

## 10.5 An ergodic argument

In 1977, Furstenberg [121] gave a spectacular new proof of Szemerédi's theorem (and hence Roth's theorem), using the methods of ergodic theory rather than Fourier analysis or combinatorics. The argument relies on very little arithmetic structure, being based almost entirely on an analysis of the mixing properties of the shift operator  $TA := A + 1$  on a set  $A$  of integers. As such it is very flexible and has led to several wide-ranging generalizations of Szemerédi's theorem, some of which we will discuss in the next chapter.

The initial ergodic arguments of Furstenberg were infinitary in nature, working with the integers  $\mathbf{Z}$ , and in fact embedding these integers in an abstract measure-preserving system  $(X, \mathcal{B}, T, \mu)$ . In several versions of the argument, the axiom of choice (in the guise of Zorn's lemma) was used to obtain a suitable structural decomposition of this measure-preserving system. More recently, however, there has been progress in establishing *finitary* versions of this argument, in which one works in a concrete and finite measure-preserving system, such as the cyclic group  $\mathbf{Z}_N$  with the standard shift  $TA := A + 1$ . These finitary arguments, which were inspired by the Szemerédi regularity lemma, to be introduced in the next section, are somewhat messier than the elegant infinitary arguments, but lead to explicit (albeit poor) quantitative bounds for  $r_k(\mathbf{Z}_N)$ . Also these finitary ergodic arguments played an essential role in the proof of the Green–Tao theorem concerning progressions in the primes.

In this section we give a finitary ergodic proof of Roth's theorem, using a formulation from [358]. The proof is not fully ergodic because we shall exploit the Fourier transform, but in the next chapter we will discuss how one can remove this dependence on the Fourier transform (and thus extend the argument to higher  $k$ ). The precise result we shall establish is

**Theorem 10.33** *For all finite groups  $Z$ , and all  $f : Z \rightarrow \mathbf{R}^+$  with  $0 \leq f(x) \leq 1$  and  $\mathbf{E}_Z(f) \geq \delta$ , we have  $\Lambda_3(f, f, f) = \Theta_\delta(1)$ .*

This is of course weaker than what one can obtain by purely Fourier-analytic methods such as Theorem 10.31, but the proof is somewhat different and is easier to extend to higher  $k$ . In particular, it replaces the density increment argument of previous sections by an *energy increment argument*. Whereas in the previous arguments one constructed a series of objects (progressions or Bohr sets) on which  $f$  had increasingly large density, here we construct a series of  $\sigma$ -algebras or *partitions* with respect to which  $f$  has increasingly large *energy*. This eventually leads to constructing a “low-complexity” approximation  $f_{U^\perp}$  to  $f$ , where the error  $f_U := f - f_{U^\perp}$  is linearly uniform and thus has negligible impact on  $\Lambda_3(f, f, f)$ .

The low-complexity approximation  $f_{U^\perp}$  turns out to be *almost periodic*, which will lead to a lower bound on  $\Lambda_3(f_{U^\perp}, f_{U^\perp}, f_{U^\perp})$ .

We turn to the details, beginning with the definition of almost periodicity. For convenience we shall take advantage of the Fourier transform to define this notion, though it is not essential (see exercises).

**Definition 10.34 (Almost periodicity)** Let  $K \geq 1$  be an integer and  $\sigma > 0$ . We say that a function  $f : Z \rightarrow \mathbf{C}$  is  $K$ -*quasiperiodic* if there exist frequencies  $\xi_1, \dots, \xi_K \in Z$  (possibly repeated) and complex numbers  $c_1, \dots, c_k$  with  $|c_1|, \dots, |c_k| \leq 1$  such that  $f = \sum_{j=1}^k c_j e_{\xi_j}$ , or in other words

$$f(x) = \sum_{j=1}^k c_j e(x \cdot \xi_j).$$

We say that a function  $f : Z \rightarrow \mathbf{C}$  is  $(K, \sigma)$ -*almost periodic* if there exists a  $K$ -quasiperiodic function  $g$  such that  $\|f - g\|_{L^2(Z)} \leq \sigma$ .

A key observation is that Theorem 10.33 is easy to prove for almost periodic functions, if  $K$  is not too large and  $\sigma$  is sufficiently small. More precisely, we have

**Proposition 10.35 (Almost periodic functions are recurrent)** Let  $f : Z \rightarrow \mathbf{R}^+$  be such that  $0 \leq f(x) \leq 1$  and  $\mathbf{E}_Z(f) \geq \delta$ . If  $f$  is  $(K, \sigma)$ -almost periodic for some  $K \geq 1$  and  $0 < \sigma < \delta^3/8$ , then

$$\Lambda_3(f, f, f) = \Omega((\delta/K)^K \delta^3).$$

This proposition should be compared with Lemma 4.44. A key point here is that the smallness condition on  $\sigma$  does not involve  $K$ . This will be important for us as  $K$  will eventually be quite large compared with  $\sigma$ .

*Proof* By definition, we can find frequencies  $\xi_1, \dots, \xi_K$  and coefficients  $c_1, \dots, c_K$  of magnitude  $O(1)$  such that

$$f(x) = \sum_{j=1}^K c_j e(x \cdot \xi_j) + g(x)$$

for all  $x \in Z$ , where  $g$  has an  $L^2(Z)$  norm of at most  $\sigma$ . Now let  $S := \{\xi_1, \dots, \xi_K\}$  and let  $\rho > 0$  be a radius to be chosen later. If  $h$  lies in the Bohr set  $\text{Bohr}_Z(S, \rho)$ , then  $e(h \cdot \xi_j) = 1 + O(\rho)$ , and hence

$$T^{jh} f(x) = f(x) + O(K\rho) + T^{jh} g(x)$$

for  $j = 1, 2$ , where  $T^h f(x) = f(x + h)$  denotes the shift by  $h$ . In particular we have

$$\|T^{jh} f - f\|_{L^2(Z)} \leq O(K\rho) + 2\sigma,$$

while from the boundedness of  $f$  we have  $\|T^{jh} f\|_{L^\infty(Z)} \leq 1$ . After a few applications of the triangle inequality and Hölder's inequality, we then conclude that

$$\|f(T^h f)(T^{2h} f) - f^3\|_{L^1(Z)} \leq O(K\rho) + 4\sigma$$

and hence by the triangle inequality again

$$\mathbf{E}_{x \in Z} f(x) T^h f(x) T^{2h} f(x) \geq \mathbf{E}_{x \in Z} (f(x)^3) - O(K\rho) - 4\sigma.$$

On the other hand, from Hölder's inequality we have

$$\mathbf{E}_{x \in Z} f(x)^3 \geq \mathbf{E}_{x \in Z} (f(x))^3 = \delta^3$$

so by hypothesis on  $\sigma$

$$\mathbf{E}_{x \in Z} f(x) T^h f(x) T^{2h} f(x) \geq \frac{1}{2} \delta^3 - O(K\rho).$$

Applying (4.25) and the positivity of  $f$ , we conclude that

$$\Lambda_3(f, f, f) = \mathbf{E}_{x, h \in Z} f(x) T^h f(x) T^{2h} f(x) \geq \rho^K \max\left(\frac{1}{2} \delta^3 - O(K\rho), 0\right).$$

The claim then follows by taking  $\rho$  to be a sufficiently small multiple of  $\delta/K$ .  $\square$

To establish Theorem 10.33 in the general case, one now needs to approximate an arbitrary function  $f$  by an almost periodic one. Indeed we will establish the following fundamental proposition:

**Proposition 10.36 (Koopman–von Neumann decomposition)** *Let  $f : Z \rightarrow \mathbf{R}^+$  be such that  $0 \leq f(x) \leq 1$ , let  $\sigma > 0$ , and let  $F : \mathbf{R}^+ \times \mathbf{R}^+ \rightarrow \mathbf{R}^+$  be an arbitrary function. Then there exists a quantity  $K = O_{\sigma, F}(1)$  and a decomposition  $f = f_{U^\perp} + f_U$  with the following properties:*

- the “anti-uniform” component  $f_{U^\perp}$  obeys the bounds  $0 \leq f_{U^\perp} \leq 1$  and  $\mathbf{E}_Z f_{U^\perp} = \mathbf{E}_Z f$ , and is  $(K, \sigma)$ -almost periodic;
- the “uniform” component  $f_U$  obeys the Fourier uniformity estimate  $\|f_U\|_{u^2(Z)} \leq \frac{1}{F(\sigma, K)}$ .

A remarkable feature of this proposition is that one can make the uniformity control on  $f_U$  arbitrarily strong by making  $F$  grow arbitrarily quickly. The price one pays for this is that the upper bound on  $K$  then deteriorates substantially.

We shall prove Proposition 10.36 in the rest of this section. For now, let us see how the proposition implies Theorem 10.33. We apply the proposition with  $\sigma := \delta^3/8$  and  $F$  to be chosen later. From Proposition 10.35 we have

$$\Lambda_3(f_{U^\perp}, f_{U^\perp}, f_{U^\perp}) = \Omega((\delta/K)^K \delta^3).$$

Since  $f, f_{U^\perp}$  are bounded between 0 and 1,  $f_U$  is bounded in magnitude by 1. Applying the Fourier uniformity estimate and Proposition 10.11, we conclude that

$$\Lambda_3(f, f, f) - \Lambda_3(f_{U^\perp}, f_{U^\perp}, f_{U^\perp}) = O\left(\frac{1}{F(\sigma, K)}\right).$$

Thus if we choose  $F$  to be sufficiently quickly growing, we can absorb the error term into the main term and conclude that

$$\Lambda_3(f, f, f) = \Omega((\delta/K)^K \delta^3).$$

Since  $K = O_{\sigma, F}(1) = O_\delta(1)$ , the claim follows.

It remains to prove Proposition 10.36. One can prove this proposition by a direct application of the Fourier transform (this is essentially the approach in [34]); however we shall use a more ergodic approach which extends more easily to progressions of longer length. A crucial tool here is that of *conditional expectation*.

**Definition 10.37 (Conditional expectation)** Define a  $\sigma$ -algebra of  $Z$  to be any collection  $\mathcal{B}$  of subsets of  $Z$  which contains  $\emptyset$  and  $Z$ , and is closed under unions, intersections, and complements. (The  $\sigma$ -algebras are in one-to-one correspondence with partitions of  $Z$ , and can be viewed as such.) If  $\mathcal{B}, \mathcal{B}'$  are two  $\sigma$ -algebras, we define  $\mathcal{B} \vee \mathcal{B}'$  to be the smallest  $\sigma$ -algebra which contains both. We say that a function  $f : Z \rightarrow \mathbf{C}$  is *measurable* with respect to  $\mathcal{B}$  if it is constant on every atom of  $\mathcal{B}$ , where an *atom* is any minimal non-empty element of  $\mathcal{B}$ . Given any  $f : Z \rightarrow \mathbf{C}$ , we define the *conditional expectation*  $\mathbf{E}(f|\mathcal{B}) : Z \rightarrow \mathbf{C}$  to be the function

$$\mathbf{E}(f|\mathcal{B})(x) := \mathbf{E}_{\mathcal{B}(x)} f = \frac{1}{|\mathcal{B}(x)|} \sum_{y \in \mathcal{B}(x)} f(y)$$

where  $\mathcal{B}(x)$  is the unique atom of  $\mathcal{B}$  which contains  $x$ ; equivalently,  $\mathbf{E}(f|\mathcal{B})$  is the orthogonal projection in  $L^2(Z)$  to the space of  $\mathcal{B}$ -measurable functions.

It turns out that certain  $\sigma$ -algebras  $\mathcal{B}$  are “compact” in the sense that conditional expectations such as  $\mathbf{E}(f|\mathcal{B})$  are automatically almost periodic. One precise formulation of this is

**Proposition 10.38 (Characters generate compact  $\sigma$ -algebras)** *Let  $\xi \in Z$  and  $0 < \varepsilon < 1$ . Then there exists a  $\sigma$ -algebra  $\mathcal{B}_{\varepsilon, \xi}$  with  $O_\varepsilon(1)$  atoms which approximately contains the character  $e_\xi(x) := e(\xi \cdot x)$  in the sense that*

$$\|e_\xi - \mathbf{E}(e_\xi|\mathcal{B}_{\varepsilon, \xi})\|_{L^\infty(Z)} = O(\varepsilon), \tag{10.20}$$

*and also has the property that every  $\mathcal{B}_{\varepsilon, \xi}$ -measurable function  $f$  with  $\|f\|_{L^\infty(Z)} \leq 1$  is  $(O_{\varepsilon, \sigma}(1), O(\sigma))$ -almost periodic for every  $\sigma > 0$ .*

*Proof* We use the first moment method. Let  $\alpha$  be a randomly selected element of the unit square  $\mathcal{Q} := \{z \in \mathbf{C} : 0 \leq \operatorname{Re}(z), \operatorname{Im}(z) < 1\}$ , and let  $\mathcal{B}_{\varepsilon, \xi}$  be the  $\sigma$ -algebra generated by the sets

$$A_{a,b,\varepsilon,\alpha} := \{x \in \mathbf{Z} : e_{\xi}(x) \in \varepsilon(\mathcal{Q} + a + bi + \alpha)\}; a, b \in \mathbf{Z}.$$

These sets, which partition  $\mathbf{Z}$ , are essentially translates of the Bohr set  $\operatorname{Bohr}_{\mathbf{Z}}(\{\xi\}, \varepsilon)$ ; at most  $O(1/\varepsilon)$  of them are non-empty. Since  $e_{\xi}$  fluctuates by at most  $O(\varepsilon)$  on each such set, we obtain the property (10.20). Now we prove the latter property. Observe that  $f$  is a linear combination of at most  $O(1/\varepsilon)$  indicator functions  $1_{A_{a,b,\varepsilon,\alpha}}$ , with bounded coefficients, so it suffices to prove the claim for the  $O(1/\varepsilon)$  non-trivial indicator functions  $1_{A_{a,b,\varepsilon,\alpha}}$ . The claim is trivial for  $\sigma \geq 1$ , and by approximating  $\sigma$  by the nearest power of 2 we thus see that it suffices to verify the claim for  $\sigma = 2^{-n}$  for integer  $n \geq 0$ . By the Borel–Cantelli lemma it will thus suffice to show that

$$\mathbf{P}(1_{A_{a,b,\varepsilon,\alpha}} \text{ is } (O_{\varepsilon,n}(1), O(2^{-n}))\text{-almost periodic}) = 1 - O(\varepsilon 2^{-n})$$

for each  $n \geq 1$  and  $a, b \in \mathbf{Z}$ .

Fix  $n, a, b$ . We rewrite

$$1_{A_{a,b,\varepsilon,\alpha}}(x) = 1_{\mathcal{Q}}\left(\frac{e(x \cdot \xi)}{\varepsilon} - a - bi - \alpha\right).$$

Let  $B$  be the  $\varepsilon 2^{-3n}$ -neighborhood of the boundary of the square  $\mathcal{Q}$ . From Urysohn's lemma followed by the Weierstrass approximation theorem, we can write

$$1_{\mathcal{Q}}(z) = P_{n,\varepsilon}(z) + O(1_B(z)) + O(2^{-n}),$$

where  $P_{n,\varepsilon}(z)$  is a polynomial of  $z$  and  $\bar{z}$  depending only on  $n$ . We conclude that

$$\begin{aligned} 1_{A_{a,b,\varepsilon,\alpha}}(x) & \qquad \qquad \qquad (10.21) \\ &= P_{n,\varepsilon}\left(\frac{e(x \cdot \xi)}{\varepsilon} - a - bi - \alpha\right) + O(\mathbf{I}(e(x \cdot \xi))/\varepsilon \in a + bi + \alpha \in B) + O(2^{-n}). \end{aligned}$$

The first term on the right-hand side can be easily verified to be  $O_{n,\varepsilon}(1)$ -quasiperiodic. An application of the first moment method easily shows that

$$\mathbf{E}(\|\mathbf{I}(e(x \cdot \xi))/\varepsilon \in a + bi + \alpha \in B\|_{L^2(\mathbf{Z})}^2) = O(\varepsilon 2^{-3n}),$$

so by Markov's inequality we see that the second term in (10.22) has an  $L^2(\mathbf{Z})$  norm of  $O(2^{-n})$  with probability  $1 - O(\varepsilon 2^{-n})$ . We thus see that  $1_{A_{a,b,\varepsilon,\alpha}}$  is  $(O_{n,\varepsilon}(1), O(2^{-n}))$ -quasiperiodic with probability  $1 - O(\varepsilon 2^{-n})$ , as desired.  $\square$

One can extend this to the  $\sigma$ -algebra generated by multiple characters:

**Corollary 10.39** *Let  $\xi_1, \dots, \xi_n \in Z$  and  $\varepsilon_1, \dots, \varepsilon_n > 0$ . Let  $\mathcal{B} := \mathcal{B}_{\varepsilon_1, \xi_1} \vee \dots \vee \mathcal{B}_{\varepsilon_n, \xi_n}$ , where  $\mathcal{B}_{\varepsilon, \xi}$  was defined in the previous proposition. Then every  $\mathcal{B}$ -measurable function  $f$  with  $\|f\|_{L^\infty(Z)} \leq 1$  is  $(O_{\varepsilon_1, \dots, \varepsilon_n, n, \sigma}(1), O_n(\sigma))$ -almost periodic for every  $\sigma > 0$ .*

*Proof* Observe that  $\mathcal{B}$  has at most  $O_{n, \varepsilon_1, \dots, \varepsilon_n}(1)$  atoms and so it suffices to verify the claim for an indicator  $f = 1_A$ , where  $A$  is an atom of  $\mathcal{B}$ . But  $1_A$  is then a product of  $n$  indicators  $1_{A_1} \dots 1_{A_n}$ , where  $A_j$  is an atom of  $\mathcal{B}_{\varepsilon, \xi_j}$ , and the claim then follows from the previous proposition and the observation that the product of bounded almost periodic functions remains bounded and almost periodic (but with slightly worse constants).  $\square$

The heart of the proof of Proposition 10.36 now lies in the following key lemma. We define the *energy*  $\mathcal{E}_f(\mathcal{B})$  of  $\mathcal{B}$  with respect to  $f$  to be the quantity

$$\mathcal{E}_f(\mathcal{B}) := \|\mathbf{E}(f|\mathcal{B})\|_{L^2(Z)}^2 = \mathbf{E}_{x \in Z} |\mathbf{E}(f|\mathcal{B})(x)|^2.$$

**Lemma 10.40 (Lack of uniformity implies energy increment)** *Let  $\varepsilon, \mu > 0$  be such that  $\varepsilon \leq \mu/4$ , and let  $f : Z \rightarrow \mathbf{R}^+$  be such that  $0 \leq f(x) \leq 1$ , and let  $\mathcal{B}$  be a  $\sigma$ -algebra such that*

$$\|f - \mathbf{E}(f|\mathcal{B})\|_{u^2(Z)} \geq \mu.$$

*Then there exists a frequency  $\xi \in Z$  such that we have the energy increment property*

$$\mathcal{E}_f(\mathcal{B} \vee \mathcal{B}_{\varepsilon, \xi}) \geq \mathcal{E}_f(\mathcal{B}) + \mu^2/4.$$

*Proof* By definition of  $u^2(Z)$ , we can find  $\xi \in Z$  such that

$$|\langle f - \mathbf{E}(f|\mathcal{B}), e_\xi \rangle_{L^2(Z)}| \geq \mu.$$

On the other hand, from (10.20) we have we see that  $e_\xi$  fluctuates by at most  $2\varepsilon$  on each atom of  $\mathcal{B}_{\varepsilon, \xi}$ , and hence on each atom of  $\mathcal{B} \vee \mathcal{B}_{\varepsilon, \xi}$ . Thus

$$\|e_\xi - \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\varepsilon, \xi})\|_{L^\infty(Z)} \leq 2\varepsilon;$$

since  $f - \mathbf{E}(f|\mathcal{B})$  is bounded in magnitude by 1, we conclude

$$|\langle f - \mathbf{E}(f|\mathcal{B}), e_\xi - \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\varepsilon, \xi}) \rangle_{L^2(Z)}| \leq 2\varepsilon.$$

Since  $\varepsilon < \mu/4$ , we deduce

$$|\langle f - \mathbf{E}(f|\mathcal{B}), \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\varepsilon, \xi}) \rangle_{L^2(Z)}| \geq \mu/2.$$

From the easily verified identity

$$\langle f - \mathbf{E}(f|\mathcal{B} \vee \mathcal{B}_{\varepsilon, \xi}), \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\varepsilon, \xi}) \rangle_{L^2(Z)} = 0$$

we thus have

$$|(\mathbf{E}(f|\mathcal{B} \vee \mathcal{B}_{\varepsilon, \xi}) - \mathbf{E}(f|\mathcal{B}), \mathbf{E}(e_\xi|\mathcal{B} \vee \mathcal{B}_{\varepsilon, \xi}))_{L^2(Z)}| \geq \mu/2$$

and hence by Cauchy–Schwarz

$$\|\mathbf{E}(f|\mathcal{B} \vee \mathcal{B}_{\varepsilon, \xi}) - \mathbf{E}(f|\mathcal{B})\|_{L^2(Z)}^2 \geq \mu^2/4.$$

The claim then follows from Pythagoras' theorem. □

We now have enough tools to prove Proposition 10.36.

*Proof of Proposition 10.36* We construct a nested pair of  $\sigma$ -algebras  $\mathcal{B} \subset \mathcal{B}'$  and an integer  $K \geq 1$  by the following double-loop algorithm.

- Step 0. Initialize  $\mathcal{B} = \{\emptyset, Z\}$ .
- Step 1. Let  $K$  be the smallest integer such that  $\mathbf{E}(f|\mathcal{B})$  is  $(K, \sigma/2)$ -almost periodic. (Note from the Fourier inversion formula that  $K$  is finite.) Set  $\mathcal{B}' := \mathcal{B}$ ; thus we trivially have  $\mathcal{E}_f(\mathcal{B}') \leq \mathcal{E}_f(\mathcal{B}) + \sigma^2/4$ .
- Step 2. If

$$\|f - \mathbf{E}(f|\mathcal{B}')\|_{u^2(Z)} \leq \frac{1}{F(\sigma, K)}$$

then we terminate the algorithm. If not, then we can apply Lemma 10.40 with  $\varepsilon := \frac{1}{4F(\sigma, K)}$  to obtain a new  $\sigma$ -algebra  $\mathcal{B}'' := \mathcal{B}' \vee \mathcal{B}_{\varepsilon, \xi}$  for some  $\xi \in Z$  such that

$$\mathcal{E}_f(\mathcal{B}'') \geq \mathcal{E}_f(\mathcal{B}') + \frac{1}{4F(\sigma, K)^2}.$$

- Step 3. If we have

$$\mathcal{E}_f(\mathcal{B}'') \leq \mathcal{E}_f(\mathcal{B}) + \sigma^2/4$$

then we set  $\mathcal{B}' := \mathcal{B}''$  and return to Step 2. If instead we have

$$\mathcal{E}_f(\mathcal{B}'') > \mathcal{E}_f(\mathcal{B}) + \sigma^2/4$$

then we set  $\mathcal{B} = \mathcal{B}''$  and return to Step 1.

Observe that every time we return from Step 3 to Step 2, the energy  $\mathcal{E}_f(\mathcal{B}')$  increases by at least  $\frac{1}{4F(\sigma, K)^2}$ , while  $K$  does not change. On the other hand, since  $f$  is bounded,  $\mathcal{E}_f(\mathcal{B}')$  varies between 0 and 1. Thus we can only return from Step 3 to Step 2 at most  $4F(\sigma, K)^2$  times before either terminating or returning to Step 1. Now, every time one returns from Step 3 to Step 1, the energy  $\mathcal{E}_f(\mathcal{B})$  increases by at least  $4/\sigma^2$ , so one can only return from Step 3 to Step 1 at most  $4/\sigma^2$  times. Thus this algorithm terminates after a finite number of steps. If we then set  $f_{U^\perp} := \mathbf{E}(f|\mathcal{B}')$  and  $f_U := f - \mathbf{E}(f|\mathcal{B}')$  we have  $f = f_U + f_{U^\perp}$ , that  $\|f_U\|_{u^2(Z)} \leq \frac{1}{F(\sigma, K)}$ , that

$0 \leq f_{U^\perp} \leq 1$ , and  $\mathbf{E}_Z f_{U^\perp} = \mathbf{E}_Z f$ . Finally, from construction we have  $\mathcal{E}_f(\mathcal{B}') \leq \mathcal{E}_f(\mathcal{B}) + \sigma^2/4$  and hence by Pythagoras's theorem  $\|f_{U^\perp} - \mathbf{E}(f|\mathcal{B})\|_{L^2(Z)} \leq \sigma/2$ . Since  $\mathbf{E}(f|\mathcal{B})$  is  $(K, \sigma/2)$ -almost periodic by construction, we conclude that  $f_{U^\perp}$  is  $(K, \sigma)$ -almost periodic.

The only remaining thing to verify is that  $K = O_{\sigma, F}(1)$ . Observe that at every stage,  $\mathcal{B}$  and  $\mathcal{B}'$  are the join of a finite number of  $\sigma$ -algebras of the form  $\mathcal{B}_{\varepsilon, \xi}$ . In particular, Corollary 10.39 applies to these  $\sigma$ -algebras. An easy induction argument then shows that at every stage of the iteration,  $\mathcal{B}$  and  $\mathcal{B}'$  are the join of at most  $O_{\sigma, F}(1)$   $\sigma$ -algebras, that the parameters  $\varepsilon$  involved are bounded from below by  $\Omega_{\sigma, F}(1)$ , and the parameter  $K$  is always bounded above by  $O_{\sigma, F}(1)$ . The claim follows.  $\square$

### Exercises

- 10.5.1 Let  $f, g : Z \rightarrow \mathbf{C}$  be functions bounded in magnitude by 1 which are both  $(K, \sigma)$ -almost periodic for some  $0 < \sigma \leq 1$ . Show that  $f + g$  is  $(2K, 2\sigma)$ -almost periodic, and that  $fg$  is  $(K^2, 4\sigma)$ -almost periodic.
- 10.5.2 Let  $f : Z \rightarrow \mathbf{C}$  be  $(K, \sigma)$ -almost periodic. Show that one can cover the set  $\{T^h f : h \in Z\} \subset L^2(Z)$  by at most  $O_{K, \sigma}(1)$  balls of radius  $2\sigma$  in the  $L^2(Z)$  metric. Conclude that

$$\mathbf{P}_{h \in Z}(\|T^h f - f\|_{L^2(Z)} \leq 4\sigma) = \Theta_{K, \sigma}(1),$$

which may help explain the terminology “almost periodic”. For a converse to this result, see Exercise 10.5.5 below.

- 10.5.3 Let  $\xi_1, \dots, \xi_n$  be a dissociated subset of  $Z$ . Using Rudin's inequality (Lemma 4.33), show that

$$\mathbf{P}_{h \in Z} \left( \sum_{j=1}^n |e(\xi_j \cdot h) - 1|^2 < n \right) \leq \exp(-\Omega(n)). \quad (10.22)$$

- 10.5.4 Let  $f : Z \rightarrow \mathbf{C}$  be such that  $\|f\|_{L^2(Z)} = \|\hat{f}\|_{l^2(Z)} \geq 4\sigma$  and  $\|\hat{f}\|_{l^\infty(Z)} \leq \delta\sigma$  for some  $\sigma, \delta > 0$ . Establish the bound

$$\mathbf{P}_{h \in Z} \left( \sum_{\xi \in Z} |e(\xi \cdot h) - 1|^2 |\hat{f}(\xi)|^2 < 4\sigma^2 \right) = O(\delta^c)$$

for some absolute constant  $c > 0$ . (Hint: normalize  $\|\hat{f}\|_{l^2(Z)} = 1$ , so  $\sigma \leq 1/4$ , and then let  $\xi_1, \dots, \xi_n$  be independent identical random variables with probability distribution  $|\hat{f}(\xi)|^2$ . Show that  $\xi_1, \dots, \xi_n$  are dissociated with probability  $1 - O(2^n \delta)$ , and apply (10.22) combined with the first moment method. Then optimize in  $n$ .)