Linear Equations in Primes

A Survey

Maxim Gerspach

April 18, 2017

Contents

1 The Primes contain arbitrarily long arithmetic progressions

1.1 Introduction

In this chapter, we closely follow [7].

We start by stating the following result due to Szemerédi [13]

Theorem 1.1. *Let N be a positive integer, let $\delta > 0$ be fixed, and let $k \geq 3$ be an integer. Then there is a positive integer $N_0 = N_0(\delta, k)$ with the following property. If $N \geq N_0$ and $A \subseteq \mathbb{Z}_N$ is any set of cardinality at least δN, then A contains an arithmetic progression of length k.*

One can show that an equivalent form of Szemerédi's Theorem is the following

Theorem 1.2. *Let $0 < \delta \leq 1$ and $k \geq 1$ be fixed. Let N be a sufficiently large integer parameter, and let $f : \mathbb{Z}_N \to \mathbb{R}^+$ be a function satisfying*

$$0 \leq f(x) \leq 1$$

for all $x \in \mathbb{Z}_N$ and

$$\mathbb{E}_{x \in \mathbb{Z}_N} f(x) \geq \delta.$$

Then we have

$$\mathbb{E}_{x,r \in \mathbb{Z}_N}[f(x)f(x+r)\cdots f(x+(k-1)r)] \geq c(k,\delta) - o_{k,\delta}(1)$$

for some constant $c(k, \delta) > 0$ independent of N and f.

Over the years, new proofs of this statement have been developed. However, it is believed that this statement holds even when the density of A is slowly decreasing with N. More precisely, there is the following

Conjecture 1.3 (Erős)**.** Let A be a set of positive integers satisfying $\sum_{a \in A} \frac{1}{a} = \infty$. Then A contains arbitrarily long arithmetic progressions.

There has been some progress towards a decreasing density of A. Namely, there is the following result due to Gowers [3, 4]:

Theorem 1.4 (Gowers)**.** *For every positive integer k there is a constant $c = c(k) > 0$ such that every subset of $\{1, \ldots, N\}$ of size at least $N(\log\log N)^{-c}$ contains an arithmetic progression of length k.*

The goal in this chapter will be to explain the main ideas coming into the proof of the following long-conjectured Theorem due to Green and Tao [7].

Main Theorem 1.5 (Green-Tao)**.** *The primes contain infinitely many arithmetic progressions of any length.*

Note that the Theorem of Gowers, despite being a fundamental breakthrough and inspiration for this result, is still far from being applicable to all sets of similar density as the primes.

It is in fact not much harder to show the following slightly stronger

Theorem 1.6. *Let A be any subset of the primes of positive relative upper density, i.e.*

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

Then A contains infinitely many arithmetic progressions of any length.

This is also due to Green and Tao. Note that the proofs of these Theorems do use some arithmetic properties of primes, and can thus not be easily extended to sets of the same (or higher) density as the primes.

1.2 Pseudorandom measures

In the following discussion N always denotes a large prime.

The fundamental notion introduced by Green and Tao to be able to transfer the problem from sets of positive densities to more general sets is the notion of pseudorandom measures. To introduce these, we first need the following definitions:

Definition 1.7. *Let $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ be a function (it might be more precise to call it a sequence of functions indexed by N). ν is said to be a measure if*

$$\mathbb{E}_{n\in\mathbb{Z}_N} \nu(n) = 1 + o(1). \tag{1.1}$$

Definition 1.8. *Let $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ be a measure. Let (t_0, d_0, L_0) be a triple of positive integers. Then we say that ν satisfies the (t_0, d_0, L_0)-linear forms condition if the following holds:*

Let $1 \le d \le d_0$, $1 \le t \le t_0$, and let $(L_{ij})_{1\le i\le t, 1\le j\le d}$ be integers bounded in absolute value by L_0. Moreover, let $b_i \in \mathbb{Z}, 1 \le i \le t$, and define the affine-linear forms $\psi_i : \mathbb{Z}^d \to \mathbb{Z}$,

$$\psi_i(x) := \sum_{j=1}^{d} L_{ij} x_j + b_i.$$

Then we have

$$\mathbb{E}_{n\in\mathbb{Z}_N^d} \left[\prod_{i=1}^{t} \nu(\psi_i(n)) \right] = 1 + o_{t_0,d_0,L_0}(1). \tag{1.2}$$

Note that every ψ_i induces a well-defined map $\mathbb{Z}_N^d \to \mathbb{Z}_N$.

Informally, this states that the function ν does not correlate with itself on affine-linear forms. We will later apply this in the context of prime numbers, where ν will essentially be concentrated on the primes. In this case, the linear forms condition essentially states

that the events "$\psi_j(x)$ is almost prime" are asymptotically independent of each other as j varies. Note that for $t = 1$, this also gives the measure condition.

We just give one special case as an example. If (t_0, d_0, L_0) is at least $(4, 3, 1)$, we in particular obtain that

$$\mathbb{E}_{h_1,h_2\in\mathbb{Z}_N}[\nu(x)\nu(x+h_1)\nu(x+h_2)\nu(x+h_1+h_2)] = 1 + o_{t_0,d_0,L_0}(1).$$

In terms of the definitions we give in section 1.5, this asserts that the Gowers norm $U^2(\mathbb{Z}_N)$ of ν is close to 1. Similar statements can of course be derived for higher Gowers norms by choosing the parameters accordingly.

Definition 1.9. *Let $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ be a measure, and let m_0 be a positive integer. ν is said to satisfy the m_0-correlation condition if for every $1 \le m \le m_0$ there exists a weight function $\tau = \tau_m : \mathbb{Z}_N \to \mathbb{R}^+$ satisfying*

$$\mathbb{E}_{n\in\mathbb{Z}_N}[\tau^q(n)] \ll_{m,q} 1 \tag{1.3}$$

for all $1 \le q < \infty$ and

$$\mathbb{E}_{n\in\mathbb{Z}_N}\left[\prod_{i=1}^{m}\nu(n+h_i)\right] \le \sum_{1\le i<j\le m}\tau(h_i - h_j) \tag{1.4}$$

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

This condition was, in some sense, constructed in such a way that it applies to the primes. For a rather uniformly distributed function ν one could expect a stronger bound on the right-hand side of the type $O_m(1)$, but the arithmetic properties of the primes lead to slight non-uniformities. For example, the number of primes $p \le N$ such that $p - h$ is also prime is not bounded uniformly in h by (a constant times) $N/\log^2 N$, as opposed to random sets of prime densities.

Definition 1.10. *Let D be a positive integer. A measure ν is called D-pseudorandom if it satisfies the $(D2^{D-1}, 3D-4, D)$-linear forms and the 2^{D-1}-correlation condition.*

The next claim illustrates the so-called transference principle very well. Compare the statement to 1.2.

Theorem 1.11. *Let $k \ge 3$ and $0 < \delta \le 1$ be fixed parameters. Suppose that $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ is k-pseudorandom. Let $f : \mathbb{Z}_N \to \mathbb{R}^+$ be a function satisfying*

$$0 \le f(x) \le \nu(x)$$

for all $x \in \mathbb{Z}_N$, and

$$\mathbb{E}_{x\in\mathbb{Z}_N} f(x) \ge \delta.$$

Then we have

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

Note that this time, f is not (necessarily) bounded by the constant function 1, but by a special type of function that has average value 1, namely a pseudorandom measure. This general idea applies for several statements in this area, where we have a (comparably simple) Theorem regarding bounded functions, and are able to apply it in some way to show it for functions which are bounded by pseudorandom measures. Of course, since the constant function 1 defines a k-pseudorandom measure for any value of k, this is indeed more general.

The proof of this Theorem constitutes the major part of the proof of the Green-Tao Theorem.

1.3 Notation

Notations such as the o-notation will usually be considered in the limit as $N \to \infty$; sometimes, we will also consider limits as certain variables go to zero, but we will explicitly say so. Variable indices in O-, o- and $\ll$-notation indicate that the statement holds when these variables are fixed.

For a given finite, non-empty set A we write $|A|$ for its cardinality and

$$\mathbb{E}_{x\in A}[f(x)] := \frac{1}{|A|}\sum_{x\in A} f(x)$$

to denote its average. For $N \in \mathbb{N}$, we write $[N] := \{1, \dots, N\}$. We denote by $\mathbb{Z}_N$ the cyclic group $\mathbb{Z}/N\mathbb{Z}$. We sometimes write $\log_m$, where the index will always mean the number of iterations of the natural logarithm, not the base.

1.4 Majorising the primes by a pseudorandom measure

In this whole chapter, we view k as fixed. Any dependencies of functions or implicit constants on k will be notationally omitted.

The goal of this section is to prove the Main Theorem 1.6 assuming Theorem 1.11 and two other propositions from complex analysis essentially due to Goldston, Pintz and Yıldırım, but also proved - in detail and adjusted to this setting - by Green and Tao in [7].

To this end, let $w = w(N) = \log_3 N$ and $W = \prod_{p\le w} p$. We define $\tilde{\Lambda} : \mathbb{N} \to \mathbb{R}^+$,

$$\tilde{\Lambda}(n) := \begin{cases} \frac{\varphi(W)}{W}\Lambda(Wn+1) & \text{if } Wn+1 \text{ is prime} \\ 0 & \text{else} \end{cases}.$$

Note that, from Dirichlet's Theorem on primes in APs, $\tilde{\Lambda}$ has average value 1. The fundamental observation now is the following

Proposition 1.12. *Set $\epsilon_k := 1/(2^k(k+4)!)$ and let N be a sufficiently large prime number. Then there exists a k-pseudorandom measure $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ such that $\nu(n) \ge k^{-1}2^{-k-5}\tilde{\Lambda}(n)$ for all $\epsilon_k N \le n \le 2\epsilon_k N$.*

Proof of Theorem 1.5 assuming Theorem 1.11 and Proposition 1.12. Define $f : \mathbb{Z}_N \to \mathbb{R}^+$ by $f(n) := k^{-1}2^{-k-5}\tilde{\Lambda}(n)$ for $\epsilon_k N \le n \le 2\epsilon_k N$ and 0 otherwise. We have

$$\mathbb{E}f = \frac{k^{-1}2^{-k-5}}{N} \sum_{\epsilon_k N \le n \le 2\epsilon_k N} \tilde{\Lambda}(n) = k^{-1}2^{-k-5}\epsilon_k(1 + o(1)).$$

By Proposition 1.12 we see that we may apply Theorem 1.11 to deduce

$$\mathbb{E}_{x,r\in\mathbb{Z}_N}[f(x)f(x+r)\cdots f(x+(k-1)r)] \ge c(k, k^{-1}2^{-k-5}\epsilon_k) - o(1).$$

The case $r = 0$ can contribute at most $O(\log^k N/N) = o(1)$ and may therefore be removed. Moreover, the fact that f vanishes outside $[\epsilon_k N, 2\epsilon_k N]$ implies that $x, x + k, \ldots, x+(k-1)r$ in fact defines an AP in $\mathbb{Z}$, not only in $\mathbb{Z}_N$ (if the pair (x, r) contributes to the above expectation). The claim now follows from the definition of f via $\tilde{\Lambda}$ by taking N sufficiently large. □

We have thus reduced the Main Theorem to the two statements 1.11 and 1.12. We start by reducing the latter in the way announced. Note that we have

$$\Lambda(n) = \sum_{d\,|\,n} \mu(d)\log(n/d) = \sum_{d\,|\,n} \mu(d)\log(n/d)_+,$$

where $\log_+$ denotes the positive part $\max(\log, 0)$ of the logarithm. Keeping this in mind, we can now make the following

Definition 1.13. *Let $R = R(N)$ be a parameter. We define the truncated divisor sum*

$$\Lambda_R(n) := \sum_{\substack{d\,|\,n \\ d\le R}} \mu(d)\log(R/d) = \sum_{d\,|\,n} \mu(d)\log(R/d)_+. \tag{1.5}$$

This enables us to define the k-pseudorandom measure which will majorise $\tilde{\Lambda}$, or more precisely, the function f from the subsequent proof.

Definition 1.14. *Let $R := N^{k^{-1}2^{-k-4}}$ and let ϵ_k be as in Proposition 1.12. Define $\nu : \mathbb{Z}_N \to \mathbb{R}^+$,*

$$\nu(n) := \begin{cases} \frac{\varphi(W)}{W}\frac{\Lambda_R(Wn+1)^2}{\log R} & \text{if } \epsilon_k N \le n \le 2\epsilon_k N \\ 1 & \text{else} \end{cases}.$$

We have to show several claims to prove Proposition 1.12.

(i) ν majorises f,

(ii) ν defines a measure,

(iii) ν satisfies the $(k2^{k-1}, 3k-4, k)$-linear forms condition and

(iv) ν satisfies the 2^{k-1}-correlation condition.

We will rely heavily on the following two propositions which are essentially due to Goldston, Pintz and Yıldırım, and which have quite technical proofs strongly using complex analysis.

Proposition 1.15. *Let d, t be positive integers. For $1 \le i \le t$, let $\psi_i : \mathbb{Z}^d \to \mathbb{Z}$,*

$$\psi_i(x) = \sum_{j=1}^{d} L_{ij} x_j + b_i$$

for some integers b_i and a collection of integer coefficients $|L_{ij}| \le \sqrt{w}/2$. Assume that the t-tuples $(L_{ij})_j$ are never identically zero, and that no two of the t-tuples are rational multiples of each other. Set $\theta_i := W\psi_i + 1$, and let $B = \prod_{i=1}^{t} I_i \subset \mathbb{R}^t$ be a product of intervals of length at least R^{10d}. Then

$$\mathbb{E}_{x \in B \cap \mathbb{Z}^d} \left[\Lambda_R(\theta_1(x))^2 \cdots \Lambda_R(\theta_t(x))^2 \right] = (1 + o_{d,t}(1)) \left(\frac{W \log R}{\varphi(W)} \right)^t. \tag{1.6}$$

Proposition 1.16. *Let m be a positive integer and let I be an interval of length at least R^{10m}. Suppose that $h_1, \dots, h_m$ are distinct integers such that $|h_i| \le N^2$, and set*

$$\Delta := \prod_{1 \le i < j \le m} |h_i - h_j|.$$

Then for sufficiently large $N = N(m)$ we have

$$\begin{aligned} &\mathbb{E}_{x \in I} \left[\Lambda_R(W(x + h_1) + 1)^2 \cdots \Lambda_R(W(x + h_m 1) + 1)^2 \right] \\ &\qquad \le (1 + o_m(1)) \left(\frac{W \log R}{\varphi(W)} \right)^m \prod_{p \mid \Delta} (1 + O_m(p^{-1/2})). \end{aligned} \tag{1.7}$$

Lemma 1.17. *Let $f := k^{-1} 2^{-k-5} \tilde{\Lambda}$ as before. Then $0 \le f(x) \le \nu(x)$ for all $x \in [\epsilon_k N, 2\epsilon_k N]$ if N is sufficiently large.*

Proof. If $Wn + 1$ is not prime, the claim is trivial. If N is sufficiently large, we may assume that $Wn + 1 > R$ for all $n \in [\epsilon_k N, 2\epsilon_k N]$. The truncated divisor sum corresponding to $\Lambda_R(Wn+1)$ then only contains the summand $d = 1$, so that $\Lambda_R(Wn+1) = \log R$. This implies

$$\nu(n) = \frac{\varphi(W)}{W} \log R \ge k^{-1} 2^{-k-5} \tilde{\Lambda}(n) = f(n)$$

for sufficiently large N and thus the claim. □

Lemma 1.18. *ν defines a measure.*

Proof. We can apply Proposition 1.15 with $d = t = 1$, $\psi_1(x) = x$ as well as $B = [\epsilon_k N, 2\epsilon_k N]$. If N is large enough, B satisfies the assumptions of that proposition and we obtain

$$\mathbb{E}_{x \in [\epsilon_k N, 2\epsilon_k N]} \left[\Lambda_R(Wx + 1)^2 \right] = (1 + o(1)) \frac{W \log R}{\varphi(W)}.$$

The claim follows immediately from the definition of ν. □

Lemma 1.19. *ν satisfies the $(k2^{k-1}, 3k-4, k)$-linear forms condition.*

Proof. Clearly, we want to apply Proposition 1.15. But due to the piecewise definition, we need to make some technical adjustments to be able to do that.

To this end, let $Q = Q(N)$ be a function slowly growing with N to be chosen later, and partition $\mathbb{Z}_N^d$ into Q^d boxes

$$B_{u_1,\dots,u_d} := \big\{x \in \mathbb{Z}_N^d : x_j \in \big[\lfloor u_j N/Q\rfloor, \lfloor (u_j+1)N/Q\rfloor\big), j = 1,\dots,d\big\},$$

where we may view the u_j as elements of $\mathbb{Z}_Q$. Note that the sizes of these boxes might differ up to at most 1 in every direction. We can thus rewrite

$$\mathbb{E}_{x\in\mathbb{Z}_N^d}[\nu(\psi_1(x))\cdots\nu(\psi_t(x))] = (1+o(1))\mathbb{E}_{u_1,\dots,u_d\in\mathbb{Z}_Q}\big[\mathbb{E}_{x\in B_{u_1,\dots,u_d}}[\nu(\psi_1(x))\cdots\nu(\psi_t(x))]\big].$$

If for every $i = 1,\dots,t$ we have that $\psi_i(B_{u_1,\dots,u_d})$ is either contained in or disjoint from $[\epsilon_k N, 2\epsilon_k N]$ then we can already say that

$$\mathbb{E}_{x\in B_{u_1,\dots,u_d}}[\nu(\psi_1(x))\cdots,\nu(\psi_t(x))] = 1+o(1),$$

due to Proposition 1.15 if Q grows slow enough that N/Q exceeds R^{10d}, noting that ν is identically 1 in the second case (regarding dependencies of implicit constants, note that d and t are bounded by functions of k only).

If there is i such that $\psi_i(B_{u_1,\dots,u_d})$ is not contained in either of the sets, then we can still trivially bound ν by $1+\frac{\varphi(W)}{W\log R}\Lambda_R^2(\theta_i(x))$, multiply out and apply Proposition 1.15 to obtain that

$$\mathbb{E}_{x\in B_{u_1,\dots,u_d}}[\nu(\psi_1(x))\cdots\nu(\psi_t(x))] = O(1).$$

It now suffices to show that the proportion of such tuples $(u_1,\dots,u_d)$ is $O(1/Q)$ to obtain the claim. To this end, let $1 \le i \le t$ as well as $x, x' \in B_{u_1,\dots,u_d}$ such that $\psi_i(x) \in [\epsilon_k N, 2\epsilon_k N]$, but $\psi_i(x')$ is not. We have

$$\psi_i(x) = \sum_{j=1}^d L_{ij}\lfloor Nu_j/Q\rfloor + b_i + O(N/Q)$$

and the same for $\psi_i(x')$. Thus, we either have

$$\epsilon_k N = \sum_{j=1}^d L_{ij}\lfloor Nu_j/Q\rfloor + b_i + O(N/Q)$$

or the same equation holds with $2\epsilon_k N$ (we assume the first). Dividing by N/Q gives

$$\sum_{j=1}^d L_{ij}u_j = \epsilon_k Q - b_i Q/N + O(1).$$

But we assumed that $(L_{ij})_j$ is not identically zero, hence this equation can only be satisfied by $O(Q^{d-1})$ values of $(u_1,\dots,u_d)$. This completes the proof of the linear forms condition. □

To verify the correlation condition, we of course aim to apply Proposition 1.16. To this end, we first need a Lemma in order to deal with the product term appearing on its right-hand side.

Lemma 1.20. *For any $m \geq 1$, there is a weight function $\tau = \tau_m : \mathbb{Z} \to \mathbb{R}^+$ such that $\tau(n) \geq 1$ for all $n \neq 0$ and such that for any distinct $h_1, \dots, h_m \in [\epsilon_k N, 2\epsilon_k N]$ we have*

$$\prod_{p \mid \Delta} (1 + O_m(p^{-1/2})) \leq \sum_{1 \leq i < j \leq m} \tau(h_i - h_j) \tag{1.8}$$

with notation as in Proposition 1.16, and such that

$$\mathbb{E}_{0<|n|\leq N}[\tau^q(n)] = O_{m,q}(1)$$

for $1 \leq q < \infty$.

Proof. Note that we have

$$\prod_{p \mid \Delta} (1+O_m(p^{-1/2})) \leq \prod_{1 \leq i < j \leq m} \prod_{p \mid |h_i - h_j|} (1+O_m(p^{-1/2})) \leq \prod_{1 \leq i < j \leq m} \Bigg(\prod_{p \mid h_i - h_j} (1 + p^{-1/2}) \Bigg)^{O_m(1)}.$$

The inequality of arithmetic and geometric mean implies that this (after rewriting) is bounded by

$$\Bigg(\prod_{1 \leq i < j \leq m} \Bigg(\prod_{p \mid h_i - h_j} (1 + p^{-1/2}) \Bigg)^{O_m(1)} \Bigg)^{1/\binom{m}{2}} \leq O_m(1) \sum_{1 \leq i < j \leq m} \Bigg(\prod_{p \mid h_i - h_j} (1 + p^{-1/2}) \Bigg)^{O_m(1)}.$$

Setting

$$\tau_m(n) \coloneqq O_m(1) \Big(\prod_{p \mid n} (1 + p^{-1/2}) \Big)^{O_m(1)}$$

(with the constants as in the last bound) implies the Lemma if we can show that for $1 \leq q < \infty$, we have

$$\mathbb{E}_{0<|n|\leq N} \Big[\Big(\prod_{p \mid n} (1 + p^{-1/2}) \Big)^{O_m(q)} \Big] = O_{m,q}(1).$$

One quickly verifies that for all but $O_{m,q}(1)$ values of p we can bound $(1+p^{-1/2})^{O_m(q)}$ by $1 + p^{-1/4}$. Moreover, we have $\prod_{p \mid n}(1 + p^{-1/4}) \leq \sum_{d \mid n} d^{-1/4}$. Hence, we obtain that

$$\begin{aligned} \mathbb{E}_{0<|n|\leq N} \Big[\Big(\prod_{p \mid n} (1 + p^{-1/2}) \Big)^{O_m(q)} \Big] &= O_{m,q}(1) \mathbb{E}_{0<|n|\leq N} \Big[\sum_{d \mid n} d^{-1/4} \Big] \\ &= O_{m,q}(1) \frac{1}{2N} \sum_{d=1}^{N} \frac{N}{d} d^{-1/4} = O_{m,q}(1), \end{aligned}$$

as we wanted. □

We are now in a position to verify the correlation condition, which completes the proof of Proposition 1.12.

Lemma 1.21. *ν satisfies the 2^{k-1}-correlation condition.*

Proof. Our goal is to verify that for all $1 \le m \le 2^{k-1}$ and $h_1, \dots, h_m \in \mathbb{Z}_N$, we have

$$\mathbb{E}_{x \in \mathbb{Z}_N}[\nu(x+h_1)\cdots\nu(x+h_m)] \le \sum_{1 \le i < j \le m} \tau(h_i - h_j), \tag{1.9}$$

for some weight function $\tau = \tau_m$ satisfying $\mathbb{E}[\tau^q] = O_{m,q}(1)$. Let τ be the weight function from Lemma 1.20, identifying $\mathbb{Z}_N$ with $(-N/2, N/2]$. Note that we can still set the value of $\tau(0)$ in any way such that the moment condition is still satisfied, so we can define

$$\tau(0) := \exp(Cm \log N / \log\log N)$$

for some absolute constant $C > 0$.

First assume that there are two values of h_i which coincide. In that case, we can trivially bound the left-hand side of (1.9) by $\|\nu\|_{L^\infty}^m$, and looking at the definition of ν, we can bound

$$\|\nu\|_{L^\infty} \ll \log N \big(\max_{n \le N} \tau(n) \big)^2,$$

and because it is well-known that the maximal order of $\log \tau(n)$ is $\log 2 \log n / \log\log n$, we see that for sufficiently large value of C,

$$\|\nu\|_{L^\infty}^m \le \tau(0).$$

This gives the claim in the case that two h_i are equal.

Now assume that all the h_i are distinct. Define

$$g(n) := \frac{\varphi(W)}{W} \frac{\Lambda_R^2(Wn+1)}{\log R} \mathbb{1}_{[\epsilon_k N, 2\epsilon_k N]}(n).$$

Then we can easily bound

$$\mathbb{E}_{x \in \mathbb{Z}_N}[\nu(x+h_1)\cdots\nu(x+h_m)] \le \mathbb{E}_{x \in \mathbb{Z}_N}[(1+g(x+h_1))\cdots(1+g(x+h_m))].$$

Expanding the right-hand side gives

$$\sum_{A \subseteq \{1,\dots,m\}} \mathbb{E}_{x \in \mathbb{Z}_N}\Big[\prod_{i \in A} g(x+h_i) \Big].$$

Applying Proposition 1.16 together with Lemma 1.20, we have

$$\mathbb{E}_{x \in \mathbb{Z}_N}\Big[\prod_{i \in A} g(x+h_i) \Big] \le (1 + o_m(1)) \sum_{1 \le i < j \le m} \tau(h_i - h_j)$$

and multiplying τ by $O_m(1)$, the claim follows. □

1.5 The Gowers norm

Definition 1.22. *Let $d \geq 1$, and $f : \mathbb{Z}_N \to \mathbb{C}$ be a function. For $t \in \mathbb{Z}_N$, we define the translated function $f_t : \mathbb{Z}_N \to \mathbb{C}$,*

$$f_t(x) := f(x+t).$$

Moreover, we define the Gowers 1-norm of f by

$$\|f\|_{U^1(\mathbb{Z}_N)} := |\mathbb{E}_{x \in \mathbb{Z}_N}[f(x)]|$$

and then the Gowers d-norm inductively by

$$\|f\|_{U^d(\mathbb{Z}_N)} := \Big(\mathbb{E}_{t \in \mathbb{Z}_N}\big[\|f \cdot \overline{f_t}\|^{2^{d-1}}_{U^{d-1}(\mathbb{Z}_N)}\big]\Big)^{1/2^d}.$$

Note that the Gowers 1-norm is not actually a norm.

We will, for most parts, only deal with real-valued functions, where we can ignore the complex conjugation. We note that almost all results we present on Gowers norms also hold in the complex-valued case, but we will not actually need to deal with complex functions in applications and it is notationally easier to deal with the real-valued case.

We will moreover shorten $\|f\|_{U^s[N]} := \|f\|_{U^s([N])}$.

Let $d \geq 1$. We will parametrise the d-dimensional cube with side lengths $h_1, \dots, h_d$ by $(\omega \cdot h)_{\omega \in \{0,1\}^d}$, where $\omega \cdot h := \omega_1 h_1 + \cdots + \omega_d h_d$. We can then define the Gowers inner product of complex-valued functions $(f_\omega)_{\omega \in \{0,1\}^d}$ on $\mathbb{Z}_N$ by

$$\langle (f_\omega)_{\omega \in \{0,1\}^d} \rangle_{U^d} := \mathbb{E}_{x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^d}\Big[\prod_{\omega \in \{0,1\}^d} \mathcal{C}^{|\omega|} f_\omega(x + \omega \cdot h)\Big] \tag{1.10}$$

with $|\omega| = \omega_1 + \cdots + \omega_d$ and where $\mathcal{C}$ denotes complex conjugation.

Lemma 1.23. *The Gowers norm of a function $f : \mathbb{Z}_N \to \mathbb{C}$ can be explicitly written as*

$$\|f\|_{U^d(\mathbb{Z}_N)} = \langle (f)_\omega \rangle^{1/2^d}_{U^d} = \mathbb{E}_{x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^d}\Big[\prod_{\omega \in \{0,1\}^d} \mathcal{C}^{|\omega|} f(x + \omega \cdot h)\Big]^{1/2^d}. \tag{1.11}$$

We leave this as an exercise to the reader; it is a simple induction argument.

Definition 1.24. *In the same spirit, we can define the U^d-norm of a function $f : \mathbb{Z} \to \mathbb{C}$ on some finite subset $A \subset \mathbb{Z}$ by*

$$\|f\|_{U^d(A)} := \mathbb{E}_{\substack{x \in \mathbb{Z}, h \in \mathbb{Z}^d: \\ x + \omega \cdot h \in A \, \forall \omega \in \{0,1\}^d}}\Big[\prod_{\omega \in \{0,1\}^d} \mathcal{C}^{|\omega|} f(x + \omega \cdot h)\Big]^{1/2^d}. \tag{1.12}$$

Example 1.25. Let $d = 2$. For functions $f_{00}, f_{10}, f_{01}, f_{11} : \mathbb{Z}_N \to \mathbb{R}$ we obtain

$$\langle f_{00}, f_{10}, f_{01}, f_{11} \rangle_{U^2} = \mathbb{E}_{x, h_1, h_2 \in \mathbb{Z}_N}[f_{00}(x) f_{10}(x + h_1) f_{01}(x + h_2) f_{11}(x + h_1 + h_2)].$$

The Gowers 2-norm of a function f is thus given by

$$\|f\|_{U^2(\mathbb{Z}_N)} = \mathbb{E}_{x,h_1,h_2\in\mathbb{Z}_N}[f(x)f(x+h_1)f(x+h_2)f(x+h_1+h_2)]^{1/4}.$$

It is not very hard to show that in fact we have

$$\|f\|_{U^2(\mathbb{Z}_N)} = \|\hat{f}\|_{L^4}.$$

For a more specific example, define

$$f(x) := e\left(\frac{ax^2+bx+c}{N}\right).$$

Then one verifies that

$$\|f\|_{U^3(\mathbb{Z}_N)} = 1.$$

This is essentially because we apply Weyl differencing to the argument three times, giving the constant phase 0. The analogous result holds for the $U^{d+1}(\mathbb{Z}_N)$-norm when replacing the degree 2 polynomial by one of degree d.

Lemma 1.26. *The Gowers norm is non-negative and, for all $d \geq 2$, in fact defines a norm.*

Proof. We have

$$\begin{aligned}\langle (f)_\omega\rangle_{U^d} &= \mathbb{E}_{x\in\mathbb{Z}_N,h'\in\mathbb{Z}_N^{d-1}h_d\in\mathbb{Z}_N}\Big[\prod_{\omega'\in\{0,1\}^{d-1}} f(x+\omega'\cdot h')f(x+\omega'\cdot h'+h_d)\Big]\\ &= \mathbb{E}_{h'\in\mathbb{Z}_N^{d-1}}\left[\left|\mathbb{E}_{y\in\mathbb{Z}_N}\Big[\prod_{\omega'\in\{0,1\}^{d-1}} f(y+\omega'\cdot h')\Big]\right|^2\right]\end{aligned}$$

which implies the non-negativity.

Taking general inner products with the same idea, we see that

$$\begin{aligned}\langle (f_\omega)_{\omega\in\{0,1\}^d}\rangle_{U^d} = \mathbb{E}_{h'\in\mathbb{Z}_N^{d-1}}\Bigg[&\mathbb{E}_{y\in\mathbb{Z}_N}\Big[\prod_{\omega'\in\{0,1\}^{d-1}} f_{(\omega',0)}(y+\omega'\cdot h')\Big]\\ &\mathbb{E}_{y\in\mathbb{Z}_N}\Big[\prod_{\omega'\in\{0,1\}^{d-1}} f_{(\omega',1)}(y+\omega'\cdot h')\Big]\Bigg]\end{aligned}$$

Cauchy-Schwarz inequality with respect to h' gives

$$|\langle (f_\omega)_{\omega\in\{0,1\}^d}\rangle_{U^d}| \leq \langle (f_{(\omega',0)})_{\omega\in\{0,1\}^d}\rangle_{U^d}^{1/2}\langle (f_{(\omega',1)})_{\omega\in\{0,1\}^d}\rangle_{U^d}^{1/2}$$

and inductively we obtain the Gowers-Cauchy-Schwarz inequality

$$|\langle (f_\omega)_{\omega\in\{0,1\}^d}\rangle_{U^d}| \leq \prod_{\omega\in\{0,1\}^d}\|f_\omega\|_{U^d}. \tag{1.13}$$

The triangle inequality is now a direct consequence of this together with the binomial formula: We have

$$|\langle (f+g)_\omega \rangle_{U^d}| \leq (\|f\|_{U^d} + \|g\|_{U^d})^{2^d}$$

which immediately resolves this claim. The $\mathbb{R}$-linearity is clear and thus it suffices to show the definiteness of the Gowers norm. To do this, we show the definiteness of the Gowers 2-norm and then prove that $\|f\|_{U^{d-1}} \leq \|f\|_{U^d}$ for all $d \geq 2$.

To see the definiteness of the Gowers 2-norm, we can apply the Gowers-Cauchy-Schwarz inequality with $f, \delta_a, \delta_b, \delta_c$ to obtain

$$\mathbb{E}_{x,h_1,h_2 \in \mathbb{Z}_N}[f(x)\delta_a(x+h_1)\delta_b(x+h_2)\delta_c(x+h_1+h_2)] \leq \|f\|_{U^2} N^{-C}$$

(where $C = 9/4$, but that is not important here). If $\|f\|_{U^2} = 0$, varying a, b and c over $\mathbb{Z}_N$ one sees that $f \equiv 0$ and thus $\|\cdot\|_{U^2}$ is definite.

We can now turn to the monotonicity property. Let $f : \mathbb{Z}_N \to \mathbb{R}$ be a function and let $f_\omega := 1$ when $\omega_d = 1$ and $f_\omega := f$ when $\omega_d = 0$. From the definition of the Gowers norm and from the Gowers-Cauchy-Schwarz inequality we obtain

$$\|f\|_{U^{d-1}}^{2^{d-1}} = |\langle (f_\omega)_{\omega \in \{0,1\}^d} \rangle_{U^d}| \leq \|f\|_{U^d}^{2^{d-1}}.$$

This implies the claim. □

We will sometimes need to transition between the $U^d(\mathbb{Z}_{N'})$- and $U^d[N]$-norms, typically for some prime $N' \in [CN, 2CN]$. The fundamental Lemma which allows us to do that is the following.

Lemma 1.27. *Let $N' \geq 1$ be an integer, and let $0 < \alpha \leq 1/2$. Let $I = [a, b]$ be an interval of integers which satisfies $\alpha N' \leq |I| \leq N'/2$. Let $f : I \to \mathbb{R}$ be a function on I, and $\tilde{f} : \mathbb{Z}_{N'} \to \mathbb{R}$ be the function obtained from f by identifying I with the corresponding image in $\mathbb{Z}_{N'}$ and setting it 0 otherwise. Then we have*

$$\|\tilde{f}\|_{U^d(\mathbb{Z}_{N'})} = c\|f\|_{U^d(I)} \tag{1.14}$$

for some $c = c(I, N', d)$ independent of f and bounded from above and below by $c' = c'(\alpha, d)$ uniformly over I and N'. In particular, in the above scenario when $N' \in [CN, 2CN]$ for some constant $C > 0$, the norms $U^d(\mathbb{Z}_{N'})$ and $U^d[N]$ are equivalent and the constants in both directions can be chosen independent of N (and N').

We will not proceed to prove this, the proof is technical and not very interesting for us.

Lemma 1.28. *Let ν be a k-pseudorandom measure. Then we have*

$$\|\nu - 1\|_{U^d} = o(1) \tag{1.15}$$

for all $1 \leq d \leq k-1$.

Proof. It suffices to prove the case $d = k-1$ by monotonicity. By definition of the Gowers norm, the claim follows if we can show that

$$\mathbb{E}_{x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1}}\left[\prod_{\omega\in\{0,1\}^{k-1}}(\nu(x+\omega\cdot h)-1)\right] = o(1).$$

The assertion will now follow by expanding the product and applying the linear forms condition to each part: The left-hand side coincides with

$$\sum_{A\subseteq\{0,1\}^{k-1}}(-1)^{|A|}\mathbb{E}_{x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1}}\left[\prod_{\omega\in A}\nu(x+\omega\cdot h)\right]. \tag{1.16}$$

Looking at the forms $(x, h_1, \dots, h_{k-1}) \mapsto x + \omega\cdot h$ as ω ranges over $\{0,1\}^{k-1}$, one may easily check that none of them are rational multiples of each other (or identically zero) and we can thus apply the linear forms condition with parameters $(2^{k-1}, k, 1)$ to conclude that each expectation is $1 + o(1)$. The claim then follows from the binomial formula. □

We can now state what was coined the "generalised von Neumann theorem" by Green and Tao. It is a very important manifestation of the transference principle, in the sense that it is a statement regarding functions bounded by pseudorandom measures which was previously known only for bounded functions (see [4, Theorem 3.2]). We will only prove the theorem in a special case, but note that the general strategy is identical; it is merely the much heavier notation which prevents us from proving the general statement.

Proposition 1.29. *Let $N(>2)$ be a prime, $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ a k-pseudorandom measure, and let $f_1, \dots, f_{k-1} : \mathbb{Z}_N \to \mathbb{R}$ be functions which satisfy*

$$|f_j(x)| \le \nu(x) + 1 \text{ for all } x \in \mathbb{Z}_N, 0 \le j \le k-1. \tag{1.17}$$

Let $c_0, \dots, c_{k-1}$ be a permutation of k consecutive elements of $\{-k+1, \dots, 0, \dots, k-1\}$. Then

$$\mathbb{E}_{x,r\in\mathbb{Z}_N}\left[\prod_{j=0}^{k-1} f_j(x + c_j r)\right] = O\left(\inf_{0\le j\le k-1}\|f_j\|_{U^{k-1}(\mathbb{Z}_N)}\right) + o(1). \tag{1.18}$$

For the proof of this statement, we first need a small Lemma, for which we only sketch the proof; the details are very simple to fill in.

Lemma 1.30. *Let ν be a k-pseudorandom measure. Then $\nu_{1/2} := (\nu+1)/2$ is also k-pseudorandom.*

Proof. Clearly $\nu_{1/2}$ defines a measure. The other conditions may easily be checked by plugging in $\nu_{1/2}$ in the required equations and expanding the product. In the linear forms condition, one obtains an average of 2^m terms, each of which is $1 + o_{t_0,d_0,L_0}$ and thus, their average satisfies this as well. In a similar spirit, the correlation condition easily follows by expanding the corresponding product. □

Proof of the case $k = 3$. Note that the conclusion does not depend on the pseudorandom measure ν. Thus, if we are able to show the statement for functions f with the property

$$|f_j(x)| \leq \nu(x)$$

then, applying it to $\nu_{1/2}$ in view of Lemma 1.30 gives the above statement. Hence, we may assume this. By reordering (and shifting), we may furthermore assume that f_0 has the smallest Gowers norm and that $c_j = j$.

We will now prove the case $k = 3$. Our goal is to prove

$$\mathbb{E}_{x,r\in\mathbb{Z}_N}\left[f_0(x)f_1(x+r)f_2(x+2r)\right] = O(\|f_0\|_{U^2(\mathbb{Z}_N)}) + o(1). \tag{1.19}$$

To this end, it is convenient to reparametrise the arithmetic progression $(x, x+r, x+2r)$ as $(2y_1 + 2y_2, y_2, -2y_1)$, so that the second term does not depend on y_1 and the third term is independent of y_2.

We are thus interested in bounding the quantity

$$J_0 := \mathbb{E}_{y_1,y_2\in\mathbb{Z}_N}\left[f_0(2y_1 + 2y_2)f_1(y_2)f_2(-2y_1)\right].$$

From the assumption on f_2 we deduce

$$|J_0| \leq \mathbb{E}_{y_1\in\mathbb{Z}_N}\left[|\mathbb{E}_{y_2\in\mathbb{Z}_N}[f_0(2y_1 + 2y_2)f_1(y_2)]|\nu(-2y_1)\right],$$

so that Cauchy-Schwarz inequality together with the measure property of ν implies

$$|J_0| \leq (1 + o(1))\mathbb{E}_{y_1\in\mathbb{Z}_N}\left[|\mathbb{E}_{y_1\in\mathbb{Z}_N}[f_0(2y_1 + 2y_2)f_1(y_2)]|^2\right]^{1/2} = (1 + o(1))J_1^{1/2},$$

where we set

$$J_1 := \mathbb{E}_{y_1,y_2,y_2'\in\mathbb{Z}_N}\left[f_0(2y_1 + 2y_2)f_0(2y_1 + 2y_2')f_1(y_2)f_1(y_2')\nu(-2y_1)\right].$$

This time bounding f_1 by ν, we now get

$$J_1 \leq \mathbb{E}_{y_2,y_2'\in\mathbb{Z}_N}\left[|\mathbb{E}_{y_1\in\mathbb{Z}_N}[f_0(2y_1 + 2y_2)f_0(2y_1 + 2y_2')\nu(-2y_1)]|\nu(y_2)\nu(y_2')\right].$$

Cauchy-Schwarz therefore implies

$$J_1 \leq (1 + o(1))\mathbb{E}_{y_2,y_2'\in\mathbb{Z}_N}\left[|\mathbb{E}_{y_1\in\mathbb{Z}_N}[f_0(2y_1 + 2y_2)f_0(2y_1 + 2y_2')\nu(-2y_1)]|^2\nu(y_2)\nu(y_2')\right]^{1/2}$$

Hence we have

$$|J_0| \leq (1 + o(1))J_2^{1/4},$$

where we define

$$J_2 := \mathbb{E}_{y_1,y_1',y_2,y_2'\in\mathbb{Z}_N}\left[f_0(2y_1 + 2y_2)f_0(2y_1 + 2y_2')f_0(2y_1' + 2y_2)f_0(2y_1' + 2y_2')\right.$$
$$\left.\nu(-2y_1)\nu(-2y_1')\nu(y_2)\nu(y_2')\right].$$

Reparametrising $(2y_1+2y_2, 2y_1'+2y_2, 2y_1+2y_2', 2y_1'+2y_2')$ by $(2x, 2(x+h_1), 2(x+h_2), 2(x+h_1+h_2))$, we can write

$$J_2 = \mathbb{E}_{x,h_1,h_2\in\mathbb{Z}_N}\left[f_0(2x)f_0(2(x+h_1))f_0(2(x+h_2))f_0(2(x+h_1+h_2))W(x,h_1,h_2)\right]$$

with

$$W(x,h_1,h_2) = \mathbb{E}_{y\in\mathbb{Z}_N}\left[\nu(-2y)\nu(-2y-2h_1)\nu(x-y)\nu(x-y+h_2)\right].$$

Note that, since $N>2$ is a prime, we have

$$\|f_0\|_{U^2(\mathbb{Z}_N)}^4 = \mathbb{E}_{x,h_1,h_2\in\mathbb{Z}_N}\left[f_0(2x)f_0(2(x+h_1))f_0(2(x+h_2))f_0(2(x+h_1+h_2))\right],$$

so that it suffices to show that

$$\mathbb{E}_{x,h_1,h_2\in\mathbb{Z}_N}\left[f_0(2x)f_0(2(x+h_1))f_0(2(x+h_2))f_0(2(x+h_1+h_2))(W(x,h_1,h_2)-1)\right] = o(1).$$

Again bounding f_0 by ν and applying Cauchy-Schwarz, we see that it suffices to verify the following two equations:

$$\mathbb{E}_{x,h_1,h_2\in\mathbb{Z}_N}\left[\nu(2x)\nu(2(x+h_1))\nu(2(x+h_2))\nu(2(x+h_1+h_2))(W(x,h_1,h_2)-1)^2\right] = o(1),$$
$$\mathbb{E}_{x,h_1,h_2\in\mathbb{Z}_N}\left[\nu(2x)\nu(2(x+h_1))\nu(2(x+h_2))\nu(2(x+h_1+h_2))\right] = 1+o(1).$$

After expanding the $W-1$ term in the first equation, one sees that to prove these two equations, it suffices to show

$$\mathbb{E}_{x,h_1,h_2}\left[\nu(2x)\nu(2(x+h_1))\nu(2(x+h_2))\nu(2(x+h_1+h_2))W(x,h_1,h_2)^q\right] = 1+o(1)$$

holds for $q=0,1,2$. But this is a direct consequence of the linear forms condition. □

The reader may verify that for complex-valued functions, the proof works just out in the way we want. For example when defining J_1, we need a complex conjugation precisely over those terms containing y_2'. Analogously, in the definition of y_2 the number of complex conjugations of each term is exactly the number of y_i'-variables contained in that expression.

1.6 Gowers anti-uniformity

Definition 1.31. *We introduce the Gowers dual norm of a function $g:\mathbb{Z}_N\to\mathbb{R}$ given by*

$$\|g\|_{U_*^{k-1}(\mathbb{Z}_N)} := \sup\{|\langle f,g\rangle| : f:\mathbb{Z}_N\to\mathbb{R}, \|f\|_{U^{k-1}(\mathbb{Z}_N)}\le 1\}. \tag{1.20}$$

Example 1.32. If $k=3$, we have $\|f\|_{U^2(\mathbb{Z}_N)} = \|\hat{f}\|_{L^4}$, which quickly implies

$$\|g\|_{U_*^2(\mathbb{Z}_N)} = \|\hat{g}\|_{L^{4/3}}$$

using Parseval's identity and the duality of the L^4- and $L^{4/3}$-norm.

It is immediate from the definition that for any $f, g : \mathbb{Z}_N \to \mathbb{R}$ we have

$$|\langle f, g\rangle| \leq \|f\|_{U^{k-1}(\mathbb{Z}_N)} \|g\|_{U_*^{k-1}(\mathbb{Z}_N)}.$$

We informally call a function f Gowers-uniform if its Gowers norm is small, and a function g Gowers anti-uniform if its Gowers dual norm is not too large. From the above inequality, we see that a function f which correlates with such a Gowers anti-uniform function (meaning that $|\langle f, g\rangle|$ is not too small) can not be Gowers-uniform. We will soon make this more precise, and we will see why it is important.

Definition 1.33. *For a function $F : \mathbb{Z}_N \to \mathbb{R}$, we define the Gowers dual function $\mathcal{D}F : \mathbb{Z}_N \to \mathbb{R}$ by*

$$\mathcal{D}F(x) := \mathbb{E}_{h \in \mathbb{Z}_N^{k-1}} \Big[\prod_{\substack{\omega \in \{0,1\}^{k-1} \\ \omega \neq 0}} F(x + \omega \cdot h) \Big]. \tag{1.21}$$

The Gowers dual of a function F which is pointwise bounded by $\nu + 1$ for some pseudorandom measure ν will be called a basic Gowers anti-uniform function.

Lemma 1.34. *Let ν be a k-pseudorandom measure and let $F : \mathbb{Z}_N \to \mathbb{R}$ be a function. Then we have*

$$\langle F, \mathcal{D}F\rangle = \|F\|_{U^{k-1}(\mathbb{Z}_N)}^{2^{k-1}} \tag{1.22}$$

and

$$\|\mathcal{D}F\|_{U_*^{k-1}(\mathbb{Z}_N)} = \|F\|_{U^{k-1}(\mathbb{Z}_N)}^{2^{k-1}-1}. \tag{1.23}$$

If we moreover have $|F(x)| \leq \nu(x) + 1$ for all $x \in \mathbb{Z}_N$, i.e. $\mathcal{D}F$ is a basic Gowers anti-uniform function, then we have

$$\|\mathcal{D}F\|_{L^\infty} \leq 2^{2^{k-1}-1} + o(1). \tag{1.24}$$

This Lemma, despite having an elementary proof, is fundamental in the approach of Green-Tao. We will also try to express it in the informal language we established above. Let us assume that F is a function which is bounded by $\nu + 1$ for some pseudorandom measure ν, as is assumed in the last claim. If F is not Gowers-uniform, then we see from the first claim that F correlates with its Gowers-dual function, and by the third claim we have that this function is in fact bounded. Moreover, the second assertion gives us a way to calculate the Gowers dual norm of its Gowers dual function, which turns out not to be too large.

Proof. The first part is very simple from the definitions:

$$\langle F, \mathcal{D}F\rangle = \mathbb{E}_{x \in \mathbb{Z}_N}[F(x)\mathcal{D}F(x)] = \mathbb{E}_{x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1}} \Big[\prod_{\omega \in \{0,1\}^{k-1}} F(x + \omega \cdot h) \Big] = \|F\|_{U^{k-1}(\mathbb{Z}_N)}^{2^{k-1}}.$$

For the second part, we may assume that F is not identically zero. By the first equation, we have

$$\|\mathcal{D}F\|_{U_*^{k-1}(\mathbb{Z}_N)} \geq \Big| \Big\langle \frac{F}{\|F\|_{U^{k-1}(\mathbb{Z}_N)}}, \mathcal{D}F \Big\rangle \Big| = \|F\|_{U^{k-1}(\mathbb{Z}_N)}^{2^{k-1}-1}.$$

To see the converse direction, let f be any function, and define $f_\omega := f$ when $\omega = 0$ and $f_\omega := F$ otherwise. Then by the definition of the Gowers inner product and by the Gowers-Cauchy-Schwarz inequality, we see that

$$|\langle f, \mathcal{D}F\rangle| = |\langle (f_\omega)_{\omega\in\{0,1\}^{k-1}}\rangle_{U^{k-1}(\mathbb{Z}_N)}| \leq \|f\|_{U^{k-1}}\|F\|_{U^{k-1}(\mathbb{Z}_N)}^{2^{k-1}-1}.$$

The definition of the Gowers dual norm implies

$$\|\mathcal{D}F\|_{U_*^{k-1}(\mathbb{Z}_N)} \leq \|F\|_{U^{k-1}(\mathbb{Z}_N)}^{2^{k-1}-1}$$

and thus the claim.

The last part is a quick consequence of the linear forms condition. Since $|F| \leq 2\nu_{1/2}$, the definition of $\mathcal{D}$ implies that it suffices to show that

$$\mathcal{D}\nu_{1/2}(x) \leq 1 + o(1)$$

uniformly over $x \in \mathbb{Z}_N$. But we have

$$\mathcal{D}\nu_{1/2}(x) = \mathbb{E}_{h\in\mathbb{Z}_N^{k-1}}\Bigg[\prod_{\substack{\omega\in\{0,1\}^{k-1}\\ \omega\neq 0}} \nu_{1/2}(x+\omega\cdot h)\Bigg],$$

so that the linear forms condition of $\nu_{1/2}$ gives the assertion. □

Recall Example 1.25; one verifies that if F is of the form $e(P(x)/N)$ for some polynomial P of degree at most $k-2$, then $\mathcal{D}F = F$. Such polynomial phases are in some sense prime examples of Gowers anti-uniform functions, and in the $k = 3$ case (but not for higher k) they are essentially the only examples of such functions, as one can see using the Fourier-analytic identities of the Gowers 2-norm that we have stated. For larger k, the Gowers norms do not seem to have any Fourier-analytic correspondences; Green and Tao coined the techniques and ideas involving higher-order Gowers norms the higher-order Fourier analysis.

We have defined the basic Gowers anti-uniform functions. This term already suggests that we want to talk about more general types of such functions, namely the algebra generated by them. This will be considered in the following proposition, which tells us that $\nu - 1$ (an object that is for example small in L^1-norm by the measure property) does not correlate with any polynomial in Gowers anti-uniform functions, not even with any continuous function Φ evaluated in these anti-uniform functions.

Before stating the Proposition, we need a small

Definition 1.35. *Let A, B be finite non-empty sets, and $u : A \to B$ be a map. We say that u is a uniform covering of B by A if the inverse images $u^{-1}(b)$ have the same cardinality $|B|/|A|$ for any $b \in B$. The fundamental property of a uniform covering is that for any map $f : B \to \mathbb{R}$ we have*

$$\mathbb{E}_{a\in A}[f(u(a))] = \mathbb{E}_{b\in B}[f(b)].$$

Proposition 1.36. *Let ν be a k-pseudorandom measure. Define $I := [-2^{2^{k-1}}, 2^{2^{k-1}}]$, and let $K \in \mathbb{N}$ and $\Phi : I^K \to \mathbb{R}$ be a fixed continuous function. Moreover, let $\mathcal{D}F_1, \cdots, \mathcal{D}F_K : \mathbb{Z}_N \to I$ be basic Gowers anti-uniform functions, and define $\psi : \mathbb{Z}_N \to \mathbb{R}$,*

$$\psi(x) := \Phi(\mathcal{D}F_1(x), \dots, \mathcal{D}F_K(x)).$$

Then we have

$$\langle \nu - 1, \psi \rangle = o_{K,\Phi}(1). \tag{1.25}$$

Moreover, if Φ ranges over a compact set $E \subset C(I^K)$ (with the topology induced by the supremum norm) then the bounds are uniform w.r.t Φ, i.e. the left-hand side is $o_{K,E}(1)$.

Proof. The general idea is first to show the theorem when Φ is a polynomial and then use the Weierstrass approximation theorem for the general case. In a similar way as we have done before, we may assume

$$|F_j(x)| \le \nu(x), \qquad 1 \le j \le K.$$

To see that this implies the statement, just apply it with this stronger assumption here to $\nu_{1/2}$, and the number of extra factors two only depends on K.

Our goal now is to show the following

Lemma 1.37. *Let $d \in \mathbb{N}$, and let $P : I^K \to \mathbb{R}$ be a polynomial of degree d with real coefficients independent of N. Then we have*

$$\|P(\mathcal{D}F_1, \dots, \mathcal{D}F_K)\|_{U_*^{k-1}(\mathbb{Z}_N)} = O_{K,d,P}(1). \tag{1.26}$$

Proof. By triangle inequality it suffices to show this when P is a monomial. By enlarging K to at most dK and repeating the F_j if necessary, we may assume $P(x_1, \dots, x_K) = x_1 \cdots x_K$. It hence suffices to show that

$$\Big\langle f, \prod_{j=1}^{K} \mathcal{D}F_j \Big\rangle = O_K(1)$$

uniformly over all f with $\|f\|_{U^{k-1}(\mathbb{Z}_N)} \le 1$. We can expand the left-hand side using the definitions to obtain that

$$\Big\langle f, \prod_{j=1}^{K} \mathcal{D}F_j \Big\rangle = \mathbb{E}_{x \in \mathbb{Z}_N} \Big[f(x) \prod_{j=1}^{K} \mathbb{E}_{h^{(j)} \in \mathbb{Z}_N^{k-1}} \Big[\prod_{\substack{\omega \in \{0,1\}^{k-1} \\ \omega \neq 0}} F_j(x + \omega \cdot h^{(j)}) \Big] \Big].$$

Setting $h^{(j)} = h + H^{(j)}$ for any $h \in \mathbb{Z}_N^{k-1}$ and then averaging over h, we can rewrite the right-hand side of this as

$$\mathbb{E}_{x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1}} \Big[f(x) \prod_{j=1}^{K} \mathbb{E}_{H^{(j)} \in \mathbb{Z}_N^{k-1}} \Big[\prod_{\substack{\omega \in \{0,1\}^{k-1} \\ \omega \neq 0}} F_j(x + \omega \cdot H^{(j)} + \omega \cdot h) \Big] \Big].$$

We can now set

$$H := (H^{(1)}, \dots, H^{(K)}) \in (\mathbb{Z}_N^{k-1})^K, \quad \omega \cdot H := (\omega \cdot H^{(1)}, \dots, \omega \cdot H^{(K)})$$

as well as

$$g_{u^{(1)},\dots,u^{(K)}}(x) := \prod_{j=1}^{K} F_j(x + u^{(j)}) \qquad \text{for } u^{(1)}, \dots, u^{(K)} \in \mathbb{Z}_N$$

and then

$$f_{0,H} := f, \quad f_{\omega,H} := g_{\omega \cdot H} \quad (\omega \neq 0).$$

With these definitions, it is elementary to verify that the above expression coincides with

$$\mathbb{E}_{H \in (\mathbb{Z}_N^{k-1})^K} \big[\langle (f_{\omega,H})_{\omega \in \{0,1\}^{k-1}} \rangle_{U^{k-1}(\mathbb{Z}_N)} \big]$$

which, due to the Gowers-Cauchy-Schwarz inequality, is bounded by

$$\mathbb{E}_{H \in (\mathbb{Z}_N^{k-1})^K} \Bigg[\|f\|_{U^{k-1}} \prod_{\substack{\omega \in \{0,1\}^{k-1} \\ \omega \neq 0}} \|g_{\omega \cdot H}\|_{U^{k-1}(\mathbb{Z}_N)} \Bigg],$$

so that it suffices to show

$$\mathbb{E}_{H \in (\mathbb{Z}_N^{k-1})^K} \Bigg[\prod_{\substack{\omega \in \{0,1\}^{k-1} \\ \omega \neq 0}} \|g_{\omega \cdot H}\|_{U^{k-1}(\mathbb{Z}_N)} \Bigg] = O_K(1).$$

If we can show that

$$\mathbb{E}_{H \in (\mathbb{Z}_N^{k-1})^K} \big[\|g_{\omega \cdot H}\|_{U^{k-1}(\mathbb{Z}_N)}^{2^{k-1}-1} \big] = O_K(1)$$

holds for all $\omega \in \{0,1\}^{k-1}, \omega \neq 0$, then Hölder's inequality implies

$$\begin{aligned} \mathbb{E}_{H \in (\mathbb{Z}_N^{k-1})^K} \Bigg[\prod_{\substack{\omega \in \{0,1\}^{k-1} \\ \omega \neq 0}} \|g_{\omega \cdot H}\|_{U^{k-1}(\mathbb{Z}_N)} \Bigg] &\leq \prod_{\substack{\omega \in \{0,1\}^{k-1} \\ \omega \neq 0}} \mathbb{E}_{H \in (\mathbb{Z}_N^{k-1})^K} \big[\|g_{\omega \cdot H}\|_{U^{k-1}(\mathbb{Z}_N)}^{2^{k-1}-1} \big]^{1/(2^{k-1}-1)} \\ &= O_K(1), \end{aligned}$$

hence gives the claim. But we can apply Hölder's inequality again to deduce that it in fact suffices to prove that we have

$$\mathbb{E}_{H \in (\mathbb{Z}_N^{k-1})^K} \Big[\|g_{\omega \cdot H}\|_{U^{k-1}(\mathbb{Z}_N)}^{2^{k-1}} \Big] = O_K(1)$$

for any $\omega \in \{0,1\}^{k-1}$, $\omega \neq 0$. Fix such an ω, and note that $H \mapsto \omega \cdot H$ defines a uniform covering of $\mathbb{Z}_N^K$ by $(\mathbb{Z}_N^{k-1})^K$. Thus the left-hand side can be rewritten as

$$\mathbb{E}_{u^{(1)},\dots,u^{(K)} \in \mathbb{Z}_N} \Big[\|g_{u^{(1)},\dots,u^{(K)}}\|_{U^{k-1}(\mathbb{Z}_N)}^{2^{k-1}} \Big],$$

and using the definitions, we can reformulate this as

$$\mathbb{E}_{x,u^{(1)},\dots,u^{(K)}\in\mathbb{Z}_N,h\in\mathbb{Z}_N^{k-1}}\Big[\prod_{\omega'\in\{0,1\}^{k-1}}\prod_{j=1}^{K}F_j(x+u^{(j)}+h\cdot\omega')\Big]$$
$$=\mathbb{E}_{x\in\mathbb{Z}_N,h\in\mathbb{Z}_N^{k-1}}\Big[\prod_{j=1}^{K}\mathbb{E}_{u^{(j)}\in\mathbb{Z}_N}\Big[\prod_{\omega'\in\{0,1\}^{k-1}}F_j(x+u^{(j)}+h\cdot\omega')\Big]\Big].$$

At this point, we make use of the assumption $|F_j|\le\nu$, which reduces our claim further to showing that

$$\mathbb{E}_{x\in\mathbb{Z}_N,h\in\mathbb{Z}_N^{k-1}}\Big[\mathbb{E}_{u\in\mathbb{Z}_N}\Big[\prod_{\omega'\in\{0,1\}^{k-1}}\nu(x+u+h\cdot\omega')\Big]^K\Big]=O_K(1).$$

Substituting $y:=x+u$, the left-hand side reads

$$\mathbb{E}_{h\in\mathbb{Z}_N^{k-1}}\Big[\mathbb{E}_{y\in\mathbb{Z}_N}\Big[\prod_{\omega'\in\{0,1\}^{k-1}}\nu(y+h\cdot\omega')\Big]^K\Big],$$

at which point we may apply the correlation condition to the inner expectation (this is in fact the only point in the proof where we make use of this condition). It implies

$$\mathbb{E}_{y\in\mathbb{Z}_N}\Big[\prod_{\omega'\in\{0,1\}^{k-1}}\nu(y+h\cdot\omega')\Big]\le\sum_{\omega'\neq\omega''\in\{0,1\}^{k-1}}\tau(h\cdot(\omega'-\omega''))$$

for a weight function τ satisfying $\mathbb{E}[\tau^q]=O_q(1)$. Triangle inequality in $L^K(\mathbb{Z}_N^{k-1})$ gives

$$\mathbb{E}_{h\in\mathbb{Z}_N^{k-1}}\Big[\Big(\sum_{\omega'\neq\omega''\in\{0,1\}^{k-1}}\tau\big(h\cdot(\omega'-\omega'')\big)\Big)^K\Big]\le\Big(\sum_{\omega'\neq\omega''\in\{0,1\}^{k-1}}\mathbb{E}_{h\in\mathbb{Z}_N^{k-1}}\big[\tau(h\cdot(\omega'-\omega''))^K\big]^{1/K}\Big)^K,$$

so that it is in fact sufficient to see that

$$\mathbb{E}_{h\in\mathbb{Z}_N^{k-1}}\Big[\tau(h\cdot(\omega'-\omega''))^K\Big]=O_K(1)$$

for $\omega'\neq\omega''$. But in this case, $h\mapsto h\cdot(\omega'-\omega'')$ is a uniform covering of $\mathbb{Z}_N$ by $\mathbb{Z}_N^{k-1}$, so the left-hand side coincides with $\mathbb{E}[\tau^K]=O_K(1)$. □

We now turn to the general case. Let $\varepsilon>0$ be arbitrary. Since the functions $\mathcal{D}F_j$ have image contained in the compact interval I, the Weierstrass approximation theorem implies the existence of some polynomial P (depending only on K and ε) which satisfies

$$\|\Phi(\mathcal{D}F_1,\dots,\mathcal{D}F_K)-P(\mathcal{D}F_1,\dots,\mathcal{D}F_K)\|_{L^\infty}\le\varepsilon.$$

Hence, the measure property of ν implies

$$|\langle\nu-1,\Phi(\mathcal{D}F_1,\dots,\mathcal{D}F_K)-P(\mathcal{D}F_1,\dots,\mathcal{D}F_K)\rangle|\le 3\varepsilon$$

for sufficiently large N. But the Lemma gives

$$\langle \nu - 1, P(\mathcal{D}F_1, \dots, \mathcal{D}F_K) \rangle = o_{K,\varepsilon}(1).$$

Combining these two equations, we see that for sufficiently large N (depending on K and ε), we have

$$|\langle \nu - 1, \Phi(\mathcal{D}F_1, \dots, \mathcal{D}F_K) \rangle| \leq 4\varepsilon,$$

which implies the claim. Now if we let Φ range over a compact, hence precompact, set, we can cover this by finitely many balls of radius ε w.r.t. the uniform topology. The same argument then implies the second statement. □

1.7 Generalised Bohr sets and σ-algebras

Recall that a σ-algebra $\mathcal{B}$ over $\mathbb{Z}_N$ is a subset of the power set $\mathcal{P}(\mathbb{Z}_N)$ of $\mathbb{Z}_N$ which

- contains the empty set,
- is closed under complementation
- and under taking countable intersections.

The atoms of a σ-algebra $\mathcal{B}$ are the minimal non-empty elements of $\mathcal{B}$ w.r.t. inclusion. These form a partition of $\mathbb{Z}_N$, as one may easily verify.

A function $f : \mathbb{Z}_N \to \mathbb{R}$ is said to be $\mathcal{B}$-measurable if all level sets $f^{-1}(x)$ lie in $\mathcal{B}$, or equivalently, if f is constant on each atom of $\mathcal{B}$. We define $L^q(\mathcal{B})$ to be the set of $\mathcal{B}$-measurable functions equipped with the L^q-norm.

Moreover, we define the conditional expectation $\mathbb{E}[f \,|\, \mathcal{B}] \in L^2(\mathcal{B})$ of a function f under some σ-algebra $\mathcal{B}$ by

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

where $\mathcal{B}(x)$ denotes the atom of $\mathcal{B}$ containing x. This coincides with the orthogonal projection of f onto $L^2(\mathcal{B})$, i.e. the function in $L^2(\mathcal{B})$ which minimises the L^2-distance to f.

For a collection of σ-algebras $\mathcal{B}_1, \dots, \mathcal{B}_K$, we define the join $\bigvee_{i=1}^K \mathcal{B}_i$ to be the σ-algebra whose atoms are the intersections of atoms of the $\mathcal{B}_i$. This is the same as the σ-algebra generated by $\mathcal{B}_1, \dots, \mathcal{B}_K$.

Our goal now is to construct a σ-algebra, such that the measurable functions of this σ-algebra can be approximated well by functions of the type $\Phi(\mathcal{D}F_1, \dots, \mathcal{D}F_K)$ we considered in Proposition 1.36. We first prove a Proposition concerning more general functions $G : \mathbb{Z}_N \to I$ as below, and then specialise to the case of basic Gowers anti-uniform functions.

Proposition 1.38. *Let ν be a k-pseudorandom measure, and let $0 < \varepsilon < 1$ and $0 < \eta < 1/2$. Moreover, let $I := [-2^{2^{k-1}}, 2^{2^{k-1}}]$ and $G : \mathbb{Z}_N \to I$ be a map. Then there exists a σ-algebra $\mathcal{B}_{\varepsilon,\eta}(G)$ with the following properties:*

(i) For any σ-algebra $\mathcal{B}$, we have

$$\|G - \mathbb{E}[G \,|\, \mathcal{B} \vee \mathcal{B}_{\varepsilon,\eta}(G)]\|_{L^\infty} \leq \varepsilon, \tag{1.27}$$

(ii) $\mathcal{B}_{\varepsilon,\eta}(G)$ is generated by at most $O(1/\varepsilon)$ atoms and

(iii) if A denotes any atom of $\mathcal{B}_{\varepsilon,\eta}(G)$ then there exists a continuous function $\Psi_A : I \to [0,1]$ s.t.

$$\|(\mathbb{1}_A - \Psi_A \circ G)(\nu + 1)\|_{L^1} = O(\eta). \tag{1.28}$$

Moreover, Ψ_A lies in a compact set $E_{\varepsilon,\eta}$ which does not depend on N, F, ν or A.

Proof. Note that we have

$$\begin{aligned}&\int_0^1 \sum_{n\in\mathbb{Z}} \mathbb{E}_{x\in\mathbb{Z}_N}\Big[\mathbb{1}\big(G(x) \in [\varepsilon(n-\eta+\alpha), \varepsilon(n+\eta+\alpha)]\big)\big(\nu(x)+1\big)\Big]\, d\alpha \\ &= \mathbb{E}_{x\in\mathbb{Z}_N}\Big[\big(\nu(x)+1\big)\int_{-\infty}^{\infty} \mathbb{1}\big(G(x) \in [-\eta+\alpha, \eta+\alpha]\big)\, d\alpha\Big] = 2\eta\mathbb{E}_{x\in\mathbb{Z}_N}[\nu(x)+1] = O(\eta).\end{aligned}$$

By the pigeonhole principle there thus exists $0 \leq \alpha \leq 1$ such that

$$\sum_{n\in\mathbb{Z}} \mathbb{E}_{x\in\mathbb{Z}_N}\Big[\mathbb{1}\big(G(x) \in [\varepsilon(n-\eta+\alpha), \varepsilon(n+\eta+\alpha)]\big)\big(\nu(x)+1\big)\Big] = O(\eta). \tag{1.29}$$

Fix such an α and define $\mathcal{B}_{\varepsilon,\eta}(G)$ to be the σ-algebra whose atoms are given by sets of the form $G^{-1}([\varepsilon(n+\alpha), \varepsilon(n+\alpha+1)])$ for $n \in \mathbb{Z}$. We may clearly assume $n = O(1/\varepsilon)$, since the atoms are empty otherwise; this establishes (ii).

To see that (i) is satisfied, let $x \in \mathbb{Z}_N$ and $y \in \mathcal{B} \vee \mathcal{B}_{\varepsilon,\eta}(G)(x) \subseteq \mathcal{B}_{\varepsilon,\eta}(G)(x)$ for any σ-algebra $\mathcal{B}$. Then we have

$$|G(x) - G(y)| \leq \varepsilon$$

and therefore

$$|G(x) - \mathbb{E}[G \,|\, \mathcal{B} \vee \mathcal{B}_{\varepsilon,\eta}(G)](x)| \leq \varepsilon,$$

which gives (i).

To establish (iii), let A be an atom of $\mathcal{B}_{\varepsilon,\eta}(G)$, and let $\psi_\eta : \mathbb{R} \to [0,1]$ be a fixed continuous function with $\psi_\eta \equiv 1$ on $[\eta, 1-\eta]$ and $\psi_\eta \equiv 0$ outside of $[-\eta, 1+\eta]$. Then, define

$$\Psi_A(x) := \psi_\eta\left(\frac{x}{\varepsilon} - n - \alpha\right).$$

It is clear that Ψ_A ranges over a compact set, because we may assume $n = O(1/\varepsilon)$ and have $\alpha \in [0,1]$. Now one easily verifies that

- $\mathbb{1}_A - \Psi_A \circ G$ vanishes outside $[-\eta, 1+\eta]$,
- $\mathbb{1}_A - \Psi_A \circ G$ vanishes on $[\eta, 1-\eta]$ and
- $\mathbb{1}_A - \Psi_A \circ G$ is bounded by 1.

Application of (1.29) now yields the claim. □

As mentioned before, and indicated by the choice of I (and G), we want to apply this result to basic Gowers anti-uniform functions. More precisely, we need the following

Proposition 1.39. *Let ν be a k-pseudorandom measure, and let $K \in \mathbb{N}$. Moreover, let $\mathcal{D}F_1, \dots, \mathcal{D}F_K : \mathbb{Z}_N \to I$ be basic Gowers anti-uniform functions, and let $0 < \varepsilon < 1$ and $0 < \eta < 1/2$. Set $\mathcal{B}_{\varepsilon,\eta}(\mathcal{D}F_j)$, $j = 1, \dots, K$ as constructed in the previous proposition, and define $\mathcal{B} := \mathcal{B}_{\varepsilon,\eta}(\mathcal{D}F_1) \vee \cdots \vee \mathcal{B}_{\varepsilon,\eta}(\mathcal{D}F_K)$. Then there exists $\eta_0 = \eta_0(\varepsilon, K)$ such that for $\eta < \eta_0$ there is $N_0 = N_0(\varepsilon, K, \eta)$, such that for $N > N_0$ we have*

$$\|\mathcal{D}F_j - \mathbb{E}[\mathcal{D}F_j \,|\, \mathcal{B}]\|_{L^\infty} \leq \varepsilon, \qquad 1 \leq j \leq K. \tag{1.30}$$

In addition, there exists an exceptional set $\Omega \in \mathcal{B}$ with

$$\mathbb{E}[(\nu + 1)\mathbb{1}_\Omega] = O_{K,\varepsilon}(\eta^{1/2}) \tag{1.31}$$

as well as

$$\|(1 - \mathbb{1}_\Omega)\mathbb{E}[\nu - 1 \,|\, \mathcal{B}]\|_{L^\infty} = O_{K,\varepsilon}(\eta^{1/2}). \tag{1.32}$$

Proof. We only sketch the proof; essentially, this follows from an (inductive) application of the result before. Indeed, (1.30) is immediate from (1.27).

Since each of the K σ-algebras $\mathcal{B}_{\varepsilon,\eta}(\mathcal{D}F_j)$ is generated by $O(1/\varepsilon)$ atoms, we have that $\mathcal{B}$ is generated by $O_{K,\varepsilon}(1)$ atoms. An atom A of $\mathcal{B}$ is said to be small if

$$\mathbb{E}[(\nu + 1)\mathbb{1}_A] \leq \eta^{1/2}.$$

We define Ω to be the union of all small atoms. Clearly, Ω lies in $\mathcal{B}$, and we have

$$\mathbb{E}[(\nu + 1)\mathbb{1}_\Omega] = \sum_{A \text{ small}} \mathbb{E}[(\nu + 1)\mathbb{1}_A] = O_{K,\varepsilon}(\eta^{1/2}),$$

which is (1.31).

For the third claim, let A be a large atom (an atom which is not small). Inductive application of Proposition 1.38 quickly implies that there is a continuous function $\Psi_A : I^K \to [0, 1]$ with the property

$$\|(\nu + 1)(\mathbb{1}_A - \Psi_A(\mathcal{D}F_1, \dots, \mathcal{D}F_K))\|_{L^1} = O_K(\eta),$$

and hence

$$\|(\nu - 1)(\mathbb{1}_A - \Psi_A(\mathcal{D}F_1, \dots, \mathcal{D}F_K))\|_{L^1} = O_K(\eta).$$

In addition, one may easily ensure that Ψ_A ranges over a compact set $E_{\varepsilon,\eta,K}$ inside $C(I^K)$. Together with Proposition 1.36, we thus obtain

$$\mathbb{E}[(\nu - 1)\Psi_A(\mathcal{D}F_1, \dots, \mathcal{D}F_K)] = o_{K,\varepsilon,\eta}(1),$$

and after some manipulations, this implies the third claim. □

1.8 The generalised Koopman-Von Neumann Theorem and completion of the proof

The following Proposition is the last ingredient we need in order to conclude the Main Theorem.

Proposition 1.40 (Generalised Koopman-Von Neumann Theorem)**.** *Let ν be a k-pseudorandom measure, and let $f : \mathbb{Z}_N \to \mathbb{R}^+$ be such that $0 \le f \le \nu$. Let $\varepsilon > 0$ be sufficiently small, and $N > N_0(\varepsilon)$ be sufficiently large. Then there exists a σ-algebra $\mathcal{B}$ and an exceptional set $\Omega \in \mathcal{B}$ such that*

$$\mathbb{E}[\nu \mathbb{1}_\Omega] = o_\varepsilon(1), \tag{1.33}$$

$$\|(1 - \mathbb{1}_\Omega)\mathbb{E}[\nu - 1 \,|\, \mathcal{B}]\|_{L^\infty} = o_\varepsilon(1), \tag{1.34}$$

$$\|(1 - \mathbb{1}_\Omega)(f - \mathbb{E}[f \,|\, \mathcal{B}])\|_{U^{k-1}(\mathbb{Z}_N)} \le \varepsilon^{1/2^k}. \tag{1.35}$$

There are several connections to ergodic theory, which we do not wish to elaborate on. We only mention the related works [1, 2].

We will not proceed to prove this statement, but note that the proof is considerably easier than the problem we initially faced. We omit it mainly due to its severe technicality. However, we do show how Proposition 1.40 implies Theorem 1.11, and therefore, as already mentioned, the Main Theorem 1.6 under application of Proposition 1.12.

Proof of Theorem 1.11 assuming Proposition 1.40. Let f, δ be as in 1.11, and let $\varepsilon > 0$ to be chosen later. Moreover, let $\mathcal{B}$ and Ω be as in Proposition 1.40. Define

$$f_U := (1 - \mathbb{1}_\Omega)(f - \mathbb{E}[f \,|\, \mathcal{B}])$$

and

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

We have

$$\mathbb{E}[f_{U^\perp}] = \mathbb{E}[(1 - \mathbb{1}_\Omega)f] \ge \mathbb{E}[f] - \mathbb{E}[\nu \mathbb{1}_\Omega] \ge \delta - o_\varepsilon(1).$$

In addition, an application of (1.34) yields

$$\|f_{U^\perp}\|_{L^\infty} \le 1 + \|(1 - \mathbb{1}_\Omega)\mathbb{E}[\nu - 1 \,|\, \mathcal{B}]\|_{L^\infty} \le 1 + o_\varepsilon(1).$$

Without changing the notation, we renormalise $f_{U^\perp}$ by this factor $1 + o_\varepsilon(1)$, so that the function is bounded by 1. This preserves the property $\mathbb{E}[f_{U^\perp}] \ge \delta - o_\varepsilon(1)$. We are now in a position to apply Szemerédi's Theorem in the version of 1.2 to $f_{U^\perp}$ with $\delta' = \delta - o_\varepsilon(1)$ to deduce that

$$\mathbb{E}_{x,r \in \mathbb{Z}_N}[f_{U^\perp}(x) f_{U^\perp}(x + r) \cdots f_{U^\perp}(x + (k-1)r)] \ge c(k, \delta) - o_\varepsilon(1) - o_{k,\delta}(1).$$

We note that there is a small technicality here, because on the right-hand side we do not obtain $c(k, \delta)$ but $c(k, \delta')$. There are different ways to handle this; one can, for

example, modify $f_U^\perp$ by adding a constant and then dividing by the new L^∞-norm in such a way that the new function still has expectation at least δ. This is possible, and in the end one obtains $c(k,\delta) - o_{k,\delta}(1) - o_\varepsilon(1)/(1-\delta)$ as a lower bound for the old $f_{U^\perp}$ after substituting back, which is fine.

The idea now is to look at

$$\mathbb{E}_{x,r\in\mathbb{Z}_N}[\tilde{f}(x)\cdots\tilde{f}(x+(k-1)r)],$$

where we denote $\tilde{f} := (1-\mathbb{1}_\Omega)f = f_U + f_{U^\perp}$, and after expanding the product one obtains a sum of 2^k terms. The main term stems from choosing $f_{U^\perp}$ in every factor, which can be estimated by the above. The other terms contain a factor f_U, and from (1.35) we have $\|f_U\|_{U^{k-1}(\mathbb{Z}_N)} \le \varepsilon^{1/2^k}$. We now want to apply the von Neumann Theorem 1.29; since $(1-\mathbb{1}_\Omega)f$ is bounded by ν and $f_{U^\perp}$ is bounded by $1+o_\varepsilon(1)$, we see that f_U is bounded by $\nu+1+o_\varepsilon(1)$. After a small renormalisation as done before (and then going back to the old function, noting that we did not modify much), we see that

$$\mathbb{E}_{x,r\in\mathbb{Z}_N}[\tilde{f}(x)\cdots\tilde{f}(x+(k-1)r)] \ge c(k,\delta) - O(\varepsilon^{1/2^k}) - o_\varepsilon(1) - o_{k,\delta}(1).$$

Now $0 \le (1-\mathbb{1}_\Omega)f \le f$, so that the same bound holds for f. But since $\varepsilon > 0$ was arbitrary, the error term on the right-hand side can be made arbitrarily small by taking $N = N(k,\delta)$ sufficiently large. □

2 Linear equations in Primes

2.1 Introduction and Main Theorem

Our main reference for this chapter is [8].

After considering arithmetic progressions in the primes, it is not far-fetched to guess that a similar result holds for more general linear forms inside the primes. In the following, we consider affine-linear forms $\psi : \mathbb{Z}^d \to \mathbb{Z}$, given by

$$\psi(n_1, \dots, n_d) = a_1 n_1 + \cdots + a_d n_d + b$$

for some integers $a_1, \dots, a_d, b$. We call $\psi(0) = b$ the inhomogenous part and $\dot{\psi} := \psi - \psi(0)$ the homogenous part. Moreover, we let

$$\Psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$$

be a system of affine-linear forms (and adapt the notation $\Psi(0)$ and $\dot{\Psi}$ in the obvious way). To avoid degeneracies, we require that in an affine-linear system, no ψ_i is constant and that no two forms are rational multiples of each other.

For a positive integer N, we further define the size $\|\Psi\|_N$ of a system Ψ by

$$\|\Psi\|_N := \sum_{i=1}^{t} \sum_{j=1}^{d} |\dot{\psi}_i(e_j)| + \sum_{i=1}^{t} \left| \frac{\psi_i(0)}{N} \right|,$$

where e_j denotes the j-th unit vector. We will look at the question whether the image of Ψ contains infinitely many prime lattice points, i.e. elements of $\Psi\mathbb{Z}^d$ such that every component is prime. When considering this problem, one will find two trivial obstructions:

Example 2.1. Consider the system of affine linear forms $\Psi : \mathbb{Z} \to \mathbb{Z}^3$ given by

$$n \mapsto (n, n+2, n+4).$$

(Note that systems of this type will usually not be treatable by the methods we employ due to the fact that it has infinite complexity). Then every point in the image has one component divisible by 3 and can therefore not contain infinitely many prime lattice point in its image. This is related to the prime k-tuple conjecture and the admissibility of k-tuples:

A k-tuple of non-negative integers $(a_1, \dots, a_k)$ is called admissible if there is no prime p such that $a_1, \dots, a_k$ include every residue class modulo p. For admissible k-tuples, the prime k-tuple conjecture claims that there are infinitely many values of n such that $n + a_1, \dots, n + a_k$ is prime (we can always assume $a_1 = 0$).

Example 2.2. Consider now, for example, the affine-linear system

$$(n_1, n_2) \mapsto (n_1, n_2, -n_1 - 2n_2 + 100).$$

This illustrates an even more trivial obstruction: The image contains only finitely many points consisting only of positive components; hence it can not contain infinitely many prime lattice points. We will however see (for some types of linear systems) that these are the only two obstructions.

To be more precise, we will look at the number of prime lattice points in the image of a convex body K contained in $[-N,N]^d$ under an affine-linear form Ψ, and try to establish an asymptote as $N \to \infty$. The reason for assuming convexity will become apparent soon, but the reader may always think of $K = [-N,N]^d$ to have a picture in mind. In addition, we will give weights to the prime points according to the von Mangoldt function, a method which is used throughout analytic number theory. We are therefore interested in estimating the expression

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda(\psi_i(n)) \tag{2.1}$$

for some convex body $K \subseteq [-N,N]^d$ and a large integer N. As asserted by the prime number theorem, Λ on average behaves like $\mathbb{1}_{\mathbb{R}^+}$, so one might compare this to the same expression with $\mathbb{1}_{\mathbb{R}^+}$ instead of Λ:

Lemma 2.3. *Let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be an affine-linear form such that $\|\Psi\|_N \leq L$. Then*

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \mathbb{1}_{\mathbb{R}^+}(\psi_i(n)) = \beta_\infty + o_{d,t,L}(N^d), \tag{2.2}$$

where $\beta_\infty := vol_{\mathbb{R}^d}(K \cap \Psi^{-1}(\mathbb{R}^+)^t)$.

We will not proceed to prove this result, mainly because it is more of a question in convex analysis. We do want to give some heuristic explanation: One may assume that the ψ_i are positive on K by intersecting it with the corresponding half planes (and still staying a convex set therefore). We then obtain that β_∞ is simply the volume of this new set K, and we can also replace $\mathbb{1}_{\mathbb{R}^+}(\psi_i(n))$ by the constant function 1 on K. The Lemma now simply claims that

$$|K \cap \mathbb{Z}^d| = vol_{\mathbb{R}^d}(K) + o_d(N^d),$$

which should heuristically be very plausible. In fact, it is not harder to show that the error is $O_d(N^{d-1})$, but we do not need this statement.

It turns out however that replacing Λ by $\mathbb{1}_{\mathbb{R}^+}$ loses the arithmetic content of the expression in question; there are local irregularities at small primes. To explain this in detail, define for $q \geq 1$ the local von Mangoldt function $\Lambda_{\mathbb{Z}_q} : \mathbb{Z} \to \mathbb{R}^+$ by

$$\Lambda_{\mathbb{Z}_q}(b) := \begin{cases} \frac{q}{\varphi(q)}, & \text{if } (b,q) = 1 \\ 0 & \text{else} \end{cases}.$$

Wit this definition, the prime number theorem in APs reads

$$\sum_{n \leq N} \Lambda(qn + b) = \Lambda_{\mathbb{Z}_q}(b)N + o_q(N). \tag{2.3}$$

We can now define the local factor β_q by

$$\beta_q := \mathbb{E}_{n \in \mathbb{Z}_q^d} \prod_{i \in [t]} \Lambda_{\mathbb{Z}_q}(\psi_i(n)). \tag{2.4}$$

An application of the Chinese Remainder Theorem quickly yields that the local factors are multiplicative (in the usual sense of number theory). Moreover, we will soon see that for large primes p, β_p is close to 1 and in view of the next statement, it therefore makes sense to speak of local irregularities at small primes.

Hardy and Littlewood conjectured (a special case of) the following

Conjecture 2.4. Let N, d, t, L be positive integers and let Ψ be a system of affine-linear forms of size $\|\Psi\|_N \leq L$. Let $K \subseteq [-N, N]^d$ be a convex set. Then we have

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda(\psi_i(n)) = \beta_\infty \prod_p \beta_p + o_{t,d,L}(N^d). \tag{2.5}$$

Example 2.5. Recall example 2.1. We noted the trivial obstruction that one of the numbers $n, n+2, n+4$ is always divisible by 3, or equivalently, not coprime to 3. This implies that

$$\beta_3 = \mathbb{E}_{n \in \mathbb{Z}_3} \Lambda_{\mathbb{Z}_3}(n) \Lambda_{\mathbb{Z}_3}(n+2) \Lambda_{\mathbb{Z}_3}(n+4) = 0.$$

The 'Main Term' in (2.5) therefore vanishes in this case, which fits into our observation that there are only finitely many (in fact: one) prime points in the image of this form.

Example 2.6. Recall now example 2.2. One verifies quickly that for any $K \subseteq [-N, N]^2$, we have

$$\beta_\infty \leq vol_{\mathbb{R}^2}(\{(x_1, x_2) \in (\mathbb{R}^+)^2 : x_1 + 2x_2 \leq 100\}).$$

The right-hand side is $O(1)$; we will soon show that the singular series $\prod_p \beta_p$ is absolutely convergent, and again we observe that in this case the 'Main Term' is much smaller than the error term. The global factor and the local factors are therefore able to detect our two trivial obstructions.

We will only proceed to prove this generalised Hardy-Littlewood conjecture for a special type of forms. To explain this more precisely, we need the notion of complexity.

Definition 2.7. *Let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms. For $1 \leq i \leq t$ and $s \geq 0$, we say that Ψ has i-complexity at most s if the following condition holds: The $t-1$ forms $\psi_j, j \neq i$ can be covered by $s+1$ sets such that ψ_i does not lie in the affine-linear span of any class (i.e. the homogeneous part of ψ_i is not a linear combination of their homogenous parts).*

The complexity of Ψ is then defined to be the smallest s such that Ψ has i-comlexity at most s for all i.

Several examples are in place to make this more vivid:

Example 2.8. We begin with the most obvious examples: The system

$$(n_1, \dots, n_d) \mapsto (n_1, \dots, n_d)$$

has complexity 0, because each form is independent of the rest. For $k \geq 2$, the form

$$(n, r) \mapsto (n, n + r, \dots, n + (k-1)r),$$

which counts arithmetic progressions of length k, has complexity $k - 2$. This is not hard to show: No two forms are affinely related, but every form is an affine-linear combination of any two other forms.

Now look at the system

$$n \mapsto (n, n + 2),$$

which corresponds to twin primes. This is a system of infinite complexity; in fact, a system has infinite complexity if and only if two of the forms are affinely related (otherwise, we can always assign individual classes to each form).

Example 2.9. Let us now look at a slightly more complicated system. Let $d \geq 2$ and $t := 2^{d-1}$. Define

$$\Psi(n_1, \dots, n_d) := \Big(n_1 + \sum_{j \in A} n_j\Big)_{A \subseteq \{2,\dots,d\}}.$$

This system counts $(d-1)$-dimensional cubes whose vertices are all prime. This system has complexity at most $d - 2$. To see this, consider the form n_1. We can cover the other $t-1$ forms by $d-1$ classes by defining class i to be the set of forms which involve n_{i+1}. Then any affine-linear combination of forms inside the i-th class has the same coefficient for n_1 and for n_{i+1}, and can therefore not be n_1.

More generally fix any set A and look at the corresponding form ψ_A. Define the i-th class to be the set of forms which have a different coefficient in front of n_{i+1} than ψ_A; it is clear that this defines a cover of the other forms. Assume first that $i+1 \in A$. Then all forms in class i do not involve n_{i+1}, hence no affine-linear combination does. Thus, ψ_A can not be in their affine-linear span. Now let us assume that $i + 1 \notin A$. Then all forms of the i-th class involve n_{i+1}. But this means that any affine-linear combination has the same coefficient for n_1 and n_{i+1}, hence never give ψ_A.

This implies that indeed the complexity is at most $d-2$. In fact, it is not hard to show that the complexity of this system is precisely $d - 2$.

We are now in a position to formulate the Main Theorem of this chapter:

Main Theorem 2.10. *The generalised Hardy-Littlewood conjecture 2.4 is true for all systems of affine-linear forms of finite complexity.*

In the course of the proof, we will heavily rely on two recent and very deep theorems, namely the Gowers inverse norm theorem and the Möbius and nilsequences theorem. Both of these theorems were formulated by Green and Tao and proved quite recently by both of them, partially together with Ziegler [5, 6, 9–11]. We will formulate both

statements later, and in such a way that they depend on a parameter s (we will call the statements GI(s) and MN(s)). We will then try to illustrate the proof of the Main Theorem for systems of complexity at most s assuming GI(s) and MN(s).

In the following, we will view t, d, L and s as fixed, and we will usually omit dependencies of these variables in our notation.

Example 2.11. Let us look at arithmetic progressions of length 4. One verifies that choosing K to be the convex set $\{(n_1, n_2) : 1 \le n_1 \le n_1 + 3n_2 \le N\}$ gives $\beta_\infty = N^2/6$, and that we have $\beta_2 = 4, \beta_3 = 9/8$ and $\beta_p = 1 - \frac{3p-1}{(p-1)^3}$ for $p \ge 5$. Therefore, the number of prime quadruplets $p_1 < \cdots < p_4 \le N$ in arithmetic progression is

$$(1 + o(1))\sigma_1 \frac{N^2}{\log^4 N},$$

where $\sigma_1 = \frac{3}{4} \prod_{p \ge 5} \left(1 - \frac{3p-1}{(p-1)^3}\right) \approx 0.4764$.

Let us now look at arithmetic progressions of length k. One verifies that the number of primes $p_1 < \cdots < p_k \le N$ in arithmetic progressions is

$$(1 + o(1)) \frac{1}{2(k-1)} \prod_p \beta_p \frac{N^2}{\log^k N},$$

where

$$\beta_p = \begin{cases} \frac{1}{p}\left(\frac{p}{p-1}\right)^{k-1} & \text{if } p \le k \\ \left(1 - \frac{k-1}{p}\right)\left(\frac{p}{p-1}\right)^{k-1} & \text{if } p > k \end{cases}.$$

This is heuristically very plausible: The map

$$(n, r) \mapsto (n, n + r, \ldots, n + (k-1)r)$$

has two free parameters n and r. The heuristic probability that a number $\le N$ is prime is $\frac{1}{\log N}$ according to the prime number theorem. This gives us the term $N^2/\log^k N$.

Now, let us turn our attention to the multiplicative constant. If *plek*, then n needs to be coprime to p and r has to be a multiple of p. The first event has probability $(p-1)/p$, while the second event has probability $1/p$. Furthermore, the value of the local von Mangoldt function in this case is $p/(p-1)$ and appears k times, which indeed gives us

$$\frac{1}{p}\left(\frac{p}{p-1}\right)^{k-1}.$$

If on the other hand p is larger than k, then n needs to be coprime to p, which again has probability $(p-1)/p$, while r needs to be chosen such that none of the numbers $n + r, \ldots, n + (k-1)r$ is divisible by p. An elementary number-theoretic argument implies that this happens for precisely $p - (k-1)$ values of r if n is coprime to p. Since the value of the von Mangoldt function is $p/(p-1)$ again and appears k times, we indeed obtain

$$\left(1 - \frac{k-1}{p}\right)\left(\frac{p}{p-1}\right)^{k-1}.$$

The factor $\frac{1}{2(k-1)}$ is the proportion of the volume inside $[-N,N]^2$ such that the numbers $n, n+r, \dots, n+(k-1)r$ are all positive, as the reader may quickly verify. This heuristically explains the above asymptotic.

It if important to verify that the infinite product $\prod_p \beta_p$ in fact converges for all systems that are of interest to us:

Lemma 2.12. *Let $N \geq 1$, and let $\Psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be an affine-linear system of finite complexity such that $\|\Psi\|_N \leq L$. Then we have $\beta_p = 1 + O(p^{-2})$. In particular, $\prod_p \beta_p$ converges absolutely, but it might vanish at small primes.*

Proof. We may assume that p is sufficiently large depending on d, t and L. Since Ψ has finite complexity, we have that no two of the ψ_i are affinely related. Writing

$$\psi_i(n) = a_1^i n_1 + \cdots + a_d^i n_d + b^i,$$

this means that no two of the vectors $(a_1^i, \dots, a_d^i)$ are rational multiples of each other. We will now show that for sufficiently large p, this implies that no two of the vectors are linearly dependent over $\mathbb{Z}_p$ (where we identify the vectors as elements of $\mathbb{Z}_p^d$ in the obvious way).

If $d = 1$ and $t \geq 2$, any two vectors are rational multiples of each other, so we may exclude this case. If $d = t = 1$ there is nothing to show. Now assume $d \geq 2$ and that there are $1 \leq i < j \leq t$ and $\lambda \in \mathbb{Z}_p$ such that $\lambda a_k^i = a_k^j$ in $\mathbb{Z}_p$ for all $k = 1, \dots, d$. Then for any two indices $k_1 \neq k_2$ we have $a_{k_1}^i a_{k_2}^j - a_{k_2}^i a_{k_1}^j = 0$ in $\mathbb{Z}_p$. But if p is sufficiently large then this is an equality in $\mathbb{Z}$ (the vector entries are bounded by L), and the corresponding vectors are rational multiples of each other, a contradiction.

Since we know that for sufficiently large p no two of the vectors are linearly dependent over $\mathbb{Z}_p$, elementary linear algebra tells us that for any $1 \leq i < j \leq t$ the proportion of $n \in \mathbb{Z}_p^d$ such that $\psi_i(n)$ and $\psi_j(n)$ are divisible by p is $O(p^{-2})$.

It is clear that the proportion of n such that $\psi_i(n) = 0$ in $\mathbb{Z}_p$ is precisely p^{-1} for any i, since we assumed the forms to be non-constant. Now, define

$$A_i := \{n \in \mathbb{Z}_p^d : \psi_i(n) = 0 \in \mathbb{Z}_p\}.$$

Then we have

$$\beta_p = \mathbb{E}_{n \in \mathbb{Z}_p^d}\Big[\prod_{i \in [t]} \Lambda_{\mathbb{Z}_p}(\psi_i(n))\Big] = \frac{1}{p^d}\left(\frac{p}{p-1}\right)^t |A_1^c \cap \cdots \cap A_t^c| = \left(\frac{p}{p-1}\right)^t \left(1 - \frac{|A_1 \cup \cdots \cup A_t|}{p^d}\right).$$

But the Bonferroni inequalities imply that

$$|A_1 \cup \cdots \cup A_t| = \sum_{i=1}^t |A_i| + O\left(\sum_{1 \leq i < j \leq t} |A_i \cap A_j|\right).$$

Above, we have established that

$$\frac{|A_i|}{p^d} = \frac{1}{p} \quad \text{and} \quad \frac{|A_i \cap A_j|}{p^d} = O(p^{-2}) \qquad (i < j).$$

Putting everything together, this implies

$$\beta_p = \left(\frac{p}{p-1}\right)^t \left(1 - tp^{-1} + O(p^{-2})\right) = \left(1 + \frac{t}{p-1} + O(p^{-2})\right)\left(1 - \frac{t}{p} + O(p^{-2})\right) = 1 + O(p^{-2})$$

as claimed. □

2.2 Normal form

We are now going to reformulate our Main Theorem in a more accessible form by introducing the normal form. We are not going to prove that it indeed reduces to the statement we are going to formulate, because this reduction is technical and has few arithmetic content.

Definition 2.13. *Let $\Psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms, and let $s \geq 0$. We say that Ψ is in s-normal form if for every $i \in [t]$ there is a collection $J_i \subseteq \{e_1, \dots, e_d\}$ of basis vectors of cardinality $|J_i| \leq s+1$ such that $\prod_{e \in J_i} \dot{\psi}_{i'}(e)$ is non-zero for $i' = i$ and zero otherwise.*

Informally, this means that for every ψ_i there is a collection of $\leq s+1$ variables such that ψ_i is the only form that truly uses all of these variables. We will give some examples to illustrate this.

Example 2.14. The system

$$(n, r) \mapsto (n, n+r, \dots, n+(k-1)r)$$

we used to parametrise arithmetic progressions of length k, has complexity $k-2$, but is not in any normal form. However, the system

$$(n_1, \dots, n_k) \mapsto (n_2 + 2n_3 + \dots + (k-1)n_k, -n_1 + n_3 + \dots + (k-2)n_k, \dots, -(k-1)n_1 - \dots - n_{k-1}),$$

which also parametrises arithmetic progressions of length k, is in $k-2$ normal form. Indeed, we can assign to ψ_i the set $J_i = \{e_1, \dots, e_{i-1}, e_{i+1}, \dots e_k\}$, and the i-th form is the only one that uses all of the corresponding variables.

Example 2.15. Recall example 2.9. The system we defined there was not in s-normal form for any s. However, the system

$$\Psi'(n_1, \dots, n_{d-1}, n'_1, \dots, n'_{d-1}) := \left(\sum_{i \in A} n_i + \sum_{i \in [d-1] \setminus A} n'_i\right)_{A \subseteq [d-1]}$$

which also counts $(d-1)$-dimensional cubes, is in $(d-2)$-normal form. Indeed the form corresponding to the set A is the only one which truly uses the $d-1$ variables $(n_i)_{i \in A}$ and $(n'_i)_{i \in [d-1] \setminus A}$.

If a system Ψ is in s-normal form, then it is easy to see that it has complexity at most s. Fix i and consider the $t-1$ forms $\psi_1, \dots, \psi_{i-1}, \psi_{i+1}, \dots, \psi_t$. The s-normal form condition associates to i a set J_i of basis vectors with the mentioned properties. To a given basis vector $e \in J_i$ we furthermore associate the collection of forms ψ_j which satisfy $\dot{\psi_j}(e) = 0$. One verifies easily using the normal form condition that this defines a cover of the other $t-1$ forms of cardinality $|J_i| \leq s+1$. Since $\dot{\psi_i}(e) \neq 0$, we obtain that ψ_i can not lie in the affine linear span of any of these classes, which implies that the i-complexity of Ψ is at most s.

We will now try to establish some kind of converse to this consideration: Namely that every system of (finite!) complexity s admits an extension that is in s-normal form.

Definition 2.16. *Let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be an affine-linear system. An extension of Ψ is an affine-linear system $\Psi' : \mathbb{Z}^{d'} \to \mathbb{Z}^t$ with $d' \geq d$ such that*

$$\Psi' \mathbb{Z}^{d'} = \Psi \mathbb{Z}^d$$

and

$$\Psi'(n_1, \dots, n_d, 0, \dots, 0) = \Psi(n_1, \dots, n_d).$$

Note that the system of Example 2.15 is not an extension of the system in Example 2.9. However, their direct sum $\Psi \oplus \Psi'$ is an extension of Ψ in $(d-1)$-normal form.

Lemma 2.17. *Let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of finite complexity s. Then there exists an extension $\Psi' : \mathbb{Z}^{d'} \to \mathbb{Z}^t$ for some $d' = O(1)$, which is in s-normal form. If we moreover have $\|\Psi\|_N = O(1)$ then the same holds for Ψ'.*

The proof is not difficult, but technical and does not really fit into our field of interest here, so we are not going to prove it. Neither will we prove that it suffices to show the following reduction of the Main Theorem; admittedly, this sufficiency takes some effort to prove.

Theorem 2.18. *Let $N \geq 1$, $s \geq 1$, and let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be an affine-linear system in s-normal form (hence of complexity at most s) satisfying $\|\Psi\|_N = O(1)$. Let $K \subseteq [-N, N]^d$ be a convex body such that $\psi_1, \dots, \psi_t > N^{9/10}$ on K. Then we have*

$$\sum_{n \in K \cap \mathbb{Z}^d} \left(\prod_{i \in [t]} \Lambda(\psi_i(n)) - \prod_p \beta_p \right) = o(N^d). \tag{2.6}$$

We note that the global factor β_∞ has not vanished into thin air: We know that $\sum_{n \in K \cap \mathbb{Z}^d} \prod_p \beta_p = \prod_p \beta_p \sum_{n \in K \cap \mathbb{Z}^d} 1$ is close to $\beta_\infty \prod_p \beta_p$.

We also note that the exponent $9/10$ is to some extent arbitrary and should not concern the reader to much. It is not hard to see that we may assume this additional condition $\psi_i > N^{9/10}$ for all i, essentially because the majority of n has this property. We will make use of this assumption later on.

Our goal now is to sketch the proof of Theorem 2.18, trying to draw parallels to the last chapter.

2.3 The W-trick

We have already introduced a variant of the following idea in Proposition 1.12, where we defined the function $\tilde{\Lambda}$. The advantage of this modified von Mangoldt function is that it is more regular than the original von Mangoldt function and therefore we could majorise it by a pseudorandom measure. The same idea also applies here!

Definition 2.19. *Let* $w := \log_3 N$ *and*

$$W := \prod_{p \le w} p = O(\log_2 N).$$

For $b \le W$, $(b, W) = 1$, *define*

$$\Lambda_{b,W}(n) := \frac{\varphi(W)}{W} \Lambda(Wn + b).$$

Moreover, let Λ' *be the restriction of* Λ *to the primes, i.e. the function that has value* $\log p$ *at a prime* p *and* 0 *otherwise. Then, set*

$$\Lambda'_{b,W}(n) := \frac{\varphi(W)}{W} \Lambda'(Wn + b).$$

Again, Dirichlet's theorem on primes in APs asserts that $\Lambda_{b,W}(n)$ has average value 1 as $n \to \infty$. It is well-known and easy to see that Λ' is close to Λ in an average sense, which gives the same result for $\Lambda'_{b,W}$.

We also note that the choice of w is quite arbitrary and the proof would still work with slightly larger w ($\frac{1}{2} \log\log N$ works), but would need stronger statements such as the Siegel-Walfisz theorem. Since our final bounds are ineffective, we do not try to optimise w, gaining simplicity in some arguments.

With these definitions, we are in a position to make another reduction of the Main Theorem:

Theorem 2.20. *Let* $N \ge 1$, $s \ge 1$, *and let* $\Psi = (\psi_1, \dots, \psi_t)$ *be a system of affine-linear forms in* s*-normal form satisfying* $\|\Psi\|_N = O(1)$. *Let* $K \subseteq [-N, N]^d$ *be a convex body on which* $\psi_1, \dots, \psi_t > N^{4/5}$. *Then for any* $b_1, \dots, b_t \le W$ *which are coprime to* W, *we have*

$$\sum_{n \in K \cap \mathbb{Z}^d} \left(\prod_{i \in [t]} \Lambda'_{b_i,W}(\psi_i(n)) - 1 \right) = o(N^d). \tag{2.7}$$

Proof that Theorem 2.20 implies 2.18. Let Ψ, K be as in the assumptions of Theorem 2.18, and let N be sufficiently large (we can always assume this). It clearly suffices to prove equation (2.6) with Λ replaced by Λ'. Moreover, as already mentioned, by (2.2) it in fact suffices to show

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda'(\psi_i(n)) = vol_d(K) \prod_p \beta_p + o(N^d) \tag{2.8}$$

(under the assumption of Theorem 2.20). Note that we have

$$\log \prod_{p>w} \beta_p = \sum_{p>w} O(p^{-2}) \to 0$$

as $N \to \infty$. This implies

$$\prod_p \beta_p = (1+o(1)) \prod_{p \le w} \beta_p = (1+o(1))\beta_W$$

using the multiplicativity of the local factors. Since $vol_d(K) = O(N^d)$, this implies

$$vol_d(K) \prod_p \beta_p = vol_d(K)\beta_W + o(N^d). \tag{2.9}$$

Now define

$$A := \{a \in [W]^d : (\psi_i(a), W) = 1 \ \forall i \in [t]\}.$$

We have

$$\beta_W = \mathbb{E}_{a \in \mathbb{Z}_W^d} \prod_{i \in [t]} \Lambda_{\mathbb{Z}_W}(\psi_i(a)) = \left(\frac{W}{\varphi(W)}\right)^t \frac{|A|}{W^d}. \tag{2.10}$$

Applying this to (2.9), we can write

$$vol_d(K) \prod_p \beta_p = vol_d(K) \left(\frac{W}{\varphi(W)}\right)^t \frac{|A|}{W^d} + o(N^d). \tag{2.11}$$

Moreover, Lemma 2.12 ensures $\beta_W \ll 1$, and together with (2.10) we obtain

$$|A| \ll \left(\frac{\varphi(W)}{W}\right)^t W^d. \tag{2.12}$$

Now focusing on the other side of our equation of interest (2.8), note that

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda'(\psi_i(n)) = \sum_{a \in [W]^d} \sum_{\substack{n \in \mathbb{Z}^d \\ Wn+a \in K}} \prod_{i \in [t]} \Lambda'(\psi_i(Wn+a)). \tag{2.13}$$

But if $a \notin A$ then there is some $i \in [t]$ such that $(\psi_i(Wn+a), W) = (\psi_i(a), W) \neq 1$. By assumption, we have $\psi_i(Wn+a) > N^{4/5}$ for $Wn+a \in K$, which gives $\Lambda'(\psi_i(Wn+a)) = 0$ in this case. Hence, the sum may just range over $a \in A$. Using simple Euclidean division, we can then write

$$\psi_i(Wn+a) = W\tilde{\psi}_{i,a}(n) + b_i(a),$$

where $0 < b_i(a) < W$ is coprime to W and where $\tilde{\psi}_{i,a}(0) = O(N/W)$. With this notion, we have derived

$$\Lambda'(\psi_i(Wn+a)) = \frac{W}{\varphi(W)} \Lambda'_{b_i(a),W}(\tilde{\psi}_{i,a}(n)).$$

Plugging this into the right-hand side of (2.13), we obtain

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]}\Lambda'(\psi_i(n)) = \left(\frac{W}{\varphi(W)}\right)^t \sum_{a\in A}\sum_{\substack{n\in\mathbb{Z}^d\\ Wn+a\in K}}\prod_{i\in[t]}\Lambda'_{b_i(a),W}(\tilde{\psi}_{i,a}(n)). \tag{2.14}$$

We are now in a position to apply Theorem 2.20 with $\tilde{\Psi}_a = (\tilde{\psi}_{1,a},\dots,\tilde{\psi}_{t,a})$ as well as $\tilde{N} = N/W$ and $\tilde{K} = (K-a)/W$ for each $a \in A$; the reader may convince himself that all assumptions are indeed satisfied. In doing so, we obtain that for all $a \in A$ we have

$$\sum_{\substack{n\in\mathbb{Z}^d\\ Wn+a\in K}}\left(\prod_{i\in[t]}\Lambda'_{b_i(a),W}(\tilde{\psi}_{i,a}(n)) - 1\right) = o\left(\left(\frac{N}{W}\right)^d\right) \tag{2.15}$$

Inserting this into (2.14) and recalling (2.12) gives

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]}\Lambda'(\psi_i(n)) = \left(\frac{W}{\varphi(W)}\right)^t \sum_{a\in A}\sum_{\substack{n\in\mathbb{Z}^d\\ Wn+a\in K}} 1 + o(N^d). \tag{2.16}$$

But (2.2) applied to $\tilde{K}$ tells us that

$$\sum_{\substack{n\in\mathbb{Z}^d\\ Wn+a\in K}} 1 = \frac{vol_d(K)}{W^d} + o\left(\left(\frac{N}{W}\right)^d\right),$$

from which we can conclude, together with (2.16) and (2.12), that

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]}\Lambda'(\psi_i(n)) = vol_d(K)\left(\frac{W}{\varphi(W)}\right)^t \frac{|A|}{W^d} + o(N^d). \tag{2.17}$$

Comparing this to (2.11), the claim follows. □

Note that the local factors are gone after passing from Λ to $\Lambda'_{b,W}$. Morally speaking, the latter function is more regular with respect to small primes. This is essentially because w grows slowly with N so that we can ignore primes $p \le w$, but it grows slow enough so that this doesn't change the function too much. Later, we will have to invert this trick however, and pass from $\Lambda'_{b,W}$ back to Λ. Note also that the right-hand side is independent of the b_i.

It is a simple task to reduce the last Theorem further to the following

Theorem 2.21. *Let $N \ge 1$, $s \ge 1$, and let $\Psi = (\psi_1,\dots,\psi_t)$ be a system of affine-linear forms in s-normal form satisfying $\|\Psi\|_N = O(1)$. Let $K \subseteq [-N,N]^t$ be any convex body on which $\psi_1,\dots,\psi_t > N^{4/5}$. Then for any $b_1,\dots,b_t \le W$ which are coprime to W, we have*

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]}(\Lambda'_{b_i,W}(\psi_i(n)) - 1) = o(N^d). \tag{2.18}$$

The proof that Theorem 2.21 implies Theorem 2.20 is a simple matter of expanding the product in 2.20 after writing $\Lambda'_{b_i,W} = (\Lambda'_{b_i,W} - 1) + 1$ and repeatedly applying 2.21 to each summand. The details are left to the reader.

2.4 Pseudorandom measures, and reduction of the Main Theorem to a Gowers norm estimate

We have introduced pseudorandom measures in the last chapter, and explained that the function $\tilde{\Lambda} = \Lambda'_{1,W}$ can be bounded by such a pseudorandom measure in Proposition 1.12. This result generalises in the following way:

Proposition 2.22. *Let $D > 1$ be arbitrary. Then there is a constant $C_0 = C_0(D)$ such that the following holds. Let $C \geq C_0$, and let $N' \in [CN, 2CN]$ be a prime. Let $0 < b_1, \dots, b_t < W$ be integers coprime to W. Then there exists a D-pseudorandom measure $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ which obeys the pointwise bounds*

$$1 + \Lambda'_{b_1,W}(n) + \cdots + \Lambda'_{b_t,W}(n) \ll_{D,C} \nu(n) \tag{2.19}$$

for all $n \in [N^{3/5}, N]$, where n is identified as an element of $\mathbb{Z}_{N'}$ in the obvious manner.

We also analysed the connections of pseudorandom measures to the Gowers norm in Proposition 1.29. The next proposition is a generalisation of this in the obvious way, and we are not going to prove it. The further effort necessary for this form of the Proposition is of a very technical nature and not of interest to us here.

Proposition 2.23. *There are constants C_1 and D (depending on our parameters s, t, d and L) such that the following is true. Let $C_1 \leq C = O(1)$ be arbitrary and $N' \in [CN, 2CN]$ be a prime. Let $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ be a D-pseudorandom measure and suppose that $f_1, \dots, f_t : [N] \to \mathbb{R}$ are functions satisfying $0 \leq |f_i(x)| \leq \nu(x)$ for all $i \in [t]$ and $x \in [N]$. Let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms in s-normal form such that $\|\Psi\|_N \leq L$. Moreover, let $K \subseteq [-N, N]^d$ be a convex body with $\Psi(K) \subseteq [N]^t$. Suppose that*

$$\min_{1 \leq j \leq t} \|f_j\|_{U^{s+1}[N]} \leq \delta \tag{2.20}$$

for some $\delta > 0$. Then we have

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} f_i(\psi_i(n)) = o_{\delta,C}(N^d) + o_C(\delta) N^d, \tag{2.21}$$

where the last o-notation is viewed in the limit as $\delta \to 0$.

This proposition allows us to reduce the Main Theorem further, and obtain a statement about the Gowers norm of $\Lambda'_{b,W} - 1$ which does in fact not depend on the affine-linear form Ψ at all anymore.

Theorem 2.24. *Let $N \geq 1$, $s \geq 1$ and let $b \leq W$ be coprime to W. Then we have*

$$\|\Lambda'_{b,W} - 1\|_{U^{s+1}[N]} = o_s(1). \tag{2.22}$$

Proof of the Main Theorem assuming Theorem 2.24. It suffices to establish Theorem 2.21, so let the assumptions be as in that theorem. Since $\|\Psi\|_N = O(1)$, we may assume

that $\Psi(K) \subseteq [N]^t$ by enlarging N by a factor of $O(1)$ if necessary. Let D be as in 2.23, and let $C = C(D) := \max(C_0(D), C_1)$, where C_0 is as in Proposition 2.22 and C_1 is as in Proposition 2.23. Let also $N' \in [CN, 2CN]$ be a prime. Then Proposition 2.22 tells us that there is a D-pseudorandom measure $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ such that (2.19) holds for $n \in [N^{3/5}, N]$. In particular, there is $c > 0$ (depending on C) such that $f_i(n) := c(\Lambda'_{b_i,W} - 1)$ is pointwise bounded in magnitude by ν.

Note that there is a small technicality here: this only holds for $n \in [N^{3/5}, N]$. But if we define f_i to be 0 for smaller n, then the pointwise bound clearly holds. The reader may verify quickly that this change does not affect the Gowers norm too much, namely only by at most $N^{-2/5}$ times some logarithmic power.

An application of Theorem 2.24 implies that the assumptions of Proposition 2.23 are satisfied for any $\delta > 0$, and hence the claim. □

Our Main Theorem has finally become independent of any form Ψ, the parameters d, t and L, convex bodies, global or local factors and is now 'only' a statement concerning the Gowers norm of a modified von Mangoldt function. Nonetheless, it is still a very difficult result and we have to rely on two recent theorems, whose statements we try to explain now. Their proofs were performed in [9, 11].

2.5 The inverse Gowers norm and Möbius and nilsequences theorems

Definition 2.25. *Let G be a connected, simply connected, Lie group. We define the central series*

$$G = G_0 = G_1 \supseteq G_2 \supseteq G_3 \supseteq \dots$$

by $G_{i+1} := [G, G_i]$ for $i \geq 2$, where the commutator group $[G, H]$ of two (Lie) groups G, H is the group generated by $\{ghg^{-1}h^{-1} : g \in G, h \in H\}$. We say that G is s-step nilpotent if $G_{s+1} = \{e\}$.

Let $\Gamma \subseteq G$ be a discrete, cocompact (i.e. the quotient is compact) subgroup. Then the quotient G/Γ is called an s-step nilmanifold. If $g \in G$ then g acts on G/Γ by left multiplication $x \mapsto g \cdot x$.

By an s-step nilsequence we mean a sequence of the form $(F(g^n x))_{n \in \mathbb{N}}$, where $x \in G/\Gamma$ and $F : G/\Gamma \to \mathbb{R}$ is a continuous function. We say that the nilsequence is 1-bounded if $|F| \leq 1$.

Our goal is to show that the function $\Lambda'_{b,W} - 1$ has a small Gowers norm (i.e. is Gowers uniform). It turns out that a function which satisfies this property is not allowed to correlate with nilsequences, a statement we will make precise now:

Proposition 2.26. *Let $s \geq 1$ and $\delta \in (0, 1)$. Let $G/\Gamma = (G/\Gamma, d_{G/\Gamma})$ be an s-step nilmanifold with some smooth metric $d_{G/\Gamma}$, and let $((F(g^n x))_n$ be a bounded s-step nilsequence with Lipschitz constant at most M. Moreover, let $f : [N] \to [-1, 1]$ be a*

function which satisfies

$$\mathbb{E}_{n\in[N]}f(n)F(g^n x) \geq \delta$$

(i.e. f correlates with F). Then we have

$$\|f\|_{U^{s+1}[N]} \gg_{s,\delta,M,G/\Gamma} 1.$$

We will not prove this result; in fact, we will not use it either, but we do believe that it fits into the framework we are discussing, and makes the central concepts more understandable. The fundamental, and extremely deep, observation is that a converse of this statement is true - nilsequences are the only obstruction to uniformity:

Theorem 2.27 (Inverse Gowers norm Theorem)**.** *Let $\delta \in (0,1]$. Then there exists a finite collection $\mathcal{M}_{s,\delta}$ of s-step nilmanifolds with the following property. Given $N \geq 1$ and $f : [N] \to [-1,1]$ such that*

$$\|f\|_{U^{s+1}[N]} \geq \delta,$$

there is $G/\Gamma \in \mathcal{M}_{s,\delta}$ and a Lipschitz, 1-bounded s-step nilsequence $(F(g^n x))_n$ on it with Lipschitz constant $O_{s,\delta}(1)$ such that

$$|\mathbb{E}_{n\in[N]}f(n)F(g^n x)| \gg_{s,\delta} 1. \tag{2.23}$$

We will not proceed in any way to prove this Theorem. Its proof is very long and extremely difficult, and can be considered the most fundamental new contribution which made the proof of the Green-Tao Theorem possible.

We have just stated that any non-uniform function has to correlate with some nilsequence. The next theorem, also a very deep result, states that the Möbius function in fact does not correlate with nilsequences (and is far from that). This essentially tells us that the Möbius function in fact has a small Gowers norm. Since Number Theory gives us a relation between the Möbius and the Von Mangoldt function, we can hope to achieve a similar result for this function, and it is also plausible that this transfers to a similar estimate for our modified von Mangoldt function, thus implying the Main Theorem.

Theorem 2.28 (Möbius and nilsequences Theorem)**.** *Let G/Γ be an s-step nilmanifold, and let $(F(g^n x))_n$ be a bounded s-step nilsequence with Lipschitz constant M. Then for any $A > 0$ we have*

$$|\mathbb{E}_{n\in[N]}\mu(n)F(g^n x)| \ll_{A,M,G/\Gamma,s} \log^{-A} N. \tag{2.24}$$

We will now see how to apply these two Theorem to deduce the Main Theorem. We assume both the Gowers inverse norm theorem and the Möbius and nilsequences theorem in the following discussion.

2.6 Self-correlation estimates of the Möbius and Liouville functions

From our last discussion, we can deduce rather quickly the following self-correlation result of the Möbius and the Liouville function on affine-linear forms. Recall that the Liouville function λ is the unique completely multiplicative function that is -1 on the primes.

Proposition 2.29. *Let N, d, t, L, s be positive integers, and let $\Psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms of complexity at most s such that $\|\Psi\|_N \leq L$. Let $K \subseteq [-N, N]^d$ be a convex body. Then we have*

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \mu(\psi_i(n)) = o_s(N^d) \tag{2.25}$$

as well as

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \lambda(\psi_i(n)) = o_s(N^d). \tag{2.26}$$

Proof. Since μ and λ are bounded by 1, we can apply the generalised von Neumann Theorem 2.23 with $\nu \equiv 1$, which is D-pseudorandom for any value of D (possibly enlarging N by a factor $O_L(1)$ to ensure $\Psi(K) \subseteq [N]^t$). It thus suffices to show

$$\|\mu\|_{U^{s+1}[N]} = o_s(1) \tag{2.27}$$

and

$$\|\lambda\|_{U^{s+1}[N]} = o_s(1). \tag{2.28}$$

Now making use of the inverse Gowers norm Theorem 2.27, we can reduce the two statements further to showing that

$$\mathbb{E}_{n \leq N} \mu(n) F(g^n x) = o_{s,M,\delta}(1) \tag{2.29}$$

as well as

$$\mathbb{E}_{n \leq N} \lambda(n) F(g^n x) = o_{s,M,\delta}(1) \tag{2.30}$$

uniformly over $G/\Gamma \in \mathcal{M}_{s,\delta}$ and 1-bounded, M-Lipschitz nilsequences $(F(g^n x))_{n \leq N}$. Now (2.29) is a corollary of the Möbius and nilsequences Theorem 2.28.

To deduce 2.30, note that

$$\lambda(n) = \sum_{d^2 \mid n} \mu\left(\frac{n}{d^2}\right).$$

For positive real X, fixed $G/\Gamma \in \mathcal{M}_{s,\delta}$ and 1-bounded M-Lipschitz nilsequence $(F(g^n x))_{n \leq N}$ on G/Γ, we thus obtain

$$\begin{aligned}\mathbb{E}_{n \leq N} \lambda(n) F(g^n x) &= \mathbb{E}_{n \leq N} \sum_{d^2 \mid n} \mu\left(\frac{n}{d^2}\right) F(g^n x) \\ &= \sum_{d \leq X} \mathbb{E}_{n \leq N} \mathbb{1}_{d^2 \mid n} \mu\left(\frac{n}{d^2}\right) F(g^n x) + \sum_{d > X} \mathbb{E}_{n \leq N} \mathbb{1}_{d^2 \mid n} \mu\left(\frac{n}{d^2}\right) F(g^n x).\end{aligned}$$

But we have

$$\left|\sum_{d>X} \mathbb{E}_{n\leq N} \mathbb{1}_{d^2\mid n}\mu\left(\frac{n}{d^2}\right)F(g^n x)\right| \leq \sum_{d>X}\frac{1}{d^2} = O(X^{-1}),$$

so that

$$\begin{aligned}\mathbb{E}_{n\leq N}\lambda(n)F(g^n x) &= \sum_{d\leq X}\mathbb{E}_{n\leq N}\mathbb{1}_{d^2\mid n}\mu\left(\frac{n}{d^2}\right)F(g^n x) + O(X^{-1})\\ &= \sum_{d\leq X}\mathbb{E}_{k\leq N/d^2}\mu(k)F(g^{d^2 k}x) + O(X^{-1}).\end{aligned}$$

Application of the Möbius and nilsequences Theorem 2.28 with g replaced by g^{d^2} yields

$$\mathbb{E}_{k\leq N/d^2}\mu(k)F(g^{d^2 k}x) = o_{G/\Gamma,M,s}(1),$$

hence we obtain

$$\mathbb{E}_{n\leq N}\lambda(n)F(g^n x) = o_{G/\Gamma,M,X,s}(1) + O(X^{-1}).$$

Next, let $\varepsilon > 0$ and set $X := 1/\varepsilon$. Then by taking N sufficiently large, we can ensure that the left-hand side is $O_{G/\Gamma,M,s}(\varepsilon)$, hence the claim, recalling that $|\mathcal{M}_{s,\delta}| = O_{s,\delta}(1)$. □

2.7 The transference principle

Recall that we are trying to prove Theorem 2.24 using the Gowers inverse norm Theorem 2.27 and the Möbius and nilsequences Theorem 2.28. Their claims seem to fit together nicely, but looking at the assumptions, we need a function to be bounded in order to apply 2.27. This is another instance where the transference principle comes into place: Our next - and final - goal is to transfer the Gowers inverse norm Theorem from a statement for bounded functions to a statement regarding functions which are bounded by pseudorandom measures. The fundamental result in this direction is the following

Proposition 2.30 (Relative inverse Gowers norm Theorem)**.** *For $s \geq 1$, $\delta \in (0,1]$ and $C \geq 20$ there exists a finite collection $\mathcal{M}_{s,\delta,C}$ of nilmanifolds with the following property. Let $N \geq 1$, and let $N' \in [CN, 2CN]$ be a prime. Moreover, suppose that $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ is an $(s+2)$-pseudorandom measure and that $f : [N] \to \mathbb{R}$ is a function satisfying $|f(n)| \leq \nu(n)$ for all $n \in [N]$ as well as*

$$\|f\|_{U^{s+1}[N]} \geq \delta.$$

Then there exists $G/\Gamma \in \mathcal{M}_{s,\delta,C}$ together with a 1-bounded s-step nilsequence $(F(g^n x))_{n\in\mathbb{N}}$ with Lipschitz constant $O_{s,\delta,C}(1)$ such that

$$|\mathbb{E}_{n\leq N} f(n)F(g^n x)| \gg_{s,\delta,C} 1. \tag{2.31}$$

The next Proposition tells us that the function $\Lambda'_{b,W} - 1$ in which we are interested does not correlate with nilsequences. Together with the last proposition, this quickly gives the Main Theorem.

Proposition 2.31. *Let $s \geq 1$, and assume the Möbius and nilsequences Theorem MN(s). Let G/Γ be an s-step nilmanifold, and let $(F(g^n x))_n$ be a bounded s-step nilsequence with Lipschitz constant M. Let $b \leq W$ be coprime to W. Then*

$$\mathbb{E}_{n \leq N}(\Lambda'_{b,W}(n) - 1)F(g^n x) = o_{M,G/\Gamma,s}(1). \tag{2.32}$$

Proof of the Main Theorem assuming Proposition 2.30 and 2.31. It suffices to establish Theorem 2.24. To this end, let $C = \max(C_0, 20)$, where $C_0 = C_0(s+2)$ is the constant appearing in Proposition 2.22, and assume that the conclusion of Theorem 2.24 is false. Then we can find a subsequence of values of N going to infinity such that $\|\Lambda'_{b,W} - 1\|_{U^{s+1}[N]} \geq \delta$ for some $\delta \in (0, 1]$.

From Proposition 2.22 we get an $(s+2)$-pseudorandom measure ν such that $c|\Lambda'_{b,W}(n) - 1| \leq \nu(n)$ for some $c = c(s) > 0$. An application of Proposition 2.30 then implies that there is some nilmanifold $G/\Gamma \in \mathcal{M}_{s,\delta,C}$ and a 1-bounded s-step nilsequence $(F(g^n x))_{n \in \mathbb{N}}$ with Lipschitz constant $O_{s,\delta}(1)$ satisfying

$$|\mathbb{E}_{n \leq N}(\Lambda'_{b,W}(n) - 1)F(g^n x)| \gg_{s,\delta} 1.$$

But this contradicts Proposition 2.31, hence we have the claim. □

Our next goal will be to deduce Proposition 2.30, and then Proposition 2.31 gives the Main Theorem. To infer this, we need the following structure theorem:

Proposition 2.32. *Let $s \geq 1$ and let $N' \geq N \geq 1$ be integers. Suppose that $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ is an $(s+2)$-pseudorandom measure, and that $f : \mathbb{Z}_{N'} \to \mathbb{R}$ is a function such that $|f(n)| \leq \nu(n)$ for all n. Then we can decompose $f = f_1 + f_2$ in such a way that*

$$\|f_1\|_{L^\infty(\mathbb{Z}_{N'})} \leq 1 \tag{2.33}$$

and

$$\|f_2\|_{U^{s+1}(\mathbb{Z}_{N'})} = o(1). \tag{2.34}$$

Moreover, if f is supported inside $\{-N, \dots, N\}$ for some $N < N'/10$ then we can choose f_1 and f_2 in a way such that their support is contained in $\{-2N, \dots, 2N\}$.

Proof. We will only prove the first part. For the second part, the idea is essentially to multiply f, f_1 and f_2 by a smooth cutoff equal to 1 on $\{-N, \dots, N\}$ and vanishing outside $\{-2N, \dots, 2N\}$ and to verify that the new functions f_1 and f_2 still satisfy the same properties.

In the course of the proof, we will make heavy use of Proposition 1.40. Note that we assumed f to be non-negative in that Proposition, while we assume here that it is only bounded in absolute value by ν. Going through the proof, one may however verify

that we can handle the positive and the negative part of f separately, and at the end the result quickly follows for f as well.

Note also that it suffices to show

$$\|f_1\|_\infty \le 1 + o(1),$$

since we can always transfer the $o(1)$ part to f_2, noting that $\|g\|_{U^{s+1}(\mathbb{Z}_{N'})} \le \|g\|_{L^\infty(\mathbb{Z}_{N'})}$.

Let $\varepsilon > 0$ be sufficiently small, $N > N_0(\varepsilon)$ sufficiently large, and write

$$f = f_1 + f_2^{(1)} + f_2^{(2)},$$

where

$$\begin{aligned} f_1 &:= (1 - \mathbb{1}_\Omega)\mathbb{E}[f \,|\, \mathcal{B}], \\ f_2^{(1)} &:= (1 - \mathbb{1}_\Omega)(f - \mathbb{E}[f \,|\, \mathcal{B}]), \\ f_1^{(2)} &:= \mathbb{1}_\Omega f. \end{aligned}$$

Then Proposition 1.40 implies

$$\|f_1\|_{L^\infty(\mathbb{Z}_{N'})} \le 1 + o_\varepsilon(1)$$

and

$$\|f_2^{(1)}\|_{U^{s+1}(\mathbb{Z}_{N'})} \le \varepsilon^{1/2^{s+2}}$$

as well as

$$\|f_2^{(2)}\|_{L^1(\mathbb{Z}_{N'})} = o_\varepsilon(1).$$

Since we have $|f_2^{(2)}| \le \nu$ pointwise, we obtain

$$\begin{aligned} \|f_2^{(2)}\|_{U^{s+1}(\mathbb{Z}_{N'})}^{2^{s+1}} &= \mathbb{E}_{n\in\mathbb{Z}_{N'}, h\in\mathbb{Z}_{N'}^{s+1}}\Big[f_2^{(2)}(n) \prod_{\substack{\omega\in\{0,1\}^{s+1}\\ \omega\neq 0}} f_2^{(2)}(n+\omega\cdot h)\Big] \\ &\le \mathbb{E}_{n\in\mathbb{Z}_{N'}}\big[|f_2^{(2)}(n)|\big] \sup_{n\in\mathbb{Z}_{N'}} \Big(\mathbb{E}_{h\in\mathbb{Z}_{N'}^{s+1}}\Big[\prod_{\substack{\omega\in\{0,1\}^{s+1}\\ \omega\neq 0}} \nu(n+\omega\cdot h)\Big]\Big) \\ &= \|\mathcal{D}\nu\|_{L^\infty(\mathbb{Z}_{N'})}\|f_2^{(2)}\|_{L^1(\mathbb{Z}_{N'})}. \end{aligned}$$

Now, (1.24) implies $\|\mathcal{D}\nu\|_{L^\infty(\mathbb{Z}_{N'})} = O_s(1)$, and hence

$$\|f_2^{(2)}\|_{U^{s+1}(\mathbb{Z}_{N'})} = o_{\varepsilon,s}(1).$$

As a consequence, we obtain that $f_2 := f_2^{(1)} + f_2^{(2)}$ satisfies

$$\|f_2\|_{U^{s+1}(\mathbb{Z}_{N'})} \le o_{\varepsilon,s}(1) + \varepsilon^{1/2^{s+2}}.$$

For a given $\varepsilon' > 0$, we can then choose ε small enough such that the second term is $< \varepsilon'/2$. Then, we can choose N large enough such that the first part is $< \varepsilon'/2$. The claim follows. □

We are now able to apply this result in order to deduce Proposition 2.30.

Proof of Proposition 2.30. Applying Proposition 2.32, we can write

$$f = f_1 + f_2,$$

where $\|f_1\|_{L^\infty(\mathbb{Z}_{N'})} \leq 1$ and $\|f_2\|_{U^{s+1}(\mathbb{Z}_{N'})} = o(1)$, and where we may assume both to have support contained in $\{-2N, \dots, 2N\}$. By the comparability of the $U^{s+1}[N]$ and $U^{s+1}(\mathbb{Z}_{N'})$, Lemma 1.27, the assumption $\|f\|_{U^{s+1}[N]} \geq \delta$ implies $\|f\|_{U^{s+1}(\mathbb{Z}_{N'})} \gg_{C,s} \delta$, and hence we also have $\|f_1\|_{U^{s+1}(\mathbb{Z}_{N'})} \gg_{C,s} \delta$. Transferring this back to $\{-2N, \dots, 2N\}$ with the same Lemma, we have derived $\|f_1\|_{U^{s+1}(\{-2N,\dots,2N\})} \gg_{C,s} \delta$.

Translating $\{-2N, \dots, 2N\}$ to $[4N+1]$, the inverse Gowers norm Theorem 2.27 gives the existence of an s-step nilmanifold $G/\Gamma \in \mathcal{M}_{s,\delta,C}$ together with a 1-bounded s-step nilsequence $(F(g^n x))_{n\in\mathbb{N}}$ on G/Γ with Lipschitz constant $O_{s,\delta,C}(1)$ such that

$$|\mathbb{E}_{-2N\leq n\leq 2N} f_1(n)F(g^n x)| \gg_{s,\delta,C} 1.$$

At the same time we have $\|f_2\|_{U^{s+1}(\mathbb{Z}_{N'})} = o(1)$, and Proposition 2.26 therefore gives

$$|\mathbb{E}_{-2N\leq n\leq 2N} f_2(n)F(g^n x)| = o_{G/\Gamma,s,\delta,C}(1).$$

If $N \geq N_0 = N_0(s, \delta, C)$ then this implies

$$|\mathbb{E}_{-2N\leq n\leq 2N} f(n)F(g^n x)| \gg_{s,\delta,C} 1,$$

noting that f vanishes outside $[N]$. If on the other hand $N < N_0 = O_{s,\delta,C}(1)$ then the claim is trivial: All norms on $[N]$ are equivalent up to factor $O_N(1) = O_{s,\delta,C}(1)$ (this is a more or less empty statement), and all functions on $[N]$ can be expressed as nilsequences on $\mathbb{R}/\mathbb{Z}$ with Lipschitz constant $O_N(1) = O_{s,\delta,C}(1)$. In particular, we can set $F(g^n x) = f(n)$, and then we have

$$|\mathbb{E}_{n\leq N} f(n)F(g^n x)| = \|f\|^2_{L^2[N]} \gg_{s,C,\delta} \|f\|^2_{U^{s+1}[N]} \gg_{s,C,\delta} 1.$$

□

Summarising what we have explained up to this point, it suffices to obtain Proposition 2.31, which essentially claims that the function $\Lambda'_{b,W} - 1$ does not correlate with nilsequences. We will soon decompose the von Mangoldt function Λ into a 'smooth' and a 'rough' part. This decomposition will induce a further decomposition on the function $\Lambda_{b,W}$, which we expect to be close to $\Lambda'_{b,W}$ in any reasonable sense. We will then try to separately verify that both the smooth and the rough part of $\Lambda_{b,W}$ do not correlate with nilsequences.

However, it turns out that it is only possible to bound the smooth component in its Gowers norm, which might at first appear to be suitable for application of Proposition 2.26; but it is not a bounded function, as required. In fact, it appears to be difficult to even bound it by a pseudorandom measure (with the idea of using some type of transference principle for 2.26)! For this reason, we introduce the notion of averaged nilsequences, which have more regularity, and we will be able to bound them in their Gowers dual norm.

Definition 2.33. *Let G/Γ be an s-step nilmanifold, and let $M > 0$. An s-step averaged nilsequence on G/Γ with Lipschitz constant at most M is a function $F(n)$ of the form*

$$F(n) = \sum_{i \in I} F_i(g_i^n x_i),$$

where I is some finite index set, and for each i, we have that $F_i(g_i^n x_i)$ is a bounded s-step nilsequence on G/Γ with Lipschitz constant at most s.

The precise use of this notion will become apparent soon; essentially, it is the following technical proposition, which we will not prove.

Proposition 2.34. *Let G/Γ be an s-step nilmanifold, and let $M > 0$. Suppose that $(F(g^n x))_{n \in \mathbb{N}}$ is a bounded s-step nilsequence on G/Γ with Lipschitz constant at most M. Let $\varepsilon \in (0,1)$ and $N \geq 1$. Then we can decompose*

$$F(g^n x) = F_1(n) + F_2(n), \tag{2.35}$$

where $F_1 : \mathbb{N} \to [-1,1]$ is an averaged nilsequence on $(G/\Gamma)^{2^{s+1}-1}$ with Lipschitz constant $O_{M,\varepsilon,G/\Gamma}(1)$ and

$$\|F_1\|_{U^{s+1}[N]^*} \ll_{M,\varepsilon,G/\Gamma} 1,$$

while $F_2 : \mathbb{N} \to \mathbb{R}$ satisfies

$$\|F_2\|_{L^\infty} = O(\varepsilon).$$

We try to illustrate the core idea of the proof: In some way, we need to determine what the special feature of nilmanifolds and nilsequences is.

Given a nilmanifold G/Γ, we write $(G/\Gamma)^{\{0,1\}^{s+1}}$ for the set of all 2^{s+1}-tuples $(x_\omega)_{\omega \in \{0,1\}^{s+1}}$. We call an element of $(G/\Gamma)^{\{0,1\}^{s+1}}$ an $(s+1)$-dimensional parallelepiped if it is of the form $(g^{n+\omega \cdot h})_{\omega \in \{0,1\}^{s+1}}$ for some $g \in G$, $x \in G/\Gamma$, $n \in \mathbb{Z}$ and $h \in \mathbb{Z}^{s+1}$.

Now the fundamental idea is that for a given vertex of such a parallelepiped, its value is determined in a continuous way by the values of its other $2^{s+1} - 1$ vertices. This constraint allows one to deduce the structural properties necessary for proving the claim.

Example 2.35. The easiest example of a nilmanifold is the torus $\mathbb{R}/\mathbb{Z}$, which is a 1-step nilmanifold. In this case, we write $g^n x$ as $x + gn$ for $g \in \mathbb{R}$, $x \in \mathbb{R}/\mathbb{Z}$ and $n \in \mathbb{N}$. A 2-dimensional parallelepiped on the torus is an element of $(\mathbb{R}/\mathbb{Z})^4$ of the form

$$(x + ng, x + (n + h_1)g, x + (n + h_2)g, x + (n + h_1 + h_2)g).$$

Denoting the components by $(y_{00}, y_{10}, y_{01}, y_{11})$, one immediately sees that

$$y_{00} = y_{10} + y_{01} - y_{11},$$

meaning that one vertex of this parallelepiped is continuously determined by the other vertices. For higher-order nilmanifolds, this is of course much more difficult to verify.

With Proposition 2.34 in our pocket, we are now able to reduce the Main Theorem even further; note that we have by now reduced it to showing Proposition 2.31.

Proposition 2.36. *Let $s \geq 1$, and let G/Γ be an s-step nilmanifold. Further, let $F_1(n)$ be an averaged s-step nilsequence with Lipschitz constant M, and let $b \leq W$ be coprime to W. Suppose the dual norm bound*

$$\|F_1\|_{U^{s+1}[N]^*} \leq M'. \tag{2.36}$$

Then we have

$$\mathbb{E}_{n\leq N}\Big[(\Lambda'_{b,W}(n)-1)F_1(n)\Big] = o_{M,M',G/\Gamma,s}(1). \tag{2.37}$$

Proof of the Main Theorem assuming Proposition 2.36. It suffices to establish Proposition 2.31, so let F be as in that Proposition. Let $\varepsilon \in (0,1)$ be arbitrary, and decompose F as in Proposition 2.34. By Proposition 2.36, the contribution of F_1 is $o_{M,G/\Gamma,s,\varepsilon}(1)$. For the contribution of F_2, we can bound

$$|\mathbb{E}_{n\leq N}(\Lambda'_{b,W}(n)-1)F_2(n)| \leq \|F_2\|_{L^\infty}(\mathbb{E}_{n\leq N}[\Lambda'_{b,W}(n)]+1) = O(\varepsilon)$$

using the Siegel-Walfisz Theorem. As a consequence, we have

$$\mathbb{E}_{n\leq N}(\Lambda'_{b,W}(n)-1)F(g^n x) = o_{M,G/\Gamma,s,\varepsilon}(1) + O(\varepsilon),$$

which gives the claim. □

2.8 A splitting of the Von Mangoldt function

Recall that our goal now is to establish Proposition 2.36 to show the Main Theorem, namely

$$\mathbb{E}_{n\leq N}(\Lambda'_{b,W}(n)-1)F_1(n) = o_{M,M',G/\Gamma,s}(1).$$

To show this, one easily verifies that it suffices to show the same statement with $\Lambda'_{b,W}$ replaced by $\Lambda_{b,W}$. Writing this out, our goal is to prove the following estimate:

$$\mathbb{E}_{n\leq N}\left[\Big(\frac{\varphi(W)}{W}\Lambda_{b,W}(Wn+b)-1\Big)F_1(n)\right] = o_{M,M',G/\Gamma,s}(1). \tag{2.38}$$

Let $\gamma = \gamma_s > 0$ to be determined later, and define $R := N^\gamma$. Note that we have

$$\Lambda(n) = -\log R \sum_{s\,|\,n}\mu(d)\frac{\log d}{\log R}.$$

Next, we make a smooth decomposition $x = \chi^s(x) + \chi^r(x)$ of the identity function on $\mathbb{R}^+$, where $\chi^r(x)$ vanishes for $x \geq 1$ and $\chi^s(x)$ vanishes for $x \leq 1/2$. With this, we can define

$$\Lambda^s(n) := -\log R \sum_{d\,|\,n}\mu(d)\chi^s\left(\frac{\log d}{\log R}\right), \tag{2.39}$$

$$\Lambda^r(n) := -\log R \sum_{d\,|\,n}\mu(d)\chi^r\left(\frac{\log d}{\log R}\right), \tag{2.40}$$

so that $\Lambda(n) = \Lambda^s(n) + \Lambda^r(n)$. It then suffices to establish the two estimates

$$\mathbb{E}_{n\leq N}\Big[\Big(\frac{\varphi(W)}{W}\Lambda^s(Wn+b)-1\Big)F_1(n)\Big] = o_{s,M'}(1) \tag{2.41}$$

and

$$\mathbb{E}_{n\leq N}\Big[\frac{\varphi(W)}{W}\Lambda^r(Wn+b)F_1(n)\Big] = o_{M,G/\Gamma,s}(1), \tag{2.42}$$

since they imply (2.38) and therefore the Main Theorem. Let us first focus on establishing (2.41). From the dual norm bound (2.36), we obtain

$$\begin{aligned}\Big|\mathbb{E}_{n\leq N}\Big[\Big(\frac{\varphi(W)}{W}\Lambda^s(Wn+b)-1\Big)F_1(n)\Big]\Big| &\leq \Big\|\frac{\varphi(W)}{W}\Lambda^s(Wn+b)-1\Big\|_{U^{s+1}[N]}\|F_1\|_{U^{s+1}[N]^*}\\ &\leq M'\Big\|\frac{\varphi(W)}{W}\Lambda^s(Wn+b)-1\Big\|_{U^{s+1}[N]}.\end{aligned}$$

As a consequence, it is sufficient to verify that

$$\Big\|\frac{\varphi(W)}{W}\Lambda^s(Wn+b)-1\Big\|_{U^{s+1}[N]} = o_s(1),$$

which, by writing out the definition of the $U^{s+1}[N]$-norm, is in turn implied by the more general bound

$$\sum_{(n,h)\in K}\prod_{\omega\in\{0,1\}^{s+1}}\left(\frac{\varphi(W)}{W}\Lambda^s(W(n+\omega\cdot h)+b)-1\right) = o(N^{s+2})$$

for any convex body $K \subseteq [-N,N]^{s+2}$. Expanding out the product on the left-hand side, we can reduce this further to showing that

$$\sum_{(n,h)\in K}\prod_{\omega\in B}\frac{\varphi(W)}{W}\Lambda^s(W(n+\omega\cdot h)+b) = vol_{s+2}(K) + o(N^{s+2})$$

for any $B \subseteq \{0,1\}^{s+1}$. Since Λ^s is a truncated von Mangoldt function and the nilsequences have disappeared in the above claim, this is accessible by much more standard sieve theory techniques, but we note that it still needs considerable efforts. The details can be found in appendix D of [8].

We now turn to the estimate (2.42). By the triangle inequality, it suffices to show

$$\mathbb{E}_{n\leq N}\frac{\varphi(W)}{W}\Lambda^r(Wn+b)F(g^nx) = o_{M,G/\Gamma,s}(1) \tag{2.43}$$

uniformly over g, x and 1-bounded, s-step nilsequences $(F(g^nx))_n$ with Lipschitz constant at most M, since we can average over such nilsequences to obtain F_1. We can trivially bound $\varphi(W)/W \leq 1$ and thus remove this factor. Next, we can write

$$\mathbb{E}_{n\leq N}\Lambda^r(Wn+b)F(g^nx) = W\mathbb{E}_{b<n\leq Wn+b}\mathbb{1}_{n\equiv b\,(W)}\Lambda^r(n)F(g^{(n-b)/W}x). \tag{2.44}$$

We know that the exponential map on a nilpotent Lie group is surjective, and one can show that the Lie group is in fact divisible, i.e. for any $g \in G$ and $m \in \mathbb{N}$ there exists an element $g^{1/m}$ such that $(g^{1/m})^m = g$. Thus, we define $g' := g^{1/W}$ and $x' := g^{-b/W}x$, so that

$$F(g'^n x'^n) = F(g^{(n-b)/W}x).$$

The point is that the left-hand side is defined for any integer n. Moreover, we can expand

$$\mathbb{1}_{n \equiv b(W)} = \frac{1}{W} \sum_{r(W)} e\left(\frac{r(n-b)}{W}\right).$$

Substituting both identities into (2.44) gives

$$\mathbb{E}_{n \le N} \Lambda^r(Wn+b) F(g^n x) = \mathbb{E}_{b < n \le WN+b} \left[\sum_{r(W)} e\left(\frac{r(n-b)}{W}\right) \Lambda^r(n) F(g'^n x') \right].$$

Clearly, $n \mapsto e(r(n-b)/W)$ induces a 1-bounded, $O(1)$-Lipschitz nilsequence on the 1-step nilmanifold $\mathbb{R}/\mathbb{Z}$. It hence suffices to show that

$$W \mathbb{E}_{b < n \le WN+b} \big[\Lambda^r(n) F(g^n x) \big] = o_{M,G/\Gamma,s}(1) \tag{2.45}$$

for all M-Lipschitz, 1-bounded nilequences $(F(g^n x))_{n \in \mathbb{N}}$ on the s-step nilmanifold G/Γ, noting the importance of the uniformity with respect to g, x and F. We will now show, that in fact

Lemma 2.37.

$$\left| \sum_{n \le N} \Lambda^r(n) F(g^n x) \right| \ll_{M,G/\Gamma,s,A} N \log^{-A} N \tag{2.46}$$

holds for any $A > 0$.

Proof. By the choice of w, this easily implies (2.45). By definition of Λ^r, we can rewrite and then rearrange the left-hand side as

$$\begin{aligned} \left| \sum_{n \le N} \Lambda^r(n) F(g^n x) \right| &= \left| \sum_{n \le N} \sum_{d \mid n} \mu(d) \chi^r\left(\frac{\log d}{\log R}\right) F(g^n x) \right| \\ &= \left| \sum_{m \le N} \sum_{d \le N/m} \mu(d) \chi^r\left(\frac{\log d}{\log R}\right) F\big((g^m)^d x\big) \right|. \end{aligned}$$

Note that, since χ^r vanishes on $[0, 1/2]$, we only have a contribution from terms with $d \ge R^{1/2}$, so that $m \le N/R^{1/2}$. Application of the summation by parts formula and the

Möbius and nilsequences Theorem then yields

$$\begin{aligned}
&\left|\sum_{d\le N/m}\mu(d)\chi^r\left(\frac{\log d}{\log R}\right)F\big((g^m)^d x\big)\right|\\
&\le\left|\chi^r\left(\frac{\log(N/m)}{\log R}\right)\sum_{d\le N/m}\mu(d)F\big((g^m)^d x\big)\right|\\
&+\left|\int_1^{N/m}\frac{d}{dt}\left(\chi^r\left(\frac{\log t}{\log R}\right)\right)\sum_{1\le d\le t}\mu(d)F\big((g^m)^d x\big)\,dt\right|\\
&\ll_{A,M,G/\Gamma,s}\log(N/m)\log^{-A}(N/m)+\int_{R^{1/2}}^{N/m}\log^{-A}t\,dt\ll_A\frac{N}{m}\log^{-A}(N/m).
\end{aligned}$$

Again, we make use of the uniformity with respect to g to apply the Möbius and nilsequences Theorem to g^m instead of g. Since $m\le N/R^{1/2}$, one easily verifies that $\log^{-A}(N/m)\ll_A\log^{-A}N$. Summing the above over all m in question implies, putting the last steps together, that

$$\left|\sum_{n\le N}\Lambda^r(n)F(g^n x)\right|\ll_{A,M,G/\Gamma,s}\sum_{m\le N}\frac{N}{m}\log^{-A}N.$$

Since this holds for any $A>0$, the claim follows. This concludes the proof of the Main Theorem. □

The applications of the Gowers inverse norm Theorem and of pseudorandom measures are not limited to the von Mangoldt function (or rather the derived functions we analysed here), but can be extended to many other types of functions as soon as one can bound them by pseudorandom measures. One example of such an approach is [12] (among several other papers by the same author), where one looks at the function which counts the number of representations of an integer by a positive definite binary quadratic form.

For a given collection $f_1,\dots,f_t$ of positive definite binary quadratic forms, denoting

$$R_{f_i}(n):=|\{(x,y):f_i(x,y)=n\}|$$

one can deduce an asymptotic for the expression

$$\mathbb{E}_{n\in K\cap\mathbb{Z}^d}\prod_{i=1}^t R_{f_i}(\psi_i(n))$$

(note the similarities of this expression with the one in the Generalised Hardy-Littlewood conjecture 2.4). The very basic idea is indeed to bound the representation functions by pseudorandom measures similar to the methods we employed in this chapter.

References

[1] Furstenberg, H. Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.*, 31 (1977), 204–256.

[2] Furstenberg, H., Katznelson, Y., and Ornstein, D. The ergodic-theoretical proof of Szemerédi's theorem. *Bull. Amer. Math. Soc.*, 7 (1982), 527–552.

[3] Gowers, W. T. A new proof of Szemerédi's Theorem for arithmetic progressions of length four. *GAFA*, 8 (1998), 529–551.

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

[5] Green, B. J., and Tao, T. C. Quadratic uniformity of the Möbius function. *Annales de l'Institut Fourier*, 58.6 (2006), 1863–1935.

[6] Green, B. J., and Tao, T. C. An inverse theorem for the Gowers $U^3(G)$ norm. *Proc. Edinburgh Math. Soc. no. 1*, 51 (2008), 73–153.

[7] Green, B. J., and Tao, T. C. The primes contain arbitrarily long arithmetic progressions. *Annals of Math.*, 167 (2008), 481–547.

[8] Green, B. J., and Tao, T. C. Linear Equations in Primes. *Annals of Math.*, 171 (2010), 753–850.

[9] Green, B. J., and Tao, T. C. The Möbius function is strongly orthogonal to nilsequences. *Annals of Math.*, 175 (2012), 541–566.

[10] Green, B. J., and Tao, T. C. The quantitative behaviour of polynomial orbits on nilmanifolds. *Annals of Math.*, 175 (2012), 465–540.

[11] Green, B. J., Tao, T. C., and Ziegler, T. An inverse theorem for the Gowers $U^{s+1}[N]$ norm. *Annals of Math. no. 2*, 176 (2012), 1231–1372.

[12] Matthiesen, L. Linear correlations amongst numbers represented by positive definite binary quadratic forms. *Acta Arith.*, 154 (2012), 235–306.

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