1 Arithmetic progressions in primes I

after B. Green and T. Tao [3][1]
A summary written by Alexander Gorodnik[2]

1.1 Introduction

This chapter contains the first part of an exposition of

Theorem 1 (Green,Tao [3])**.** *Let A be a subset of positive upper density in the set $\mathcal{P}$ of prime numbers, i.e.,*

$$\limsup_{N\to\infty} \frac{|A\cap[1,N]|}{|\mathcal{P}\cap[1,N]|} > 0.$$

Then for any $k\ge 3$, A contains an arithmetic progression of length k.

We start our discussion with the celebrated theorem of Szemerédi, which can be stated in several equivalent forms (see Section 1.1.1 below for basic notation):

Theorem 2 (Szemerédi)**.** *1. Let A be a set of positive integers such that*

$$\limsup_{N\to\infty} \frac{1}{N}|A\cap[1,N]| > 0.$$

Then for any $k\ge 3$, A contains an arithmetic progression of length k.

2. *Given $k\ge 3$ and $\delta>0$, there exists $N_0=N_0(k,\delta)>0$ such that for every $N>N_0$ and every $A\subset[1,N]$ with $|A|>\delta N$, the set A contains an arithmetic progression of length k.*

3. *Given $k\ge 3$ and $\delta>0$, there exists $c=c(k,\delta)>0$ such that for a function $f:\mathbb{Z}_N\to\mathbb{R}$ satisfying $0\le f\le 1$ and $\mathbb{E}(f)\ge\delta$, we have*

$$\mathbb{E}(f(x)f(x+r)\cdots f(x+(k-1)r)|x,r\in\mathbb{Z}_N)\ge c$$

for sufficiently large N.

[1] These notes are prepared for the workshop "Analysis and Ergodic Theory" held at Lake Arrowhead, September 2006

[2] The author is partially supported by NSF 0400631

Although there are explicit estimates on the constant N_0 in Theorem 2(2), the set $\mathcal{P}$ of prime numbers is too sparse to deduce existance of arithmetic progressions in $\mathcal{P}$ directly using known estimates. Recall that according to the prime number theorem,

$$|\mathcal{P} \cap [1, N]| \sim \frac{N}{\log N} \quad \text{as } N \to \infty.$$

The starting point of the Green-Tao argument is to consider $\mathcal{P}$ as a subset of the set $\mathcal{P}_R$ of almost primes, which consists of numbers all of whose prime factors are at least R. When $R = N^\alpha$ for small $\alpha > 0$, the set $\mathcal{P}_R \cap [1, N]$ is relatively well understood (for example, one can prove Theorem 1 for $\mathcal{P}_R$ using sieve methods), and by Mertens' theorem,

$$\limsup_{N\to\infty} \frac{|\mathcal{P} \cap [1, N]|}{|\mathcal{P}_R \cap [1, N]|} > 0.$$

The proof of Theorem 1 uses information about the structure of the set of almost primes ingeniously combined with positive density argument as in Theorem 2. In fact, it is proved that a subset of positive upper density in a so-called pseudorandom set, which will be defined in Section 1.1.2, contains arbitrary long arithmetic progressions.

Theorem 3 (relative Szemerédi theorem)**.** *Fix $k \ge 3$ and $\delta > 0$, and let $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ be a k-pseudorandom function such that $\mathbb{E}(\nu) = 1 + o(1)$. Then for any function $f : \mathbb{Z}_N \to \mathbb{R}^+$ satisfying $0 \le f \le \nu$ and $\mathbb{E}(f) \ge \delta$, we have*

$$\mathbb{E}(f(x)f(x + r) \cdots f(x + (k - 1)r)|x, r \in \mathbb{Z}_N) \ge c - o_{k,\delta}(1) \qquad (1)$$

as $N \to \infty$, where $c > 0$ is the same as in Theorem 2(3).

At this stage, we allow the reader to think naively that ν is the normalized characteristic function of the set of almost primes in $\mathbb{Z}_N$ and f is the normalized characteristic function of the set of primes in $\mathbb{Z}_N$. However, to give a rigorous argument, one needs to consider "smoothened" version of these function and to eliminate the irregularities coming for congruence properties of primes. The construction of functions ν and f for which Theorem 3 implies Theorem 1 will be given in the following chapter. In this chapter, we discuss the proof of Theorem 3.

Sketch of the proof of Theorem 3. The main ingredient of the proof is Theorem 9 below which implies that any nonnegative function f which is majorated by a k-pseudorandom ν has a decomposition

$$f = f_U + f_{U^\perp} + E \tag{2}$$

with the error term E satisfying

$$E \geq 0 \quad \text{and} \quad \mathbb{E}(E) = o(1). \tag{3}$$

Decomposition (2) is analogous to the ordinary Koopman–von-Neumann decomposition (see Remark 4), and it has the following properties:

$$f_U + f_{U^\perp} \geq 0, \tag{4}$$
$$0 \leq f_{U^\perp} \leq 1 + o(1), \tag{5}$$
$$\mathbb{E}(f) = \mathbb{E}(f_{U^\perp}) + o(1), \tag{6}$$
$$\mathbb{E}(f_0(x)f_1(x+r)\cdots f_{k-1}(x+(k-1)r)|x, r \in \mathbb{Z}_N) \quad \text{is small,} \tag{7}$$

where each f_i is either f_U or $f_{U^\perp}$, and $f_i \neq f_{U^\perp}$ for some i.

Now it follows from (3) and (4) that

$$\begin{aligned} &\mathbb{E}(f(x)\cdots f(x+(k-1)r)|x, r \in \mathbb{Z}_N) \\ \geq &\mathbb{E}((f_U + f_{U^\perp})(x)\cdots(f_U + f_{U^\perp})(x+(k-1)r)|x, r \in \mathbb{Z}_N). \end{aligned}$$

Because of (7), it suffices to prove a lower estimate for

$$\mathbb{E}(f_{U^\perp}(x)f_{U^\perp}(x+r)\cdots f_{U^\perp}(x+(k-1)r)|x, r \in \mathbb{Z}_N).$$

Using that the function $f_{U^\perp}$ satisfies (5) and (6), this lower estimate follows from the Szemerédi theorem (Theorem 2(3)). □

Remark 4 (comparison with ergodic theory)**.** When $k = 3$, decomposition (2) is a finitary analog of the Koopman–von-Neumann decomposition of $L^2(X)$, defined for a probability measure-preserving system $(X, \mathcal{B}, \mu, T)$, into weakly mixing and compact (almost periodic) parts. Note that the later decomposition can be used to show that for any nonnegative $f \in L^\infty(X)$ with $\int_X f\, d\mu > 0$, one has

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \int_X f(x)f(T^i x)f(T^{2i}x)d\mu(x) > 0.$$

This implies the Szemerédi theorem (Theorem 2) with $k = 3$.

For general k, similar decompositions appeared in the Furstenberg's proof of the Szemerédi theorem [1] and in the works of Host, Kra [2] and Ziegler [4]. These decompositions are of the form

$$f = f_U + f_{U^\perp} \quad \text{with} \quad f_U = f - \mathbb{E}(f|\mathcal{B}') \text{ and } f_{U^\perp} = \mathbb{E}(f|\mathcal{B}'),$$

where $\mathcal{B}'$ is a T-invariant sub-σ-algebra of $\mathcal{B}$, such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \int_X f_0(x) f_1(T^i x) \cdots f_{k-1}(T^{(k-1)i} x)\, d\mu(x) = 0,$$

where each f_i is either f_U or $f_{U^\perp}$, and $f_i \neq f_{U^\perp}$ for some i, and the average

$$\frac{1}{N} \sum_{i=0}^{N-1} \int_X f_{U^\perp}(x) f_{U^\perp}(T^i x) \cdots f_{U^\perp}(T^{(k-1)i} x)\, d\mu(x)$$

can be analyzed using some additional structure.

1.1.1 Notation

- $k \geq 3$ is the length of the arithmetic progression (fixed).
- N is a large prime number ($N \to \infty$).
- $o(1)$ and $O(1)$ denote the quantities that go to zero and bounded respectively as $N \to \infty$.
- $\mathbb{Z}_N$ is the field of residues mod N.
- For a function $f : \mathbb{Z}_N^l \to \mathbb{R}$ and $A \in \mathbb{Z}_N^l$,

  $$\mathbb{E}(f(x) \,|\, x \in A) := \frac{1}{|A|} \sum_{x \in A} f(x) \quad \text{and} \quad \|f\|_{L^q} := \mathbb{E}(|f(x)|^q \,|\, x \in \mathbb{Z}_N^l)^{1/q}.$$

  In particular, $\mathbb{E}(f) := \mathbb{E}(f(x)|x \in \mathbb{Z}_N^l)$. Also, $\|\cdot\|_{L^\infty}$ denotes the sup-norm.

- Given a σ-algebra $\mathcal{B}$ on $\mathbb{Z}_N$, we define the conditional expectation $\mathbb{E}(f|\mathcal{B})$, $f \in L^2(\mathbb{Z}_N)$, to be the orthogonal projection on the space of $\mathcal{B}$-measurable functions, i.e.,

  $$\mathbb{E}(f|\mathcal{B})(x) = \mathbb{E}(f(y) \,|\, y \in \mathcal{B}(x))$$

  where $\mathcal{B}(x)$ denotes the atom of the σ-algebra $\mathcal{B}$ containing x.

- For σ-algebras $\mathcal{B}_1$ and $\mathcal{B}_2$, $\mathcal{B}_1 \vee \mathcal{B}_2$ denotes the smallest σ-algebra containing both $\mathcal{B}_1$ and $\mathcal{B}_2$.

1.1.2 Pseudorandom measures

A function $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ is called a *measure* if

$$\mathbb{E}(\nu) = 1 + o(1).$$

A measure ν satisfies (m_0, t_0, L_0)-*linear form condition* if for any $m \le m_0$, $t \le t_0$, and linear forms

$$\psi_i(x) = \sum_{j=1}^{t} L_{ij} x_j + b_i, \quad i = 1, \dots, m,$$

where L_{ij} are rational numbers[3] with the numerator and denominator bounded by L_0 in absolute value and $b \in \mathbb{Z}_N$, such that the t-tuples $(L_{ij}; j = 1, \dots, t)$ are not zero and not rational multiples of each other, we have

$$\mathbb{E}(\nu(\psi_1(x)) \cdots \nu(\psi_m(x)) | x \in \mathbb{Z}_N^t) = 1 + o_{m_0,t_0,L_0}(1).$$

Roughly speaking, the measure ν will be supported on almost primes and the linear form condition says that events "$\psi_j(x)$ is almost prime" are essentially independent.

The linear form condition is motivated by the Hardy-Littlewood prime tuples conjecture. Let $\Lambda(n)$ denote the Mangoldt function, which is equal to $\log p$ if n is a power of prime p and zero otherwise, and $\Lambda_p(n)$ denote the local Mangoldt function, which is equal to $\frac{p}{p-1}$ if $(n, p) = 1$ and zero otherwise. Let ψ_i's be the linear forms as above with positive L_{ij} and b_i.

[3]Recalling that N is a large prime, we can view L_{ij}'s as elements of $\mathbb{Z}_N$.

Conjecture 5 (Hardy-Littlewood prime tuple conjecure)**.**

$$\mathbb{E}(\Lambda(\psi_1(x))\cdots\Lambda(\psi_m(x))|x \in \mathbb{Z}_N^t) = \prod_p \alpha_p + o_{m_0,t_0,L_0}(1)$$

where

$$\alpha_p = \mathbb{E}(\Lambda_p(\psi_1(x))\cdots\Lambda_p(\psi_m(x))|x \in \mathbb{Z}_p^t).$$

The linear form condition is analogous to the following property that holds for any weakly mixing probability measure-preserving system $(X, \mathcal{B}, \mu, T)$: for any $f_1, \ldots, f_m \in L^\infty(X)$,

$$\frac{1}{N}\sum_{i=0}^{N-1}\int_X f_1(T^i x)\cdots f_m(T^{mi}x)\,d\mu(x) \to \left(\int_X f_1 d\mu\right)\cdots\left(\int_X f_m d\mu\right). \quad (8)$$

A measure ν satisfies m_0*-correlation condition* if for every $m = 2, \ldots, m_0$, there exists a function $\tau = \tau_m : \mathbb{Z}_N \to \mathbb{R}^+$ such that

$$\mathbb{E}(\tau^q) = O_{m,q}(1)$$

for all $q \geq 1$ and

$$\mathbb{E}(\nu(x+h_1)\cdots\nu(x+h_m)|x \in \mathbb{Z}_N) \leq \sum_{1\leq i<j\leq m}\tau(h_i - h_j)$$

for all $h_1, \ldots, h_m \in \mathbb{Z}_N$.

A measure ν is called k*-pseudorandom* if it satisfies $(k2^{k-1}, 3k-4, k)$-linear form condition and 2^{k-1}-correlation condition.

1.2 Tools

1.2.1 Gowers uniformity norms

We define *d-th Gowers uniformity norm* $\|\cdot\|_{U^d}$ inductively. Denote by T_h the shift operator: $(T_h f)(x) = f(x+h)$. For $f : \mathbb{Z}_N \to \mathbb{R}$, we set

$$\|f\|_{U^1} = |\mathbb{E}(f(x)|x \in \mathbb{Z}_N)|,$$
$$\|f\|_{U^d} = \mathbb{E}\left(\|f\cdot(T_h f)\|_{U^{d-1}}^{2^{d-1}}|h \in \mathbb{Z}_N\right)^{1/2^d}.$$

Explicitly,

$$\|f\|_{U^d} = \mathbb{E}\left(\prod_{\omega\in\{0,1\}^d} f(x+\omega h) \,|\, x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^d\right)^{1/2^d}.$$

One can show that for $d \geq 2$, $\|\cdot\|_{U^d}$ is a genuine norm.

Given a family of functions f_ω, $\omega \in \{0,1\}^d$, we define *d-dimensional Gowers inner product*:

$$\langle (f_\omega)_{\omega\in\{0,1\}^d}\rangle_{U^d} = \mathbb{E}\left(\prod_{\omega\in\{0,1\}^d} f_\omega(x+\omega h) \,|\, x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^d\right).$$

Then we have *Gowers Cauchy-Schwarz inequality:*

$$|\langle (f_\omega)_{\omega\in\{0,1\}^d}\rangle_{U^d}| \leq \prod_{\omega\in\{0,1\}^d} \|f_\omega\|_{U^d}.$$

Gowers uniformity norms can be used to control the averages as in (1):

Theorem 6 (generalized von Neumann theorem)**.** *For a k-pseudorandom measure* $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ *and functions* $f_0, \ldots, f_{k-1} : \mathbb{Z}_N \to \mathbb{R}$ *such that* $|f_i| \leq \nu + 1$*, we have*

$$\mathbb{E}\left(\prod_{i=0}^{k-1} f_i(x+ih) \,|\, x,h\in\mathbb{Z}_N\right) = O\left(\min_{0\leq i\leq k-1} \|f_i\|_{U^{k-1}}\right) + o(1).$$

Theorem 6 is a finitary analog of (8). The proof uses van der Corput type argument and the linear form condition.

We call a function f *Gowers uniform* if $\|f\|_{U^{k-1}}$ is small. Theorem 6 shows that if at least one of f_i's is Gowers uniform functions, the average $\mathbb{E}\left(\prod_{i=0}^{k-1} f_i(x+ih)|x,h\in\mathbb{Z}_N\right)$ is negligible. This is used to arrange (7).

1.2.2 Obstructions to uniformity and dual functions

For a function $F : \mathbb{Z}_N \to \mathbb{R}$, define the *dual function* of F:

$$\mathcal{D}F(x) = \mathbb{E}\left(\prod_{\omega\in\{0,1\}^{k-1}:\omega\neq 0^{k-1}} F(x+\omega h) \mid h\in\mathbb{Z}_N^{k-1}\right).$$

Note that

$$\langle F, \mathcal{D}F \rangle = \|F\|_{U^{k-1}}^{2^{k-1}}. \tag{9}$$

Hence, if the Gowers norm of F is large, F correlates with its dual function, and the dual functions provide obstructions to uniformity.

The following two properties of dual functions are used in the proof.

Proposition 7. *For a function $F : \mathbb{Z}_N \to \mathbb{R}$ such that $|F| \leq \nu + 1$, we have*

$$\|\mathcal{D}F\|_{L^\infty} \leq 2^{2^{k-1}-1} + o(1).$$

Proposition 7 is deduced from the linear form condition.

Proposition 8. *Let $I = [-2^{2^k}, 2^{2^k}]$. Given function $F_1, \ldots, F_n : \mathbb{Z}_N \to \mathbb{R}$ such that $|\mathcal{D}F_i| \leq 2^{2^k}$ and a continuous function $\Phi : I^n \to \mathbb{R}$, we define*

$$\psi(x) = \Phi(\mathcal{D}F_1(x), \ldots, \mathcal{D}F_n(x)).$$

Then

$$\langle \nu - 1, \psi \rangle = o_{n,\Psi}(1).$$

Proposition 8 is deduced from the correlation condition using the Gowers Cauchy-Schwarz inequality.

1.2.3 σ-Algebras generated by generalized Bohr sets

Fix $\varepsilon, \eta > 0$. Given a function $G : \mathbb{Z}_N \to I := [-2^{2^k}, 2^{2^k}]$, one defines a σ-algebra $\mathcal{B}_{\varepsilon,\eta}(G)$ on $\mathbb{Z}_N$ that satisfies the following properties:

1. For any σ-algebra $\mathcal{B}$ on $\mathbb{Z}_N$,

$$\|G - \mathbb{E}(G|\mathcal{B} \vee \mathcal{B}_{\varepsilon,\eta}(G))\|_{L^\infty} \leq \varepsilon. \tag{10}$$

2. The σ-algebra $\mathcal{B}_{\varepsilon,\eta}(G)$ is generated by at most $O(1/\varepsilon)$ atoms.

3. If A is any atom of $\mathcal{B}_{\varepsilon,\eta}(G)$, then there exists a continuous function $\Psi_A : I \to [0,1]$ such that

$$\|(1_A - \Psi_A(G))(\nu + 1)\|_{L^1} = O(\eta). \tag{11}$$

   In fact, this implies that if $G_1, \ldots, G_n : \mathbb{Z}_N \to I$ and A is an atom of $\mathcal{B}_{\varepsilon,\eta}(G_1) \vee \cdots \vee \mathcal{B}_{\varepsilon,\eta}(G_n)$, then there exists a continuous function $\Psi_A : I^n \to [0,1]$ such that

$$\|(1_A - \Psi_A(G_1, \ldots, G_n))(\nu + 1)\|_{L^1} = O_n(\eta). \tag{12}$$

Roughly speaking, the σ-algebra $\mathcal{B}_{\varepsilon,\eta}(G)$ is generated by the level sets of the function G: the atoms of $\mathcal{B}_{\varepsilon,\eta}(G)$ are $G^{-1}([\varepsilon(n+\alpha),\varepsilon(n+1+\alpha)))$, $n\in\mathbb{Z}$, for suitably chosen α.

1.3 Structure theorem

Theorem 9 (generalized Koopman–von Neumann structure theorem)**.** *Let ν be a k-pseudorandom measure and $f:\mathbb{Z}_N\to\mathbb{R}$ such that $0\le f\le\nu$. Let $\varepsilon>0$ be a small parameter and $N>N_0(\varepsilon)$ sufficiently large. Then there exists a σ-algebra $\mathcal{B}$ and an exceptional set $\Omega\in\mathcal{B}$ such that*

$$\mathbb{E}(1_\Omega\nu)=o_\varepsilon(1), \tag{13}$$

$$\|(1-1_\Omega)\mathbb{E}(\nu-1|\mathcal{B})\|_{L^\infty}=o_\varepsilon(1), \tag{14}$$

$$\|(1-1_\Omega)(f-\mathbb{E}(f|\mathcal{B}))\|_{U^{k-1}}\le\varepsilon^{1/2^k}. \tag{15}$$

Now setting

$$f_U=(1-1_\Omega)(f-\mathbb{E}(f|\mathcal{B})),\ f_{U^\perp}=(1-1_\Omega)\mathbb{E}(f|\mathcal{B}),\ E=1_\Omega f,$$

we have decomposition (2) satisfying (3)–(7). Note that (3)–(6) follow directly from Theorem 9, and (7) is derived using Theorem 6.

Sketch of the proof of Theorem 9. In the proof, we use a parameter $\eta\to 0^+$.

First, we set $\mathcal{B}_0=\{\emptyset,\mathbb{Z}_N\}$ and $\Omega_0=\emptyset$. Then (13) and (14) obviously hold. If (15) fails, we set

$$F_1:=(1-1_{\Omega_0})(f-\mathbb{E}(f|\mathcal{B}_0)),$$
$$\mathcal{B}_1:=\mathcal{B}_0\vee\mathcal{B}_{\varepsilon,\eta}(\mathcal{D}F_1),$$

and define the exceptional set Ω_1 to be the union of the atoms $A\in\mathcal{B}_1$ such that $\mathbb{E}(1_A(\nu+1))\le\eta^{1/2}$. Then

$$\mathbb{E}(1_{\Omega_1}(\nu+1))=O_\varepsilon(\eta^{1/2}),$$

and using (11) and Proposition 8, one shows that

$$\|(1-1_{\Omega_1})\mathbb{E}(\nu-1|\mathcal{B}_1)\|_{L^\infty}=O_\varepsilon(\eta^{1/2}).$$

Continuing this procedure, we construct sequences of functions

$$F_n:=(1-1_{\Omega_{n-1}})(f-\mathbb{E}(f|\mathcal{B}_{n-1})),$$

σ-algebras

$$\mathcal{B}_n := \mathcal{B}_{n-1} \vee \mathcal{B}_{\varepsilon,\eta}(\mathcal{D}F_n),$$

and exceptional sets $\Omega_n \in \mathcal{B}_n$ satisfying

$$\mathbb{E}(1_{\Omega_n}(\nu+1)) = O_{n,\varepsilon}(\eta^{1/2}), \tag{16}$$

$$\|(1-1_{\Omega_n})\mathbb{E}(\nu-1|\mathcal{B}_n)\|_{L^\infty} = O_{n,\varepsilon}(\eta^{1/2}). \tag{17}$$

Note that one can check inductively that

$$\|(1-1_{\Omega_n})\mathbb{E}(f|\mathcal{B}_n)\|_{L^\infty} \leq 1 + O_{n,\varepsilon}(\eta^{1/2}), \tag{18}$$

$$|F_n| \leq (1+O_{n,\varepsilon}(\eta^{1/2}))(\nu+1). \tag{19}$$

By (19), (12) and Proposition 8 can be applied at every step to deduce (17).
It remains to show that after finitely many steps, we get

$$\|F_n\|_{U^{k-1}} \leq \varepsilon^{1/2^k}.$$

This follows from the following claim compared with estimate (18) when η is chosen sufficiently small.

Claim (energy increment property). *If* $\|F_n\|_{U^{k-1}} > \varepsilon^{1/2^k}$, *then*

$$\|(1-1_{\Omega_n})\mathbb{E}(f|\mathcal{B}_n)\|_{L^2}^2 > \|(1-1_{\Omega_{n-1}})\mathbb{E}(f|\mathcal{B}_{n-1})\|_{L^2}^2 + 2^{-2^k+1}\varepsilon.$$

Heuristically, the claim follows from the observation that if F_n is not Gowers uniform, then by (9) it has a nontrivial correlation with the dual function $\mathcal{D}F_n$.

Since the contribution of the exceptional sets can be controlled using (16), we sketch the proof assuming that the exceptional sets are empty. Using that by (9),

$$\langle F_n, \mathcal{D}F_n\rangle = \|F_n\|_{U^{k-1}}^{2^{k-1}} > \varepsilon^{1/2},$$

and by (10),

$$\langle F_n, \mathcal{D}F_n - \mathbb{E}(\mathcal{D}F_n|\mathcal{B}_n)\rangle = O(\varepsilon),$$

we deduce that

$$\langle F_n, \mathbb{E}(\mathcal{D}F_n|\mathcal{B}_n)\rangle = \langle f - \mathbb{E}(f|\mathcal{B}_{n-1}), \mathbb{E}(\mathcal{D}F_n|\mathcal{B}_n)\rangle > \varepsilon^{1/2} + O(\varepsilon).$$

In the last inequality, we can replace f by $\mathbb{E}(f|\mathcal{B}_n)$. Then by the Cauchy-Schwarz inequality and Proposition 7,

$$\|\mathbb{E}(f|\mathcal{B}_n) - \mathbb{E}(f|\mathcal{B}_{n-1})\|_{L^2} \cdot (2^{2^{k-1}-1} + O_{n,\varepsilon}(\eta^{1/2})) > \varepsilon^{1/2} + O(\varepsilon).$$

Since $\mathbb{E}(f|\mathcal{B}_n) - \mathbb{E}(f|\mathcal{B}_{n-1}) \perp \mathbb{E}(f|\mathcal{B}_{n-1})$, the claim now follows from the Pythagoras theorem (with an appropriate choice of parameter η). $\square$

References

[1] H. Furstenberg, Recurrence in ergodic theory and combinatorial number theory. M. B. Porter Lectures. Princeton University Press, Princeton, N.J., 1981.

[2] B. Host and B. Kra, Nonconventional ergodic averages and nilmanifolds. Ann. of Math. (2) 161 (2005), no. 1, 397–488.

[3] B. Green and T. Tao, The primes contain arbitrary long arithmetic progressions. To appear in Ann. Math.

[4] T. Ziegler, Universal characteristic factors and Furstenberg averages. To appear in JAMS.

Alexander Gorodnik, CalTech
email: **gorodnik@caltech.edu**