has been established for $k - 1$. Splitting f_k as the sum of a constant and function with 0 integral, we reduce the proof to those two cases. If f_k is a constant, then (6.2) follows immediately by induction.

Assume next that $\int_X f_k \, \mathsf{d}\mu = 0$. Since the right hand side of (6.2) is 0, we will use the van der Corput trick. Let $u_n := \prod_{i=1}^k T^{ni} f_i$. We have

$$\langle u_{n+h}, u_n \rangle = \int_X \prod_{i=1}^k T^{(n+h)i} f_i \cdot \overline{T^{ni} f_i} \, \mathsf{d}\mu = \int_X \prod_{i=1}^k T^{ni} \big(T^{hi} f_i \cdot \overline{f_i} \big) \, \mathsf{d}\mu = \int_X T^h f_1 \cdot \overline{f_1} \prod_{i=2}^k T^{n(i-1)} \big(T^{hi} f_i \cdot \overline{f_i} \big) \, \mathsf{d}\mu,$$

where the last equality follows from the fact that T^n preserves the measure. After using induction hypothesis on the $k - 1$ functions $\big\{ T^h f_2 \cdot \overline{f_2}, \ldots, T^{h(k-1)} f_k \cdot \overline{f_k} \big\}$ and taking averages we get

$$\begin{aligned} UC\text{-}\lim_n \langle u_{n+h}, u_n \rangle &= UC\text{-}\lim_n \int_X T^h f_1 \cdot \overline{f_1} \prod_{i=2}^k T^{n(i-1)} \big(T^{hi} f_i \cdot \overline{f_i} \big) \, \mathsf{d}\mu \\ &= \prod_{i=1}^k \int_X T^{hi} f_i \cdot \overline{f_i} \, \mathsf{d}\mu. \end{aligned}$$

Finally, taking an average on h and using Theorem 5.14 and condition (4) from Theorem 5.11 we obtain

$$\left| C\text{-}\lim_h \, UC\text{-}\lim_n \langle u_{n+h}, u_n \rangle \right| = \left| C\text{-}\lim_h \prod_{i=1}^k \int_X T^{hi} f_i \cdot \overline{f_i} \, \mathsf{d}\mu \right| \le \prod_{i=1}^{k-1} \|f_i\|^2 \cdot C\text{-}\lim_h \big| \langle T^{hk} f_k, f_k \rangle \big| = 0$$

□

On the other end of the spectrum, we have Kronecker systems.

Theorem 6.6. *Let $\mathbf{X}$ be a Kronecker system, let $k \in \mathbb{N}$ and let $f \in L^\infty(X)$. Then for every $\epsilon > 0$, the set*

$$\left\{ n \in \mathbb{N} : \int_X \prod_{i=0}^k T^{ni} f \, \mathsf{d}\mu > \int_X f^{k+1} \, \mathsf{d}\mu - \epsilon \right\}$$

is syndetic.

To see why Theorem 6.6 implies that Theorem 6.1 holds for Kronecker systems, apply Theorem 6.6 to the indicator function 1_A of a set $A \in \mathcal{B}$ with $\mu(A) > 0$ and let S be the syndetic set of n for which $\mu(A \cap T^{-n} A \cap \cdots \cap T^{-kn} A) > \mu(A)/2$. Let $L \in \mathbb{N}$ be bound on the gaps of S (so that every interval of length L contains an element of S). Then for $N - M$ large enough we have $|[M, N] \cap S| \ge (N - M - L)/L > (N - M)/(2L)$ and hence

$$UC\text{-}\lim_n \mu(A \cap \cdots \cap T^{-kn} A) \ge \lim_{N - M \to \infty} \frac{1}{N - M} \sum_{n \in S \cap [M, N]} \mu(A \cap \cdots \cap T^{-kn} A) \ge \frac{\mu(A)}{4L} > 0.$$

Proof of Theorem 6.6. Let $\mathbf{X}$, k, f and $\epsilon > 0$ be as in the statement of the theorem. By re-scaling we may assume that $\|f\|_\infty \leq 1$. Since f is a compact function, it follows from Exercise 5.23 that the set $S := \{n \in \mathbb{N} : \|T^n f - f\| < \epsilon/k^2\}$ is syndetic. Observe that for each $n \in S$ and $i \in \{0, 1, \dots, k\}$ we have

$$\|T^{in} f - f\| \leq \|T^{in} f - T^{(i-1)n} f\| + \|T^{(i-1)n} f - T^{(i-2)n} f\| + \cdots + \|T^n f - f\| = i\|T^n f - f\| \leq \frac{\epsilon}{k}.$$

Using the Cauchy-Schwarz inequality repeatedly we conclude that for every $n \in S$

$$\begin{aligned} \int_X \prod_{i=0}^{k} T^{ni} f \, \mathsf{d}\mu &= \int_X f \cdot \prod_{i=1}^{k} T^{ni} f \, \mathsf{d}\mu \geq \int_X f^2 \cdot \prod_{i=2}^{k} T^{ni} f \, \mathsf{d}\mu - \frac{\epsilon}{k} \\ &\geq \int_X f^3 \cdot \prod_{i=3}^{k} T^{ni} f \, \mathsf{d}\mu - \frac{2\epsilon}{k} \geq \cdots \geq \int_X f^{k+1} \, \mathsf{d}\mu - \epsilon. \end{aligned}$$

□

We have now shown that either a system $\mathbf{X}$ is weak mixing, and hence Theorem 6.1 holds, or it is not weak mixing, and hence it has a non-trivial Kronecker factor, where Theorem 6.1 holds. In either case, we have proved that any ergodic system has a non-trivial factor where Theorem 6.1 holds. The basic idea of the proof of Theorem 6.1 for general systems is to keep finding larger factors where the conclusion holds, ultimately covering all of $\mathbf{X}$. To make the necessary definitions we will take advantage of a useful theorem of Rokhlin.

6.3. **Rokhlin's skew-product lemma.** Two probability spaces $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{D}, \nu)$ are isomorphic if the (trivial) measure preserving systems $(X, \mathcal{B}, \mu, Id)$ and $(Y, \mathcal{D}, \nu, Id)$ are isomorphic. It is a known result that if μ is a Borel measure on a compact metric space with no point masses, then $(X, \mathcal{B}, \mu)$ is isomorphic to $[0, 1]$ with the Borel σ-algebra and the Lebesgue measure. In particular any two such probability spaces are isomorphic! It follows in particular that, when understood as probability spaces with the (appropriate) Lebesgue measure, $[0, 1]$ is isomorphic to $[0, 1]^2$. The following lemma improves upon these ideas and provides a useful way to understand factors.

Lemma 6.7. *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system and let $\mathbf{Y} = (Y, \mathcal{C}, \nu, S)$ be a factor, with factor map $\pi : X \to Y$. Then there exists a probability space $(Z, \mathcal{D}, \lambda)$ and a measurable map $\rho : Y \to \mathrm{Aut}(Z)$ (called a co-cycle) taking values in the set of measure preserving transformations of $(Z, \mathcal{D}, \lambda)$ such that $\mathbf{X}$ is isomorphic to $(Y \times Z, \mathcal{C} \otimes \mathcal{D}, \nu \otimes \lambda, R)$, where $R(y, z) = \big(Sy, \rho(y)(z)\big)$ (in other words, $\mathbf{X}$ is a skew-product over $\mathbf{Y}$).*

Example 6.8. *Take $X = [0, 1]^2$ with the (Borel) Lebesgue measure and let $T : (y, x) \mapsto (y + \alpha, x + y)$ for some fixed irrational α. Let $Y = [0, 1]$, also endowed with the (Borel) Lebesgue measure and let $S : y \mapsto y + \alpha$. Then the projection $\pi : X \to Y$ onto the first coordinate is a factor map of measure preserving systems.*

In this case we can take $Z = [0, 1]$ and $\rho(y)$ to be the rotation on Z by y (in other words $\rho(y) : z \mapsto z + y \bmod 1$).

6.4. **Relative weak mixing and compactness.** Recall that a m.p.s. $\mathbf{X}$ is weak mixing if and only if the product $\mathbf{X} \times \mathbf{X}$ is ergodic. The following definition extends this concept to a relative notion.

Definition 6.9. *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ and $\mathbf{Y} = (Y, \mathcal{C}, \nu, S)$ be ergodic systems and let $\pi : \mathbf{X} \to \mathbf{Y}$ be a factor map. Let $(Z, \mathcal{D}, \lambda)$ and ρ be given by Lemma 6.7. The* **relative product** *of $\mathbf{X}$ with itself over $\mathbf{Y}$ is the system*

$$\mathbf{X} \times_{\mathbf{Y}} \mathbf{X} := (Y \times Z \times Z, \mathcal{C} \otimes \mathcal{D} \otimes \mathcal{D}, \nu \otimes \lambda \otimes \lambda, T \times_{\mathbf{Y}} T),$$

where $T \times_{\mathbf{Y}} T : (y, z_1, z_2) \mapsto (Sy, \rho(y)z_1, \rho(y)z_2)$.

We say that $\mathbf{X}$ is **weak mixing relative to** *(or a* **weak mixing extension** *of) $\mathbf{Y}$ is the relative product $\mathbf{X} \times_{\mathbf{Y}} \mathbf{X}$ is ergodic.*

Exercise 6.10. *Show that a system is weak mixing if and only if it is relative weak mixing with respect to the trivial factor (i.e. the factor to the one-point system).*

Next recall that a system is a Kronecker system if and only if every L^2 function is compact, i.e., for any $f \in L^2$ the orbit $\{T^n f : n \in \mathbb{N}\}$ is pre-compact. For a subset of a Hilbert space, pre-compact is equivalent to totally bounded, so f is compact if and only if for any $\epsilon > 0$ there are finitely many functions $g_1, \dots, g_r \in L^2$ such that $\{T^n f : n \in \mathbb{N}\} \subset \bigcup_{i=1}^r B(g_i, \epsilon)$. This inclusion can be written as

$$\forall n \in \mathbb{N} \quad \min_{1 \le i \le r} \|T^n f - g_i\|_{L^2(\mu)} < \epsilon.$$

We can now relativize the notion of a compact (or Kronecker) factor.

Definition 6.11. *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ and $\mathbf{Y} = (Y, \mathcal{C}, \nu, S)$ be ergodic systems and let $\pi : \mathbf{X} \to \mathbf{Y}$ be a factor map. Let $(Z, \mathcal{D}, \lambda)$ be given by Lemma 6.7 and consider the measures $\mu_y = \delta_y \otimes \lambda$ for each $y \in Y$.*

A function $f \in L^2(X)$ is **compact relative to Y** *if for every $\epsilon > 0$ there are finitely many functions $g_1, \dots, g_r \in L^2(X)$ such that*

$$\forall n \in \mathbb{N} \quad \min_{1 \le i \le r} \|T^n f - g_i\|_{L^2(\mu_y)} < \epsilon \quad \text{for } \nu\text{-a.e. } y.$$

The system $\mathbf{X}$ is **compact relative to** *(or a* **compact extension of***)* $\mathbf{Y}$ *if there is a dense set of relatively compact functions in $L^2(X)$.*

Exercise 6.12. *Show that an ergodic system is a Kronecker system if and only if it is compact relative to the trivial (one point) factor.*

Exercise 6.13. *Show that the system $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ given by $X = [0,1]^2$, $\mathcal{B} = \text{Borel}$, $\mu = \text{Lebesgue}$ and $T : (x, y) \mapsto (x + \alpha, y + x)$, where α is irrational, is a compact extension of the rotation by α (i.e. the system $\mathbf{Y} = (Y, \mathcal{D}, \nu, S)$ where $Y = [0,1]$, $\mathcal{D} = \text{Borel}$, $\nu = \text{Lebesgue}$ and $S : x \mapsto x + \alpha$).*

Note that in Definition 6.11 we do not require that every function $f \in L^2(X)$ be compact relative to $\mathbf{Y}$ but only that a dense subset of L^2 has this property. The next exercise helps explain why.

The notation $\lfloor x \rfloor$ for a real number x, denotes the largest integer n such that $n \le x$.

Exercise 6.14. *Let $\mathbf{X}$ be as in the previous exercise. Show that the function $f(x, y) = e(y\lfloor 1/x \rfloor)$ is not conditionally compact with respect to $\mathbf{Y}$.*

It is often helpful to think of "compact" as a generalization of "finite". The next exercise explains in which sense the notion of "relatively compact" generalizes the notion of "relatively finite".

Exercise 6.15. *Let $\mathbf{X}$ and $\mathbf{Y}$ be ergodic systems and let $\pi : \mathbf{X} \to \mathbf{Y}$ be a factor map. Let $(Z, \mathcal{D}, \lambda)$ be given by Lemma 6.7 and suppose that Z is finite. Show that $\mathbf{X}$ is a compact extension of $\mathbf{Y}$.*

6.5. Sketch of the proof.

Definition 6.16 (Sz systems)**.** *An ergodic system $(X, \mathcal{B}, \mu, T)$ is called* **Sz** *(for Szemerédi) if it satisfies the conclusion of the Theorem 6.1*

We already saw that every ergodic system has a non-trivial factor that is ***Sz*** (either the whole system if it is weak mixing, or its non-trivial Kronecker factor otherwise). The idea of the proof of Theorem 6.1 is that any proper factor which is Sz is contained in a strictly larger factor which is also Sz. There are 3 main components. The first states that the Sz property can be lifted by weak mixing extensions.

Theorem 6.17. *If an ergodic system $\mathbf{X}$ is a weak mixing extension of a system $\mathbf{Y}$ and $\mathbf{Y}$ is Sz, then so is $\mathbf{X}$.*

The proof of Theorem 6.17 is very similar to the proof of Theorem 6.5 which dealt with the "absolute" case. In particular it combines a similar induction with the van der Corput trick.

Similarly, the next results states that the Sz property can be lifted by compact extensions.

Theorem 6.18. *If an ergodic system $\mathbf{X}$ is a compact extension of a system $\mathbf{Y}$ and $\mathbf{Y}$ is Sz, then so is $\mathbf{X}$.*

The proof of Theorem 6.18 draws on the ideas from the proof of Theorem 6.6 which dealt with the "absolute" case, but is more complicated and requires additional insights. One useful (but not entirely necessary) tool is the van der Waerden theorem, in the finitistic form described in Exercise 1.5.

The third major step is a relative version of the Jacobs-de Leeuw-Glicksberg decomposition. Recall that a consequence of this decomposition is that whenever a system $\mathbf{X}$ is not weak mixing, it has a non-trivial factor $\mathbf{Y}$ which is a Kronecker system.

Theorem 6.19. *If a non-trivial extension $\pi : \mathbf{X} \to \mathbf{Y}$ of ergodic systems is not relatively weak mixing, then there exists an intermediate relatively compact extension, i.e., there exists a system $\mathbf{Z}$ and factor maps $\pi_1 : \mathbf{X} \to \mathbf{Z}$ and $\pi_2 : \mathbf{Z} \to \mathbf{Y}$ such that $\pi = \pi_2 \circ \pi_1$, π_2 is non-trivial and $\mathbf{Z}$ is a compact extension of $\mathbf{Y}$.*

We are now ready to finish the proof of Theorem 6.1. The idea is to consider a maximal factor of $\mathbf{X}$ which is a Sz system. Here maximal means with respect to the natural partial order on all factors of $\mathbf{X}$ given by $\mathbf{Y} \prec \mathbf{Z}$ if $\mathbf{Y}$ is a factor of $\mathbf{Z}$.[5] An equivalent description of this partial order is obtained by corresponding each factor to a subset of $L^2(X)$ (as explained after Theorem 6.2); then $\mathbf{Y} \prec \mathbf{Z}$ if and only if $L^2(Y) \subset L^2(Z)$.

To consider a maximal Sz factor of $\mathbf{X}$ one can use Zorn's lemma (or, alternatively, a transfinite induction). To be able to apply this lemma, one needs to show that the Sz property is "closed", in the following sense.

Lemma 6.20. *Let $\mathbf{X}$ be an ergodic system and suppose that there is a totally ordered family of factors $\mathbf{Y}_\alpha$ such that $\bigcup_\alpha L^2(\mathbf{Y}_\alpha)$ is dense in $L^2(\mathbf{X})$. If every $\mathbf{Y}_\alpha$ is Sz, then so is $\mathbf{X}$.*

Lemma 6.20 implies that Zorn's lemma can be applied and hence that $\mathbf{X}$ has a maximal Sz factor, say $\mathbf{Y}$. If $\mathbf{X}$ is a non-trivial extension of $\mathbf{Y}$ there are two cases. In the first case, $\mathbf{X}$ is a weak mixing extension of $\mathbf{Y}$ and hence by Theorem 6.17 $\mathbf{X}$ is Sz. In the second case, $\mathbf{X}$ is not a weak mixing extension of $\mathbf{Y}$, and hence by Theorem 6.19 there is a non-trivial extension $\mathbf{Z}$ of $\mathbf{Y}$ which is a compact extension of $\mathbf{Y}$ and a factor of $\mathbf{X}$. By Theorem 6.18, $\mathbf{Z}$ is Sz, but this contradicts the fact that $\mathbf{Y}$ was the maximal factor of $\mathbf{X}$ that was Sz.

Therefore, $\mathbf{X}$ must be Sz itself and this finishes the proof.

References

[1] V. Bergelson. Ergodic Ramsey theory–an update. In *Ergodic theory of $\mathbb{Z}^d$ actions*, volume 228 of *London Math. Soc. Lecture Note Ser.*, pages 1–61. Cambridge Univ. Press, Cambridge, 1996.

[2] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*, 9(3):725–753, 1996.

[3] V. Bergelson and J. Moreira. Van der Corput's difference theorem: some modern developments. *Indag. Math. (N.S.)*, 27(2):437–479, 2016.

[4] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinaĭ. *Ergodic theory*, volume 245 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1982. Translated from the Russian by A. B. Sosinskiĭ.

[5] J. G. van der Corput. Diophantische Ungleichungen. I. Zur Gleichverteilung Modulo Eins. *Acta Math.*, 56(1):373–456, 1931.

[6] Manfred Einsiedler and Thomas Ward. *Ergodic theory with a view towards number theory*, volume 259 of *Graduate Texts in Mathematics*. Springer-Verlag London, Ltd., London, 2011.

[7] Tanja Eisner, Bálint Farkas, Markus Haase, and Rainer Nagel. *Operator theoretic aspects of ergodic theory*, volume 272 of *Graduate Texts in Mathematics*. Springer, Cham, 2015.

[8] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.*, 31:204–256, 1977.

[9] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory.* Princeton University Press, Princeton, N.J., 1981.

[10] E. Glasner. *Ergodic theory via joinings*, volume 101 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2003.

[11] T. Gowers. A new proof of Szemerédi's theorem. *GAFA*, 11:465–588, 2001.

[12] R. L. Graham, B. L. Rothschild, and J. H. Spencer. *Ramsey theory.* John Wiley & Sons, Inc., New York, second edition, 1990.

[13] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math. (2)*, 161(1):397–488, 2005.

[14] B. Host and B. Kra. *Nilpotent structures in ergodic theory*, volume 236 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2018.

[15] T. Kamae and M. Mendès France. Van der Corput's difference theorem. *Israel J. Math.*, 31(3-4):335–342, 1978.

[16] A. Y. Khinchin. *Three pearls of number theory.* Graylock Press, Rochester, N. Y., 1952.

[17] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences.* John Wiley & Sons, New York-London-Sydney, 1974.

[18] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:245–252, 1953.

[19] A. Sárközy. On difference sets of sequences of integers. I. *Acta Math. Acad. Sci. Hungar.*, 31(1–2):125–149, 1978.

[20] I. Schur. Über die kongruenz $x^m + y^m \equiv z^m (\text{mod } p)$. *Jahresbericht der Deutschen Math. Verein.*, 25:114–117, 1916.

[21] E. Szemerédi. On the sets of integers containing no k elements in arithmetic progressions. *Acta Arith.*, 27:299–345, 1975.

[22] B. L. van der Waerden. Beweis einer baudetschen vermutung. *Nieuw. Arch. Wisk.*, 15:212–216, 1927.

[23] P. Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1982.

[24] H. Weyl. Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.*, 77(3):313–352, 1916.

[5] To more precise we also need the factor maps between $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ to be compatible.