Primes in Arbitrarily Long Arithmetic Progression

By Zhuo (Aubrey) Yang*

Advisor: Yum-Tong Siu and Leslie G. Valiant

March 29, 2012

1 Abstract

It has been a long conjecture that there are arbitrarily long arithmetic progressions of primes. As of now, the longest known progression of primes is of length 26 and was discovered by Benoat Perichon and PrimeGrid in April, 2010 ([1]): $43142746595714191+23681770\cdot 223092870n$ for $n = 0, 1, \cdots, 25$. Many mathematicians have spent years trying to prove (or disprove) this conjecture, and even more ambitiously, to come up with an explicit algorithm of finding arithmetic progressions of primes of length k for any k. The recent Green-Tao Theorem (2004) proves this conjecture. It states that the prime numbers contain infinitely many arithmetic progressions of length k for any k. There are two main stages in the proof. The first stage is a generalized version of Szemerédi's Theorem, which asserts that any subset of the positive integers with positive density relative to a sufficiently pseudorandom measure contains arbitrarily long arithmetic progressions. The second stage is to use the argument by Goldston and Yıldırım to construct a pseudorandom measure applicable to our problem, and use a form of Dirichlet's Theorem to conclude that the primes contain arbitrarily long arithmetic progressions. In this paper, we will present the proof of the Green-Tao Theorem, with the focus on proving the generalized Szemerédi's Theorem. The technique is to decompose a function into a *Gowers uniform* (error) component and *Gowers anti-uniform* (structured) component, and then estimate each part separately. The first half of the paper involves the study of the Gowers uniformity norms from a combinatorial point of view, and the second half of the paper investigates the Furstenberg tower from a ergodic theoretic point of view. In the end we will briefly discuss Szemerédi's theorem itself, which can be proved by several methods.

*This is an undergraduate thesis submitted to the Harvard University Department of Computer Science in partial fulfillment of the requirements for the joint degree of A.B. in Mathematics and Computer Science.

2 Introduction

2.1 Background

There are many conjectures involving the primes both historically and more recently. For instance, it has been a long conjecture that there are infinitely many twin primes. More recently in 1923, Hardy and Littlewood made a very general conjecture which predicts that there are infinitely prime k-tuples of the form $p, p+a_1, \cdots, p+a_k$ for any positive integers $a_1, \cdots, a_k$, unless a trivial divisibility condition holds. Let us also mention a conjecture that there are arbitrarily long arithmetic progressions of *consecutive* primes, which if proven to be true, would contain the Green-Tao Theorem in particular. The first theoretical development on the conjecture of arithmetic progressions of primes was by van der Corput, who used analytic methods to prove that there are infinitely arithmetic progressions of primes of length 3. The recent Green-Tao Theorem proved the conjecture for any k. However, it must be noted that the Green-Tao Theorem is merely an existence theorem; it does not give an explicit algorithm of finding such progressions of primes.

2.2 Basic Idea

As in almost any problems involving the primes, the first step is to transform the conjecture into an analytic problem. We will adopt the following notation throughout the paper: Let $\mathbb{P}$ denote the set of primes. Let k be the length the arithmetic progression of primes whose existence we try to prove. Throughout the paper, k will stay fixed. Let N be a large prime compared to k, and define $\mathbb{Z}_N := \mathbb{Z}/N\mathbb{Z}$. For $f : \mathbb{Z}_N \to \mathbb{R}^+$, define $\mathbb{E}(f(x)|x \in \mathbb{Z}_N) := \frac{1}{N}\sum_{x=0}^{N-1} f(x)$ in the usual manner. Note that if we carefully choose f such that $\text{supp}(f) \subset \mathbb{P}$, and form the product $f(x)f(x+r)\cdots f(x+(k-1)r)$, then the product is nonzero only when each of $x, x+r, \cdots, x+(k-1)r$ is in $\text{Supp}(f) \subset \mathbb{P}$, and thus $x, x+r, \cdots, x+(k-1)r$ is a progression of k primes. Hence it will suffice to prove that $\mathbb{E}(f(x)f(x+r)\cdots f(x+(k-1)r)|x, r \in \mathbb{Z}_N)$ is positive when N is large. In fact, this shows something stronger: if $\mathbb{E}(f(x)f(x+r)\cdots f(x+(k-1)r)|x, r \in \mathbb{Z}_N) \geq c(k)$ as $N \to \infty$ where $c(k)$ is a constant depending on k but not on N, then there will be at least $c(k)N^2 - o_k(1)$ arithmetic progressions of k primes, all of which are smaller than N. Taking $N \to \infty$, it immediately follows that there are infinitely many arithmetic progressions of k primes. There is one caveat here: since the domain of f is $\mathbb{Z}_N$, it could be the case that $x, x+r, \cdots, x+(k-1)r$ is only an arithmetic progression in $\mathbb{Z}_N$ but not a genuine arithmetic progression. In the proof, we will resolve this issue by showing that each $x, x+r, \cdots, x+(k-1)r$ that contributes to $\mathbb{E}(f(x)f(x+r)\cdots f(x+(k-1)r)|x, r \in \mathbb{Z}_N)$ will be a genuine arithmetic progression, except for the case where $r = 0$ which will not affect our conclusion.

Notational Remark. Throughout the paper, unless otherwise specified, the 'big-oh' and 'little-oh' will always be with respect to $N \to \infty$.

3 Outline of the Proof

We are going to pursue the idea presented in the previous section. Roughly, there are 2 major steps: the first is to prove the result $\mathbb{E}(f(x)f(x+r)\cdots f(x+(k-1)r)|x,r\in\mathbb{Z}_N)\geq c(k)-o_k(1)$ whenever f has certain properties, and the second is to construct a specific function f and verify that it indeed has those properties. The properties here, as we shall see, are that f is bounded above by a pseudorandom measure (which we will define precisely) and the expectation of f is bounded below as $N\to\infty$.

3.1 Generalized Szemerédi's Theorem

The key theorem that we are going to quote is the following Szemerédi's Theorem:

Theorem 3.1 (Szemerédi's Theorem). *Let $\nu_{const}\equiv 1$ be the constant function $\mathbb{Z}_N\to\mathbb{R}$. Fix $0<\delta\leq 1$ and $k\geq 1$. Suppose $f:\mathbb{Z}_N\to\mathbb{R}^+$ satisfies*

$$0\leq f(x)\leq\nu_{const}(x)\text{ for all }x\in\mathbb{Z}_N$$

and

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

Then

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

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

We are not going to prove the above theorem in this paper. Interested readers can find Szemerédi's original combinatorial proof in [9] and Tao's recent ergodic theory proof in [10]. But since it plays such an important role in the main arguments of this paper, we will briefly discuss the idea of the proof in section 12. The main result that we are going to prove assuming this theorem is the following *Generalized Szemerédi's Theorem*:

Theorem 3.2 (Generalized Szemerédi's Theorem). *Fix $k\geq 3$ and $0<\delta\leq 1$. If $\nu:\mathbb{Z}_N\to\mathbb{R}^+$ is k-pseudorandom, and if $f:\mathbb{Z}_N\to\mathbb{R}^+$ satisfies*

$$0\leq f(x)\leq\nu(x)\text{ for all }x\in\mathbb{Z}_N$$

and

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

Then

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

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

Notice the only (but rather important) difference in Theorem 3.2 is that we replace the constant measure ν_{const} with a k-pseudorandom measure ν. In the next few sections, we will prove Theorem 3.2 assuming Theorem 3.1.

3.2 Application to Finding Arithmetic Progressions in Primes

After proving Theorem 3.2, we will discuss the following proposition:

Proposition 3.3 (Goldston-Yıldırım). *Let $w(N)$ be a fixed arbitrary function of N that goes to infinity slowly such that $1/w(N) = o(1)$. (For example, $w(N)\log w(N) = O(\log\log N)$ will suffice. In fact the $o(1)$ term is not necessary and $w(N)$ can actually be taken to be a large constant independent of N). Write $\epsilon_k := \frac{1}{2^k(k+4)!}$, and $R := N^{k^{-1}2^{-k-4}}$. Let $W := \prod_{p\leq w(N)} p$, $\Lambda_R(n) := \sum_{d|n,d\leq R}\mu(d)\log(R/d)$ where μ is the Mobius function. Let $\phi(\cdot)$ be the Euler totient function. Define $f : \mathbb{Z}_N \to \mathbb{R}^+$ as*

$$f(n) := \left\{ \begin{array}{cc} k^{-1}2^{-k-5}\frac{\phi(W)}{W}\log(Wn+1) & \text{if } Wn+1 \text{ is a prime and } \epsilon_k N \leq n \leq 2\epsilon_k N \\ 0 & \text{otherwise .} \end{array} \right\}$$

Define $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ as

$$\nu(n) := \left\{ \begin{array}{cc} \frac{\phi(W)}{W}\frac{\Lambda_R(Wn+1)^2}{\log(R)} & \text{if } \epsilon_k N \leq n \leq 2\epsilon_k N \\ 1 & \text{otherwise .} \end{array} \right\}$$

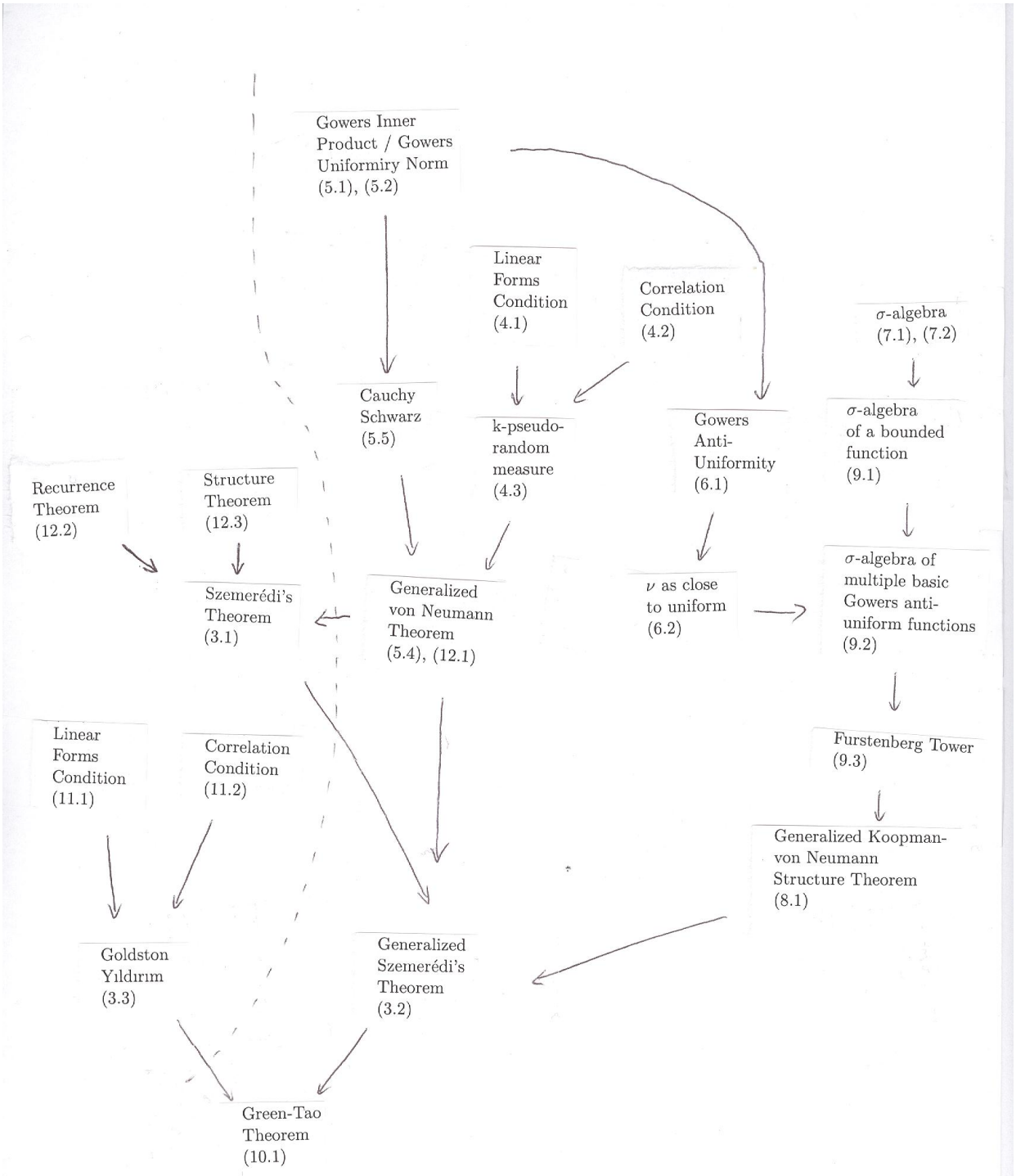
Then

$$0 \leq f(x) \leq \nu(x) \text{ for all } x \in \mathbb{Z}_N$$

$$\nu \text{ is a } k\text{-pseudorandom measure.}$$

Then we will apply Theorem 3.2 to the above choice of f and ν, from which it follows easily that the prime numbers contain an arithmetic progression of length k.

The following flow chart gives a pictorial description of the relationship among the main ideas in this paper. We will prove everything to the right of the dotted line, and only discuss the ideas in the proof of the left part because of limited space in this paper.


Gowers Inner Product / Gowers Uniformiry Norm (5.1), (5.2)
Linear Forms Condition (4.1)
Correlation Condition (4.2)
σ-algebra (7.1), (7.2)
Cauchy Schwarz (5.5)
k-pseudo-random measure (4.3)
Gowers Anti-Uniformity (6.1)
σ-algebra of a bounded function (9.1)
Recurrence Theorem (12.2)
Structure Theorem (12.3)
Szemerédi's Theorem (3.1)
Generalized von Neumann Theorem (5.4), (12.1)
ν as close to uniform (6.2)
σ-algebra of multiple basic Gowers anti-uniform functions (9.2)
Linear Forms Condition (11.1)
Correlation Condition (11.2)
Furstenberg Tower (9.3)
Generalized Koopman-von Neumann Structure Theorem (8.1)
Goldston Yıldırım (3.3)
Generalized Szemerédi's Theorem (3.2)
Green-Tao Theorem (10.1)

4 k-pseudorandom Measure

In this section, we will define precisely what a k-pseudorandom measure is. To do that, we need the following 2 definitions:

Definition 4.1 (Linear forms condition). *Suppose $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ satisfies $\mathbb{E}(\nu) = 1 + o(1)$. Let m_0, t_0, L_0 be fixed positive integers. Then we say ν satisfies the (m_0, t_0, L_0)-linear forms condition if the following holds: Take any positive integer $m \leq m_0$, positive integer $t \leq t_0$. Take any $b_i \in \mathbb{Z}_N$ where i ranges from 1 to m. Take any rational numbers $(L_{ij})_{1\leq i\leq m, 1\leq j\leq t}$ where the numerator and denominator of each L_{ij} is at most L_0 in absolute value. Denote $L = (L_{ij})_{1\leq i\leq m,1\leq j\leq t}$, $b = (b_i)_{1\leq i\leq m}$ such that L is a m-by-t matrix, and b is a m-by-1 vector. Take any linear function $\psi : \mathbb{Z}_N^t \to \mathbb{Z}_N^m$ of the form $\psi(x) = Lx + b$, where L_{ij} are interpreted as elements of $\mathbb{Z}_N$ (assuming N is a prime and $N \gg L_0$ so that $\mathbb{Z}_N$ is a field and inverses are well defined). Suppose no two rows of L are rational multiples of each other, and no row consists entirely of 0. Then*

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

where the 'little-oh' is with respect to $N \to \infty$ as throughout the paper, and ψ_i is the i-th component of ψ.

To understand the linear forms condition, first note that if we let $m = 1$, we then have $\mathbb{E}(\nu) = 1+o(1)$, which is the definition of a *measure* on $\mathbb{Z}_N$. For the general case, the linear forms condition is essentially saying the random variables $\nu \circ \psi_j$ (as functions from $\mathbb{Z}_N$ to $\mathbb{R}$) are uncorrelated with each other, when the ψ_j's are linear and with different linear coefficients (up a rational multiple). In connection to ergodic theory, if there is no restriction on the size of m, then the linear forms condition is closely related to the *weak mixing* property of a dynamic system[1]. As we will see in section 5 and 6, the linear forms condition will be invoked several times to estimate certain products.

Definition 4.2 (Correlation condition). *Suppose $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ satisfies $\mathbb{E}(\nu) = 1 + o(1)$. Let m_0 be a fixed positive integer. Then we say ν satisfies the m_0-correlation condition if for every positive integer $m \leq m_0$, there exists a weight function $\tau_m : \mathbb{Z}_N \to \mathbb{R}^+$ that satisfies the 'moment conditions':*

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

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

$$\mathbb{E}(\nu(x + h_1)\nu(x + h_2) \cdots \nu(x + h_m)|x \in \mathbb{Z}_N) \leq \sum_{1\leq i<j\leq m} \tau_m(h_i - h_j)$$

[1]Since a significant portion of the argument has a close connection with ergodic theory, we would like to set up some standard terminology from ergodic theory. We say $(X, \mathcal{X}, \mu)$ is a *measure space* if $\mathcal{X}$ is a σ-algebra over X and μ is a measure on X. If a map $T : X \to X$ is measurable with respect to $\mathcal{X}$ and $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}$, then we say T is a *measure preserving transformation*, and $(X, \mathcal{X}, \mu, T)$ is a *measure preserving dynamical system*. A measure preserving transformation T is *ergodic* if for every $E \in \mathcal{B}$ such that $T^{-1}(E) = E$, we have $\mu(E) = 0$ or $\mu(E) = 1$. A dynamic system $(X, \mathcal{X}, \mu, T)$ is *weak mixing* if $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} |\mu(A \cap T^{-i}B) - \mu(A)\mu(B)| = 0$.

for all $h_1, h_2, \cdots, h_m \in \mathbb{Z}_N$.

As we shall see in section 6, the correlation condition will be used to estimates certain products involving Gowers anti-uniform functions. It is essentially saying that the correlation between the random variables $\nu \circ \psi_j$ satisfies certain bounds, when the ψ_j's are linear functions with the same linear coefficient.

With the above definitions, we are now ready to define what a k-pseudorandom measure is.

Definition 4.3 (k-pseudorandom measure). *Suppose $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ satisfies $\mathbb{E}(\nu) = 1 + o(1)$. Then ν is a k-pseudorandom measure if it satisfies the $(k \cdot 2^{k-1}, 3k-4, k)$-linear forms condition and the 2^{k-1}-correlation condition.*

Remark. It is clear that if ν satisfies the (m_0, t_0, L_0)-linear forms condition, then for any $m \leq m_0, t \leq t_0, L \leq L_0$, ν also satisfies the (m, t, L)-linear forms condition. Hence the exact values $(k \cdot 2^{k-1}, 3k-4, k)$ are not very important; it turns out that it is a convenient (and sufficient) choice so that we can make use of this notion in our argument. Here is an example of the use of the linear forms condition: suppose ν satisfies the (3,2,1)-linear forms condition, then take arbitrary $b \in \mathbb{Z}_N$, and take $\psi_1(x) = x_1 + b, \psi_2(x) = x_2 + b, \psi_3(x) = x_1 + x_2 + b$, then apply the linear forms condition to ψ we can immediately conclude $\mathbb{E}(\nu(x_1+b)\nu(x_2+b)\nu(x_1+x_2+b)|x_1, x_2 \in \mathbb{Z}_N) = 1 + o(1)$ for all $b \in \mathbb{Z}_N$. The notion of the *linear forms condition* is closely related to the Gowers uniformity norms, which we shall define shortly. One might suspect that the correlation condition could be obtained from the linear forms condition. But note that the linear component of $x + h_1, x + h_2, \cdots, x + h_m$ are all the same (equal to 1), thus one cannot apply the linear forms condition.

The following simple result will be useful to the argument that we are going to present.

Lemma 4.4. *If ν is a k-pseudorandom measure, then $\nu_{1/2} := (\nu + 1)/2$ is also a k-pseudorandom measure.*

Proof. Clearly $\nu_{1/2} \geq 0$ and $\mathbb{E}(\nu_{1/2}) = 1 + o(1)$. To verify the linear forms condition, note that

$$\begin{aligned}
&\mathbb{E}(\nu_{1/2}(\psi_1(x))\nu_{1/2}(\psi_2(x)) \cdots \nu_{1/2}(\psi_m(x))|x \in \mathbb{Z}_N^t) \\
=&\mathbb{E}(\frac{\nu(\psi_1(x)) + 1}{2} \frac{\nu(\psi_2(x)) + 1}{2} \cdots \frac{\nu(\psi_m(x)) + 1}{2}|x \in \mathbb{Z}_N^t) \\
=&2^{-m}\mathbb{E}((\nu(\psi_1(x)) + 1)(\nu(\psi_2(x)) + 1) \cdots (\nu(\psi_m(x)) + 1)|x \in \mathbb{Z}_N^t) \\
=&2^{-m} \sum_{S \subset \{1,2,\cdots,m\}} \mathbb{E}(\prod_{i \in S} \nu(\psi_i(x))) \\
=&2^{-m} \sum_{S \subset \{1,2,\cdots,m\}} (1 + o(1)) \\
=&1 + o(1)
\end{aligned}$$

as required. The correlation condition can be verified in a very similar manner:

$$\begin{aligned}
&\mathbb{E}(\nu_{1/2}(x+h_1)\nu_{1/2}(x+h_2)\cdots\nu_{1/2}(x+h_m)|x\in\mathbb{Z}_N)\\
=&\mathbb{E}(\frac{\nu(x+h_1)+1}{2}\frac{\nu(x+h_2)+1}{2}\cdots\frac{\nu(x+h_m)+1}{2}|x\in\mathbb{Z}_N)\\
=&2^{-m}\mathbb{E}((\nu(x+h_1)+1)(\nu(x+h_2)+1)\cdots(\nu(x+h_m)+1)|x\in\mathbb{Z}_N)\\
=&2^{-m}\sum_{S\subset\{1,2,\cdots,m\}}\mathbb{E}(\prod_{i\in S}\nu(x+h_i))\\
\leq&2^{-m}\sum_{S\subset\{1,2,\cdots,m\}}\sum_{i,j\in S,i<j}\tau(h_i-h_j)\\
\leq&\sum_{1\leq i<j\leq m}\tau_m(h_i-h_j)
\end{aligned}$$

□

A direct implication of the lemma is that a function f bounded by $\nu+1$ is equivalent to bounded by ν, up a factor of 2. As we will see, constant factors do not matter for most of the arguments in the paper; the important part is the asymptotics. However, it is helpful to have the $\nu+1$ bound, because as we shall see in section 8, we will be considering functions of the form $f_j := f - \mathbb{E}(f|\mathcal{B})$ where f is bounded by ν and $\mathbb{E}(f|\mathcal{B})$ is bounded by 1, and thus f_j is bounded by $\nu+1$. The definition of $\mathbb{E}(f|\mathcal{B})$ will be made explicit in section 7, but for now one can think of it as the projection of $f\in L^1(\mathbb{Z}_N)$ onto a subspace of $L^1(\mathbb{Z}_N)$.

The following simple fact will be useful and appear several times throughout the paper: if $\Phi: A\to B$ is surjective and all fibers $\{\Phi^{-1}(b): b\in B\}$ have the same cardinality, then for any $f: B\to\mathbb{R}$ we have $\mathbb{E}(f(\Phi(a))|a\in A)=\mathbb{E}(f(b)|b\in B)$. This can be verified easily by expanding the definition of the expection operator explicitly. This fact allows us to pass from averaging over A to averaging over B and vice versa.

We also adopt the following standard inner product notation on $L^2(\mathbb{Z}_N)$ space: $\langle f,g\rangle := \mathbb{E}(fg)$ for any $f,g\in L^2(\mathbb{Z}_N)$.

5 Gowers Uniformity Norms and Generalized von Neumann Theorem

The technique to prove the generalized Szemerédi's Theorem is to decompose a function into a Gowers uniform component f_U and a Gowers anti-uniform component $f_{U^\perp}$. As we shall see later, we can decompose f as $f=f_U+f_{U^\perp}$, where f_U has small enough Gowers uniform norm so that its contribution can be neglected by the Generalized von Neumann Theorem that we will prove in

this section, and $f_{U^\perp}$ will satisfy the condition for the traditional Szemerédi's Theorem so that we we can apply the theorem to conclude that $f_{U^\perp}$ has positive contribution. In this section we are concerned with the Gowers uniform component and the Generalized von Neumann Theorem.

First, we will need some background and basic properties of the Gowers uniformity norm.

Definition 5.1 (Gowers Inner Product). *Let d be a positive integer. Denote $\{0,1\}^d := \{(w_1, \cdots, w_d) : w_j \in \{0,1\}\}$. For any $h = (h_1, \cdots, h_d)$, define $w \cdot h := \sum_{j=1}^d w_j h_j$. Suppose $(f_w)_{w\in\{0,1\}^d}$ is a 2^d-tuple of real-valued functions in $L^\infty(\mathbb{Z}_N)$ indexed by w. Then the d-dimensional Gowers inner product $\langle (f_w)_{w\in\{0,1\}^d}\rangle_{U^d}$ is defined by*

$$\langle (f_w)_{w\in\{0,1\}^d}\rangle_{U^d} := \mathbb{E}(\prod_{w\in\{0,1\}^d} f_w(x + w\cdot h)|x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^d).$$

This is a generalization of an inner product (from a bilinear form to a multilinear form). Certain positivity properties hold for this $\langle\cdot\rangle_{U^d}$ as expected. Let $w' := (w_1, \cdots, w_{d-1})$ be the first $d-1$ coordinates of w except for the last one. So we can write $w = (w', w_d)$. Similarly, let $h' := (h_1, \cdots, h_{d-1})$ so that $h = (h', h_d)$. Then

$$\begin{aligned}\langle (f_w)_{w\in\{0,1\}^d}\rangle_{U^d} &= \mathbb{E}(\prod_{w\in\{0,1\}^d} f_w(x + w\cdot h)|x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^d)\\ &= \mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w',0}(x + w'\cdot h')f_{w',1}(x + w'\cdot h' + h_d)|x \in \mathbb{Z}_N, h' \in \mathbb{Z}_N^{d-1}, h_d \in \mathbb{Z}_N)\end{aligned}$$

Note that if f_w is independent of the last coordinate of w, i.e. $f_{w',0} = f_{w',1} = f_{w'}$ for all $w' \in \{0,1\}^{d-1}$, then we further have

$$\begin{aligned}&\langle (f_w)_{w\in\{0,1\}^d}\rangle_{U^d}\\ =&\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w',0}(x + w'\cdot h')f_{w',1}(x + w'\cdot h' + h_d)|x \in \mathbb{Z}_N, h' \in \mathbb{Z}_N^{d-1}, h_d \in \mathbb{Z}_N)\\ =&\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w'}(x + w'\cdot h')f_{w'}(x + w'\cdot h' + h_d)|x \in \mathbb{Z}_N, h' \in \mathbb{Z}_N^{d-1}, h_d \in \mathbb{Z}_N)\\ =&\mathbb{E}(\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w'}(x + w'\cdot h')f_{w'}(x + w'\cdot h' + h_d)|x \in \mathbb{Z}_N, h_d \in \mathbb{Z}_N)|h' \in \mathbb{Z}_N^{d-1})\\ =&\mathbb{E}(\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w'}(x + w'\cdot h')f_{w'}(y + w'\cdot h')|x \in \mathbb{Z}_N, y \in \mathbb{Z}_N)|h' \in \mathbb{Z}_N^{d-1})\\ =&\mathbb{E}([\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w'}(z + w'\cdot h')|z \in \mathbb{Z}_N)]^2|h' \in \mathbb{Z}_N^{d-1})\\ \geq&0.\end{aligned}$$

In particular, when $(f_w)_{w\in\{0,1\}^d} = (f)_{w\in\{0,1\}^d}$, so f_w is completely independent of w (not just the last coordinate), we have $\langle (f)_{w\in\{0,1\}^d}\rangle_{U^d} \geq 0$. So for any $f : \mathbb{Z}_N \to \mathbb{R}$, we can define the *Gowers uniformity norm* $||f||_{U^d}$ of f by

$$||f||_{U^d} := \langle (f)_{w\in\{0,1\}^d}\rangle_{U^d}^{1/2^d}.$$

The power $1/2^d$ is chosen such that $||\lambda f||_{U^d} = |\lambda| \cdot ||f||_{U^d}$.

In general, when f_w does depend on the last coodinate of w, we still have the following inequality by applying Cauchy-Schwarz of the form $\mathbb{E}(f(h')g(h')) \leq [\mathbb{E}(f(h')^2)]^{1/2}[\mathbb{E}(g(h')^2)]^{1/2}$:

$$\begin{aligned}
&\langle (f_w)_{w\in\{0,1\}^d}\rangle_{U^d} \\
=&\mathbb{E}(\prod_{w\in\{0,1\}^d} f_w(x + w\cdot h)|x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^d) \\
=&\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w',0}(x + w'\cdot h')f_{w',1}(x + w'\cdot h' + h_d)|x \in \mathbb{Z}_N, h' \in \mathbb{Z}_N^{d-1}, h_d \in \mathbb{Z}_N) \\
=&\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w',0}(x + w'\cdot h')f_{w',1}(y + w'\cdot h')|x \in \mathbb{Z}_N, h' \in \mathbb{Z}_N^{d-1}, y \in \mathbb{Z}_N) \\
=&\mathbb{E}(\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w',0}(x + w'\cdot h')f_{w',1}(y + w'\cdot h')|x \in \mathbb{Z}_N, y \in \mathbb{Z}_N)|h' \in \mathbb{Z}_N^{d-1}) \\
=&\mathbb{E}(\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w',0}(x + w'\cdot h')f_{w',1}(y + w'\cdot h')|x \in \mathbb{Z}_N, y \in \mathbb{Z}_N)|h' \in \mathbb{Z}_N^{d-1}) \\
=&\mathbb{E}([\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w',0}(x + w'\cdot h')|x \in \mathbb{Z}_N)][\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w',1}(y + w'\cdot h')|y \in \mathbb{Z}_N)]|h' \in \mathbb{Z}_N^{d-1}) \\
\leq&\{\mathbb{E}([\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w',0}(x + w'\cdot h')|x \in \mathbb{Z}_N)]^2|h' \in \mathbb{Z}_N^{d-1})\}^{1/2} \\
&\cdot \{\mathbb{E}([\mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f_{w',1}(y + w'\cdot h')|y \in \mathbb{Z}_N)]^2|h' \in \mathbb{Z}_N^{d-1})\}^{1/2} \\
=&\langle (f_{w',0})_{w\in\{0,1\}^d}\rangle_{U^d}^{1/2} \cdot \langle (f_{w',1})_{w\in\{0,1\}^d}\rangle_{U^d}^{1/2}.
\end{aligned}$$

The way to interpret $(f_{w',0})_{w\in\{0,1\}^d}$ is that as w ranges over $\{0,1\}^d$, $f_{w',0}$ looks at the first $d-1$ coordinates of w and always puts 0 in the last coordinate regardless of what the last coordinate of w is. Similarly for $(f_{w',1})_{w\in\{0,1\}^d}$.

By the same argument, if we itereate over the each coordinate of w exactly once, we will obtain the *Gowers Cauchy Schwarz Inequality*:

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

With this, it is easy to verify that $|\langle (f+g)_{w\in\{0,1\}^d}\rangle_{U^d}| \leq (||f||_{U^d} + ||g||_{U^d})^{2^d}$:

$$\begin{aligned}
|\langle (f+g)_{w\in\{0,1\}^d}\rangle_{U^d}| =& \mathbb{E}(\prod_{w\in\{0,1\}^d} (f(x+w\cdot h)+g(x+w\cdot h))|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^d)\\
=& \mathbb{E}(\sum_{S\subset\{0,1\}^d}[\prod_{w\in S} f(x+w\cdot h)][\prod_{w\in\{0,1\}^d\setminus S} g(x+w\cdot h)]|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^d)\\
=& \sum_{S\subset\{0,1\}^d} \mathbb{E}([\prod_{w\in S} f(x+w\cdot h)][\prod_{w\in\{0,1\}^d\setminus S} g(x+w\cdot h)]|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^d)\\
\leq& \sum_{S\subset\{0,1\}^d} (\prod_{w\in S} ||f||_{U^d})(\prod_{w\in\{0,1\}^d\setminus S} ||g||_{U^d})\\
=& (||f||_{U^d} + ||g||_{U^d})^{2^d}.
\end{aligned}$$

Raising both sides to the $1/2^d$ power, we obtain the triangle inequality $||f+g||_{U^d} \leq ||f||_{U^d} + ||g||_{U^d}$.

One should note that U^1 is not a genuine norm since $||f||_{U^1} = |\mathbb{E}(f)|$, and thus we may have $||f||_{U^1} = 0$ without having $f \equiv 0$. However, for $d = 2$, we have the following Proposition.

Proposition 5.2. *U^2 is a genuine norm.*

Proof. We have already verified the positive homogeneity and triangle inequality. It remains to prove that if $||f||_{U^2} = 0$ then $f \equiv 0$. It turns out that an argument from Fourier analysis can help us establish this result. Consider the Fourier transform of f,

$$\hat{f}(\xi) := \mathbb{E}(f(x)e^{-2\pi i x\xi/N}|x\in\mathbb{Z}_N)$$

We claim that

$$||f||_{U^2} = (\sum_{\xi\in\mathbb{Z}_N} |\hat{f}(\xi)|^4)^{\frac{1}{4}}.$$

To see this, note that

$$\begin{aligned}
|\hat{f}(\xi)|^2 =& |\mathbb{E}(f(x)e^{-2\pi i x\xi/N}|x\in\mathbb{Z}_N)|^2\\
=& \mathbb{E}(f(x)e^{-2\pi i x\xi/N}|x\in\mathbb{Z}_N)\overline{\mathbb{E}(f(x')e^{-2\pi i x'\xi/N}|x'\in\mathbb{Z}_N)}\\
=& \mathbb{E}(f(x)e^{-2\pi i x\xi/N}|x\in\mathbb{Z}_N)\mathbb{E}(f(x')e^{2\pi i x'\xi/N}|x'\in\mathbb{Z}_N)\\
=& \mathbb{E}(f(x)f(x')e^{-2\pi i x\xi/N}e^{2\pi i x'\xi/N}|x\in\mathbb{Z}_N, x'\in\mathbb{Z}_N)\\
=& \mathbb{E}(f(x)f(x+h)e^{2\pi i h\xi/N}|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N)
\end{aligned}$$

Thus

$$\begin{aligned}|\hat{f}(\xi)|^4 =&(|\hat{f}(\xi)|^2)^2\\ =&\mathbb{E}(f(x)f(x+h)e^{2\pi ih\xi/N}|x\in\mathbb{Z}_N,h\in\mathbb{Z}_N)\mathbb{E}(f(x')f(x'+h')e^{2\pi ih'\xi/N}|x'\in\mathbb{Z}_N,h'\in\mathbb{Z}_N)\\ =&\mathbb{E}(f(x)f(x+h)f(x')f(x'+h')e^{2\pi i(h+h')\xi/N}|x\in\mathbb{Z}_N,h\in\mathbb{Z}_N,x'\in\mathbb{Z}_N,h'\in\mathbb{Z}_N)\end{aligned}$$

So

$$\begin{aligned}&\mathbb{E}(|\hat{f}(\xi)|^4|\xi\in\mathbb{Z}_N)\\ =&\mathbb{E}(f(x)f(x+h)f(x')f(x'+h')e^{2\pi i(h+h')\xi/N}|x\in\mathbb{Z}_N,h\in\mathbb{Z}_N,x'\in\mathbb{Z}_N,h'\in\mathbb{Z}_N,\xi\in\mathbb{Z}_N)\\ =&\mathbb{E}(\mathbb{E}(f(x)f(x+h)f(x')f(x'+h')e^{2\pi i(h+h')\xi/N}|\xi\in\mathbb{Z}_N)|x\in\mathbb{Z}_N,h\in\mathbb{Z}_N,x'\in\mathbb{Z}_N,h'\in\mathbb{Z}_N)\\ =&\mathbb{E}(f(x)f(x+h)f(x')f(x'+h')\mathbf{1}[h+h'=0]|x\in\mathbb{Z}_N,h\in\mathbb{Z}_N,x'\in\mathbb{Z}_N,h'\in\mathbb{Z}_N)\\ =&\frac{1}{N}\mathbb{E}(f(x)f(x+h)f(x')f(x'-h)|x\in\mathbb{Z}_N,h\in\mathbb{Z}_N,x'\in\mathbb{Z}_N)\\ =&\frac{1}{N}\mathbb{E}(f(x)f(x+h_1)f(x+h_2)f(x+h_1+h_2)|x\in\mathbb{Z}_N,h_1\in\mathbb{Z}_N,h_2\in\mathbb{Z}_N)\\ =&\frac{1}{N}||f||_{U^d}^4.\end{aligned}$$

Rearranging, the claim thus follows.

With the claim, we know that $||f||_{U^2}$ implies $\hat{f}\equiv 0$. On the other hand, note that

$$\begin{aligned}\mathbb{E}(\hat{f}(\xi)e^{2\pi it\xi/N}|\xi\in\mathbb{Z}_N) =&\mathbb{E}(\mathbb{E}(f(x)e^{-2\pi ix\xi/N}|x\in\mathbb{Z}_N)e^{2\pi it\xi/N}|\xi\in\mathbb{Z}_N)\\ =&\mathbb{E}(f(x)e^{2\pi i(t-x)\xi/N}|x,\xi\in\mathbb{Z}_N)\\ =&\mathbb{E}(f(x)\mathbf{1}[t=x]|x\in\mathbb{Z}_N)\\ =&\frac{1}{N}f(t)\end{aligned}$$

Thus $\hat{f}\equiv 0$ implies $f\equiv 0$. Hence $||f||_{U^2}$ implies $f\equiv 0$, completing the proof. □

It is a coincidence that we can use Fourier analysis to study the U^2 norm. For instance, if we apply the same methods to the U^1 norm, we would only get $\mathbb{E}(|\hat{f}(\xi)|^2|\xi\in\mathbb{Z}_N)=\frac{1}{N}\mathbb{E}(f(x)^2|x\in\mathbb{Z}_N)$ which leads to no conclusion about the U^1 norm. In fact this is expected since we already know that the U^1 norm is not a genuine norm. Fourier analysis would not apply to higher degree norms either. However, we do have have the following facts about the U^d norm for $d\geq 2$.

First, it is clear that $||\nu_{const}||_{U^d}=||1||_{U^d}=1$. Now consider $(f_w)_{w\in\{0,1\}^d}$ satisfies $f_{w',0}=f$ and

$f_{w',1} = 1$. Then according to the preceding result,

$$\begin{aligned}\langle (f_w)_{w\in\{0,1\}^d}\rangle_{U^d} &\leq \langle (f_{w',0})_{w\in\{0,1\}^d}\rangle_{U^d}^{1/2} \cdot \langle (f_{w',1})_{w\in\{0,1\}^d}\rangle_{U^d}^{1/2} \\ &= \langle (f)_{w\in\{0,1\}^d}\rangle_{U^d}^{1/2} \cdot \langle (1)_{w\in\{0,1\}^d}\rangle_{U^d}^{1/2} \\ &= (||f||_{U^d}^{2^d})^{1/2} \cdot 1 \\ &= ||f||_{U^d}^{2^{d-1}}.\end{aligned}$$

On the other hand, the left hand side also equals

$$\begin{aligned}\langle (f_w)_{w\in\{0,1\}^d}\rangle_{U^d} &= \mathbb{E}(\prod_{w'\in\{0,1\}^{d-1}} f(x+w'\cdot h')|x\in\mathbb{Z}_N, h'\in\mathbb{Z}_N^{d-1}) \\ &= ||f||_{U^{d-1}}^{2^{d-1}}.\end{aligned}$$

Therefore we have $||f||_{U^{d-1}}^{2^{d-1}} \leq ||f||_{U^d}^{2^{d-1}}$, and thus $||f||_{U^{d-1}} \leq ||f||_{U^d}$. Since U^2 is a genuine norm, it follows that U^d are genuine norms for all $d \geq 2$.

The following useful lemma says that a k-pseudorandom measure is close to the constant measure in the U^d norms for $d \leq k$.

Lemma 5.3. *Suppose ν is a k-pseudorandom measure. Then $||\nu - 1||_{U^d} = o(1)$ for all $d \leq k-1$.*

Proof.

$$\begin{aligned}||\nu-1||_{U^{k-1}}^{2^{k-1}} &= \mathbb{E}(\prod_{w\in\{0,1\}^{k-1}} (\nu(x+w\cdot h)-1)|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1}) \\ &= \mathbb{E}(\sum_{S\subset\{0,1\}^{k-1}} (-1)^{|\{0,1\}^{k-1}|-|S|}(\prod_{w\in S}\nu(x+w\cdot h))|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1}) \\ &= \mathbb{E}(\sum_{S\subset\{0,1\}^{k-1}} (-1)^{|S|}(\prod_{w\in S}\nu(x+w\cdot h))|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1}) \\ &= \sum_{S\subset\{0,1\}^{k-1}} (-1)^{|S|}\mathbb{E}(\prod_{w\in S}\nu(x+w\cdot h)|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1}).\end{aligned}$$

We can now use the linear forms condition for ν to estimate $\mathbb{E}(\prod_{w\in S}\nu(x+w\cdot h)|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1})$: as w ranges over S, each of $x+w\cdot h$ is a linear function of the vector $(x,h) = (x, h_1, \cdots, h_{k-1})$, and it is clear that none of these forms is a rational multiple of any other since no w is a multiple of any other. Invoking the $(2^{k-1}, k, 1)$-linear forms condition for ν, we have $\mathbb{E}(\prod_{w\in S}\nu(x+w\cdot h)|x\in$

$\mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1}) = 1 + o(1)$. Hence

$$\begin{aligned}
||\nu - 1||_{U^{k-1}}^{2^{k-1}} &= \sum_{S \subset \{0,1\}^{k-1}} (-1)^{|S|} \mathbb{E}(\prod_{w \in S} \nu(x + w \cdot h) | x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1}) \\
&= \sum_{S \subset \{0,1\}^{k-1}} (-1)^{|S|} (1 + o(1)) \\
&= \sum_{S \subset \{0,1\}^{k-1}} (-1)^{|S|} + o(1) \\
&= o(1).
\end{aligned}$$

Since k is fixed throughout, we have $||\nu - 1||_{U^{k-1}} = o(1)$. By the monotonicity of U^d norms, we have $||\nu - 1||_{U^d} = o(1)$ for all $d \leq k - 1$. □

The key results from the above discussion of the Gowers uniformity norms are summarized as the following:

- The Gowers uniformity norm $|| \cdot ||_{U^d}$ is a genuine norm for all $d \geq 2$.
- A multidimensional version of the Cauchy-Schwarz inequality holds: $|\langle (f_w)_{w \in \{0,1\}^d} \rangle_{U^d}| \leq \prod_{w \in \{0,1\}} ||f_w||_{U^d}$.
- A k-pseudorandom measure ν is close to the constant measure in the Gowers uniformity norm: $||\nu - 1||_{U^d} = o(1)$ for all $d \leq k - 1$.

We are now ready to state and prove the Generalized von Neumann Theorem[2].

Theorem 5.4 (Generalized von Neumann). *Suppose ν is a k-pseudorandom measure, and $f_0, f_1, \cdots, f_{k-1} \in L^1(\mathbb{Z}_N)$ satisfy*

$$|f_j(x)| \leq \nu(x) \forall x \in \mathbb{Z}_N, 0 \leq j \leq k - 1.$$

Then

$$\mathbb{E}(\prod_{j=0}^{k-1} f_j(x + jr) | x, r \in \mathbb{Z}_N) = O(\inf_{0 \leq j \leq k-1} ||f_j||_{U^{k-1}}) + o(1).$$

Remark. This is a powerful result. It asserts that if each f_j is bounded by a pseudorandom measure, then the expectation of their product evaluated at an arithmetic progression is bounded above by the minimal U^d norm among the f_j's. In particular, if one of the f_j's has small U^d norm, then the product must be small, regardless of the behavior of the other f_j's (as long as they remain bounded by ν). As we will see in section 8, this is an important step in proving the Generalized Szemerédi's theorem.

[2]The theorem is generalized in the sense that f is bounded by a pseudorandom measure rather than a constant.

Proof. The actual proof of the theorem is somewhat technical, but the main technique is again successive applications of the Cauchy-Schwarz inequality plus the linear forms condition on ν. But note that since there are only 2 variables in the expectation, it would be helpful to introduce more variables, so that in each step, we use Cauchy Schwarz in one of the new variables to estimate one of the f_j's. By successively applying Cauchy Schwarz, in the final step we would have estimated all but one of the f_j's, which is assumed to have the smallest U^d norm. At that point we can invoke the linear forms condition on ν to estimate the last f_j, which would give us the desired estimate. Of course, we would need to ensure that the expectation is invariant after we make such change of variables. This is achieved if the transformation is surjective and all fibers have the same cardinality, as noted in the end of section 4.

Without loss of generality, assume $\inf_{0\leq j\leq k-1}||f_j||_{U^{k-1}} = ||f_0||_{U^{k-1}}$. Adopt the following notation: Suppose $d \leq k-1, y = (y_1, \cdots, y_{k-1}) \in \mathbb{Z}_N^{k-1}, y' = (y'_{k-d}, \cdots, y'_{k-1}) \in \mathbb{Z}_N^d$. For $S \subset \{k-d, \cdots, k-1\}$, define $y^{(S)} = (y_1^{(S)}, \cdots, y_{k-1}^{(S)})$ as

$$y_i^{(S)} := \left\{ \begin{array}{ll} y_i & \text{if } i \notin S \\ y'_i & \text{if } i \in S \end{array} \right\}.$$

The following version of the Cauchy Schwarz inequality will be crucial to the proof of Theorem 5.4.

Proposition 5.5 (Cauchy Schwarz). *Suppose $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ satisfies $\mathbb{E}(\nu) = 1 + o(1)$. Suppose $\phi_0, \phi_1, \cdots, \phi_{k-1} : \mathbb{Z}_N^{k-1} \to \mathbb{R}^+$ where ϕ_i does not depend on the ith coordinate of the input for all $i = 0, 1, \cdots, k-1$. Suppose $f_0, f_1, \cdots, f_{k-1}$ satisfy $|f_i(x)| \leq \nu(x)$ for all x and all i. For each $d \leq k-1$, define*

$$J_d := \mathbb{E}(\prod_{S\subset\{k-d,\cdots,k-1\}} [\prod_{i=0}^{k-d-1} f_i(\phi_i(y^{(S)}))] \cdot [\prod_{i=k-d}^{k-1} \nu^{1/2}(\phi_i(y^{(S)}))] | y \in \mathbb{Z}_N^{k-1}, y' \in \mathbb{Z}_N^d)$$

and

$$P_d := \mathbb{E}(\prod_{S\subset\{k-d,\cdots,k-1\}} \nu(\phi_{k-d-1}(y^{(S)})) | y \in \mathbb{Z}_N^{k-1}, y' \in \mathbb{Z}_N^d)$$

Then for each $d \leq k-2$, we have $|J_d|^2 \leq P_d J_{d+1}$.

Remark. This is the version of the Cauchy Schwarz inequality that allows us to estimate f_{k-d} in the d-th step. Note that J_d involves only up to f_{k-d-1}, which means all f_j for $j \geq k-d$ have been estimated. The result of the proposition allows us to compare J_0 with J_{k-1} (after we iterate over d), which can then be used to obtain the estimate in Theorem 5.4.

Proof. The idea of the proof is to split J_d into 2 parts such that the first part involves f_{k-d-1} and the second does not. For the second part, we average over y_{k-d-1} to get rid of the y_{k-d-1} dependence. We then apply Cauchy Schwarz in all the variables except y_{k-d-1} to obtain an estimate for $|J_d|^2$.

Finally we use the fact that f_{k-d-1} is bounded by ν to conclude that the first part is bounded by P_d, and identify the second part as J_{d+1} to conclude the result.

Define the following quantities:

$$G(y,y') := \prod_{S\subset\{k-d,\cdots,k-1\}} f_{k-d-1}(\phi_{k-d-1}(y^{(S)}))\nu^{-1/2}(\phi_{k-d-1}(y^{(S)}))$$

$$\tilde{H}(y,y') := \prod_{S\subset\{k-d,\cdots,k-1\}} [\prod_{i=0}^{k-d-2} f_i(\phi_i(y^{(S)}))\cdot \prod_{i=k-d-1}^{k-1} \nu^{1/2}(\phi_i(y^{(S)}))]$$

and

$$\begin{aligned} H(y,y') :=& \mathbb{E}(\prod_{S\subset\{k-d,\cdots,k-1\}} [\prod_{i=0}^{k-d-2} f_i(\phi_i(y^{(S)}))\cdot \prod_{i=k-d-1}^{k-1} \nu^{1/2}(\phi_i(y^{(S)}))]|y_{k-d-1}\in\mathbb{Z}_N) \\ =& \mathbb{E}(\tilde{H}(y,y')|y_{k-d-1}\in\mathbb{Z}_N) \end{aligned}$$

Then

$$G(y,y')\tilde{H}(y,y') = \prod_{S\subset\{k-d,\cdots,k-1\}} [\prod_{i=0}^{k-d-1} f_i(\phi_i(y^{(S)}))]\cdot[\prod_{i=k-d}^{k-1} \nu^{1/2}(\phi_i(y^{(S)}))]$$

which is the inside of the expectation in the definition of J_d. Thus

$$J_d = \mathbb{E}(G(y,y')\tilde{H}(y,y')|y\in\mathbb{Z}_N^{k-1}, y'\in\mathbb{Z}_N^d).$$

Since ϕ_i does not depend on y_i, we observe that $G(y,y')$ does not depend on y_{k-d-1}, thus we can also write

$$\begin{aligned} J_d &= \mathbb{E}(G(y,y')\tilde{H}(y,y')|y_1,\cdots,y_{k-1},y'_{k-d},\cdots,y'_{k-1}\in\mathbb{Z}_N) \\ &= \mathbb{E}(\mathbb{E}(G(y,y')\tilde{H}(y,y')|y_{k-d-1}\in\mathbb{Z}_N)|y_1,\cdots,y_{k-d-2},y_{k-d},\cdots,y_{k-1},y'_{k-d},\cdots,y'_{k-1}\in\mathbb{Z}_N) \\ &= \mathbb{E}(G(y,y')\mathbb{E}(\tilde{H}(y,y')|y_{k-d-1}\in\mathbb{Z}_N)|y_1,\cdots,y_{k-d-2},y_{k-d},\cdots,y_{k-1},y'_{k-d},\cdots,y'_{k-1}\in\mathbb{Z}_N) \\ &= \mathbb{E}(G(y,y')H(y,y')|y_1,\cdots,y_{k-d-2},y_{k-d},\cdots,y_{k-1},y'_{k-d},\cdots,y'_{k-1}\in\mathbb{Z}_N) \end{aligned}$$

Thus applying Cauchy Schwarz to G and H, we get

$$\begin{aligned} |J_d|^2 \leq& \mathbb{E}(G(y,y')^2|y_1,\cdots,y_{k-d-2},y_{k-d},\cdots,y_{k-1},y'_{k-d},\cdots,y'_{k-1}\in\mathbb{Z}_N) \\ &\cdot \mathbb{E}(H(y,y')^2|y_1,\cdots,y_{k-d-2},y_{k-d},\cdots,y_{k-1},y'_{k-d},\cdots,y'_{k-1}\in\mathbb{Z}_N) \end{aligned}$$

Note that since $f_{k-d-1}(x) \leq \nu(x)$, we have

$$\begin{aligned} G(y,y') &= \prod_{S\subset\{k-d,\cdots,k-1\}} f_{k-d-1}(\phi_{k-d-1}(y^{(S)}))\nu^{-1/2}(\phi_{k-d-1}(y^{(S)})) \\ &\leq \prod_{S\subset\{k-d,\cdots,k-1\}} \nu(\phi_{k-d-1}(y^{(S)}))\nu^{-1/2}(\phi_{k-d-1}(y^{(S)})) \\ &= \prod_{S\subset\{k-d,\cdots,k-1\}} \nu^{1/2}(\phi_{k-d-1}(y^{(S)})) \end{aligned}$$

Thus

$$\begin{aligned} &\mathbb{E}(G(y,y')^2|y_1,\cdots,y_{k-d-2},y_{k-d},\cdots,y_{k-1},y'_{k-d},\cdots,y'_{k-1}\in\mathbb{Z}_N) \\ \leq&\mathbb{E}(\prod_{S\subset\{k-d,\cdots,k-1\}} \nu(\phi_{k-d-1}(y^{(S)}))|y_1,\cdots,y_{k-d-2},y_{k-d},\cdots,y_{k-1},y'_{k-d},\cdots,y'_{k-1}\in\mathbb{Z}_N) \\ =&P_d \end{aligned}$$

where the last equality follows because ϕ_{k-d-1} does not depend on y_{k-d-1}.

Moreover, by expanding the square of the expectation for $\tilde{H}$, we have

$$\begin{aligned} H(y,y')^2 =&[\mathbb{E}(\tilde{H}(y,y')|y_{k-d-1}\in\mathbb{Z}_N)]^2 \\ =&[\mathbb{E}(\tilde{H}(y_1,\cdots,y_{k-d-2},y_{k-d-1},y_{k-d},\cdots,y_{k-1},y'_{k-d},\cdots,y'_{k-1})|y_{k-d-1}\in\mathbb{Z}_N)]^2 \\ =&\mathbb{E}(\tilde{H}(y_1,\cdots,y_{k-d-2},y_{k-d-1},y_{k-d},\cdots,y_{k-1},y'_{k-d},\cdots,y'_{k-1}) \\ &\cdot\tilde{H}(y_1,\cdots,y_{k-d-2},y'_{k-d-1},y_{k-d},\cdots,y_{k-1},y'_{k-d},\cdots,y'_{k-1})|y_{k-d-1},y'_{k-d-1}\in\mathbb{Z}_N) \end{aligned}$$

On the other hand, by definition

$$J_{d+1} =\mathbb{E}(\prod_{S\subset\{k-d-1,\cdots,k-1\}}[\prod_{i=0}^{k-d-2} f_i(\phi_i(y^{(S)}))][\prod_{i=k-d-1}^{k-1}\nu^{1/2}(\phi_i(y^{(S)}))]|y\in\mathbb{Z}_N^{k-1},y'\in\mathbb{Z}_N^{d+1})$$

Partitioning the power set of $\{k-d-1,k-d,\cdots,k-1\}$ into 2 collections such that one contains sets that contain $k-d-1$ and the other contains sets that do not contain $k-d-1$, we can further write

$$\begin{aligned} J_{d+1} =&\mathbb{E}(\prod_{S\subset\{k-d,\cdots,k-1\}}[\prod_{i=0}^{k-d-2} f_i(\phi_i(y^{(S)}))][\prod_{i=k-d-1}^{k-1}\nu^{1/2}(\phi_i(y^{(S)}))] \\ &\cdot\prod_{S\subset\{k-d-1,\cdots,k-1\},k-d-1\in S}[\prod_{i=0}^{k-d-2} f_i(\phi_i(y^{(S)}))][\prod_{i=k-d-1}^{k-1}\nu^{1/2}(\phi_i(y^{(S)}))]|y\in\mathbb{Z}_N^{k-1},y'\in\mathbb{Z}_N^{d+1}) \end{aligned}$$

Notice that the first product equals $\tilde{H}(y_1, \cdots, y_{k-d-2}, y_{k-d-1}, y_{k-d}, \cdots, y_{k-1}, y'_{k-d}, \cdots, y'_{k-1})$ and the second product equals $\tilde{H}(y_1, \cdots, y_{k-d-2}, y'_{k-d-1}, y_{k-d}, \cdots, y_{k-1}, y'_{k-d}, \cdots, y'_{k-1})$. Thus

$$\begin{aligned}
J_{d+1} =&\mathbb{E}(\tilde{H}(y_1, \cdots, y_{k-d-2}, y_{k-d-1}, y_{k-d}, \cdots, y_{k-1}, y'_{k-d}, \cdots, y'_{k-1}) \\
&\cdot \tilde{H}(y_1, \cdots, y_{k-d-2}, y'_{k-d-1}, y_{k-d}, \cdots, y_{k-1}, y'_{k-d}, \cdots, y'_{k-1}) | y \in \mathbb{Z}_N^{k-1}, y' \in \mathbb{Z}_N^{d+1}) \\
=&\mathbb{E}(\mathbb{E}(\tilde{H}(y_1, \cdots, y_{k-d-2}, y_{k-d-1}, y_{k-d}, \cdots, y_{k-1}, y'_{k-d}, \cdots, y'_{k-1}) \\
&\cdot \tilde{H}(y_1, \cdots, y_{k-d-2}, y'_{k-d-1}, y_{k-d}, \cdots, y_{k-1}, y'_{k-d}, \cdots, y'_{k-1}) | y_{k-d-1}, y'_{k-d-1} \in \mathbb{Z}_N) \\
&| y_1, \cdots, y_{k-d-2}, y_{k-d}, \cdots, y_{k-1}, y'_{k-d}, \cdots, y'_{k-1} \in \mathbb{Z}_N) \\
=&\mathbb{E}(H(y, y')^2 | y_1, \cdots, y_{k-d-2}, y_{k-d}, \cdots, y_{k-1}, y'_{k-d}, \cdots, y'_{k-1} \in \mathbb{Z}_N)
\end{aligned}$$

Thus we have $|J_d|^2 \leq P_d J_{d+1}$, completing the proof of Proposition 5.5. □

We can rewrite $|J_d|^2 \leq P_d J_{d+1}$ as $J_{d+1} \geq P_d^{-1} J_d^2$, and by iteratively applying this inequality starting from $d = k-2$ until $d = 0$, we get

$$J_{k-1} \geq P_{k-2}^{-1} J_{k-2}^2 \geq P_{k-2}^{-1} P_{k-3}^{-2} J_{k-3}^4 \geq \cdots \geq [\prod_{d=0}^{k-2} P_d^{-2^{k-2-d}}] \cdot J_0^{2^{k-1}}.$$

Rearranging, we get

$$J_0^{2^{k-1}} \leq J_{k-1} \prod_{d=0}^{k-2} P_d^{2^{k-2-d}}. \tag{1}$$

To prove Theorem 5.4, we need to cleverly choose ϕ_i and then apply Inequality (1). We will show that the term involving P_d is $1 + o(1)$ and therefore can be discarded. Also, J_0 is equal to the product that we want to estimate in the theorem, thus the task is then to estimate J_{k-1}. Since J_{k-1} involves only f_0 but not other f_j's, we can directly compare it with $||f_0||_{U^{k-1}}$. We can bound the difference by certain products involving ν (since f_0 is bounded by ν), and invoke the linear forms condition to conclude the difference is $o(1)$, which will then imply Theorem 5.4.

For $i = 0, 1, \cdots, k-1$, take

$$\phi_i(y_1, \cdots, y_{k-1}) = \sum_{j=1}^{k-1} (1 - \frac{i}{j}) y_j.$$

One can easily check that ϕ_i does not depend on y_i. If we define

$$x = \sum_{j=1}^{k-1} y_j, \ r = -\sum_{j=1}^{k-1} \frac{y_j}{j}$$

then we have $\phi_i(y) = x + ir$, which we would hope to use to count the number of configurations $(x, x+r, \cdots, x+(k-1)r)$. Now consider the map $\Phi : \mathbb{Z}_N^{k-1} \to \mathbb{Z}_N^2$ defined by

$$(y_1, \cdots, y_{k-1}) \mapsto (\sum_{j=1}^{k-1} y_j, -\sum_{j=1}^{k-1} \frac{y_j}{j}) = (x, r)$$

It is clear that Φ is surjective. Since it is linear, it follows that all fibers have the same cardinality. Thus

$$\mathbb{E}(\prod_{i=0}^{k-1} f_i(x+ir) | x, r \in \mathbb{Z}_N) = \mathbb{E}(\prod_{i=0}^{k-1} f_i(\phi_i(y)) | y \in \mathbb{Z}_N^{k-1})$$

Thus to prove Theorem 5.4, it is equivalent to prove

$$\mathbb{E}(\prod_{i=0}^{k-1} f_i(\phi_i(y)) | y \in \mathbb{Z}_N^{k-1}) = O(||f_0||_{U^{k-1}}) + o(1).$$

Note that the left hand side is just J_0, so it is equivalent to prove

$$J_0 = O(||f_0||_{U^{k-1}}) + o(1). \tag{2}$$

Recall inequality (1) $J_0^{2^{k-1}} \leq J_{k-1} \prod_{d=0}^{k-2} P_d^{2^{k-2-d}}$, where $P_d := \mathbb{E}(\prod_{S \subset \{k-d, \cdots, k-1\}} \nu(\phi_{k-d-1}(y^{(S)}))$ $|y \in \mathbb{Z}_N^{k-1}, y' \in \mathbb{Z}_N^d)$. If we let $y'' := (y, y') \in \mathbb{Z}_N^{(k-1)+d}$ and $\psi_S(y'') := \phi_{k-d-1}(y^{(S)})$, then we can rewrite P_d as

$$P_d := \mathbb{E}(\prod_{S \subset \{k-d, \cdots, k-1\}} \nu(\psi_S(y'')) | y'' \in \mathbb{Z}_N^{k-1+d}).$$

It is clear that for each $S \subset \{k-d, \cdots, k-1\}$, ψ_S is a linear map whose coefficients are less than or equal to k in absolute value, and no ψ_S is a rational multiple of any other. Thus by invoking the $(2^d, k-1+d, k)$-linear forms condition on ν (since ν is a k-pseudorandom measure), we conclude that $P_d = 1 + o(1)$ for each d. Thus from inequality (1) we get

$$J_0^{2^{k-1}} \leq J_{k-1}(1 + o(1)) \tag{3}$$

So now our task is to bound J_{k-1}. By definition,

$$J_{k-1} = \mathbb{E}(\prod_{S \subset \{1, \cdots, k-1\}} [f_0(\phi_0(y^{(S)}))] \cdot [\prod_{i=1}^{k-1} \nu^{1/2}(\phi_i(y^{(S)}))] | y \in \mathbb{Z}_N^{k-1}, y' \in \mathbb{Z}_N^{k-1})$$

Consider the map $w : 2^{\{1, \cdots, k-1\}} \to \{0,1\}^{k-1}$ by $S \mapsto (\mathbf{1}[1 \in S], \mathbf{1}[2 \in S], \cdots, \mathbf{1}[k-1 \in S])$, where $\mathbf{1}[\cdot]$ is the indicator function. Note that w is a bijection, and as S ranges over the power set

$2^{\{1,\cdots,k-1\}}$, $w(S)$ will range over $\{0,1\}^{k-1}$. Furthermore, observe that $y_j^{(S)} = y_j + w_j(S)(y'_j - y_j)$, so if we let $x = \sum_{j=1}^{k-1} y_j$ as before and define $h = (h_1, \cdots, h_{k-1})$ where $h_j = y'_j - y_j$, then

$$\phi_i(y^{(S)}) = \sum_{j=1}^{k-1}(1 - \frac{i}{j})y_j^{(S)} = \sum_{j=1}^{k-1}(1 - \frac{i}{j})(y_j + w_j(S)(y'_j - y_j)) = \sum_{j=1}^{k-1}(1 - \frac{i}{j})(y_j + w_j(S)h_j)$$

In particular, we for $i = 0$, we have

$$\phi_0(y^{(S)}) = \sum_{j=1}^{k-1}(y_j + w_j(S)h_j) = \sum_{j=1}^{k-1} y_j + \sum_{j=1}^{k-1} w_j(S)h_j = x + w \cdot h.$$

Indexing using $w \in \{0,1\}^{k-1}$ instead of $S \subset \{1, \cdots, k-1\}$, we can rewrite J_{k-1} as

$$\begin{aligned}
J_{k-1} =& \mathbb{E}(\prod_{w\in\{0,1\}^{k-1}} [f_0(x + w \cdot h)] \cdot [\prod_{i=1}^{k-1} \nu^{1/2}(\phi_i(y^{(S)}))] | y \in \mathbb{Z}_N^{k-1}, y' \in \mathbb{Z}_N^{k-1}) \\
=& \mathbb{E}([\prod_{w\in\{0,1\}^{k-1}} f_0(x + w \cdot h)] \cdot [\prod_{w\in\{0,1\}^{k-1}} \prod_{i=1}^{k-1} \nu^{1/2}(\phi_i(y^{(S)}))] | y \in \mathbb{Z}_N^{k-1}, y' \in \mathbb{Z}_N^{k-1}) \\
=& \mathbb{E}([\prod_{w\in\{0,1\}^{k-1}} f_0(x + w \cdot h)] \cdot [\prod_{w\in\{0,1\}^{k-1}} \prod_{i=1}^{k-1} \nu^{1/2}(\phi_i(y^{(S)}))] | y \in \mathbb{Z}_N^{k-1}, h \in \mathbb{Z}_N^{k-1})
\end{aligned}$$

where the last equality holds because fixing y, there is a bijection between y' and h, so the quantity is invariant when we change from averaging over y' to averaging over h. Now define

$$\tilde{W} := \prod_{w\in\{0,1\}^{k-1}} \prod_{i=1}^{k-1} \nu^{1/2}(\phi_i(y^{(S)}))$$

and

$$W := \mathbb{E}(\tilde{W} | y_1, \cdots, y_{k-2} \in \mathbb{Z}_N) = \mathbb{E}(\prod_{w\in\{0,1\}^{k-1}} \prod_{i=1}^{k-1} \nu^{1/2}(\phi_i(y^{(S)})) | y_1, \cdots, y_{k-2} \in \mathbb{Z}_N).$$

So we can write

$$\begin{aligned}
J_{k-1} =& \mathbb{E}(\tilde{W}[\prod_{w\in\{0,1\}^{k-1}} f_0(x + w \cdot h)] | y \in \mathbb{Z}_N^{k-1}, h \in \mathbb{Z}_N^{k-1}) && (4) \\
=& \mathbb{E}(\mathbb{E}(\tilde{W}[\prod_{w\in\{0,1\}^{k-1}} f_0(x + w \cdot h)] | y_1, \cdots, y_{k-2} \in \mathbb{Z}_N) | y_{k-1} \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1}) && (5) \\
=& \mathbb{E}(\mathbb{E}(\tilde{W}[\prod_{w\in\{0,1\}^{k-1}} f_0(x + w \cdot h)] | y_1, \cdots, y_{k-2} \in \mathbb{Z}_N) | x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1}) && (6) \\
=& \mathbb{E}(W[\prod_{w\in\{0,1\}^{k-1}} f_0(x + w \cdot h)] | x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1}) && (7)
\end{aligned}$$

where (6) follows because after averaging over $y_1, \cdots, y_{k-2}$, there is a bijection between y_{k-1} and x; (7) follows because fixing x and h, $\prod_{w\in\{0,1\}^{k-1}} f_0(x+w\cdot h)$ is a constant, thus can be taken out of the inner expectation operator.

Note that by definition,

$$\mathbb{E}(\prod_{w\in\{0,1\}^{k-1}} f_0(x+w\cdot h)|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1}) = ||f_0||_{U^{k-1}}^{2^{k-1}}$$

Combined with (2), (3), and (7), it suffices to prove that

$$\mathbb{E}(|W-1|[\prod_{w\in\{0,1\}^{k-1}} f_0(x+w\cdot h)]|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1}) = o(1).$$

Since f_0 is bounded by ν by assumption, it suffices to prove that

$$\mathbb{E}(|W-1|[\prod_{w\in\{0,1\}^{k-1}} \nu(x+w\cdot h)]|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1}) = o(1). \tag{8}$$

First, note that

$$\begin{aligned}
\tilde{W} &= \prod_{w\in\{0,1\}^{k-1}}\prod_{i=1}^{k-1}\nu^{1/2}(\phi_i(y^{(S)})) \\
&= \prod_{i=1}^{k-1}\prod_{w\in\{0,1\}^{k-1}}\nu^{1/2}(\phi_i(y^{(S)})) \\
&= \prod_{i=1}^{k-1}[\prod_{w\in\{0,1\}^{k-1},w_i=0}\nu^{1/2}(\phi_i(y^{(S)}))][\prod_{w\in\{0,1\}^{k-1},w_i=1}\nu^{1/2}(\phi_i(y^{(S)}))].
\end{aligned}$$

Since ϕ_i does not depend on i, the first term equals the second term, thus

$$\tilde{W} = \prod_{i=1}^{k-1}[\prod_{w\in\{0,1\}^{k-1},w_i=0}\nu^{1/2}(\phi_i(y^{(S)}))]^2 = \prod_{i=1}^{k-1}\prod_{w\in\{0,1\}^{k-1},w_i=0}\nu(\phi_i(y^{(S)})).$$

So

$$W = \mathbb{E}(\tilde{W}|y_1,\cdots,y_{k-2}\in\mathbb{Z}_N) = \mathbb{E}(\prod_{i=1}^{k-1}\prod_{w\in\{0,1\}^{k-1},w_i=0}\nu(\phi_i(y^{(S)}))|y_1,\cdots,y_{k-2}\in\mathbb{Z}_N)$$

Define

$$Q_q := \mathbb{E}(W^q[\prod_{w\in\{0,1\}^{k-1}}\nu(x+w\cdot h)]|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1})$$

for $q = 0, 1, 2$. Then

$$Q_0 = \mathbb{E}(\prod_{w\in\{0,1\}^{k-1}} \nu(x + w \cdot h)|x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1})$$

Note that $x + w \cdot h$ is a linear form in the variable $(x, h) = (x, h_1, \cdots, h_{k-1})$, and as w ranges over $\{0, 1\}^{k-1}$, none of the forms is a multiple of any other. Thus by invoking the $(2^{k-1}, k, 1)$-linear forms condition on ν, we conclude $Q_0 = 1 + o(1)$.

Similarly, we have

$$\begin{aligned}
Q_1 =&\mathbb{E}(W[\prod_{w\in\{0,1\}^{k-1}} \nu(x + w \cdot h)]|x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1})\\
=&\mathbb{E}([\mathbb{E}(\prod_{i=1}^{k-1} \prod_{w\in\{0,1\}^{k-1},w_i=0} \nu(\phi_i(y^{(S)}))|y_1, \cdots, y_{k-2} \in \mathbb{Z}_N)][\prod_{w\in\{0,1\}^{k-1}} \nu(x + w \cdot h)]|x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1})\\
=&\mathbb{E}([\prod_{i=1}^{k-1} \prod_{w\in\{0,1\}^{k-1},w_i=0} \nu(\phi_i(y^{(S)}))][\prod_{w\in\{0,1\}^{k-1}} \nu(x + w \cdot h)]|y_1, \cdots, y_{k-2} \in \mathbb{Z}_N, x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1})
\end{aligned}$$

Invoking the $((k-1)2^{k-2} + 2^{k-1}, (k-2) + 1 + (k-1), k)$-linear forms condition on ν with variables $y_1, \cdots, y_{k-2}, x, h_1, \cdots, h_{k-1}$ and forms $\phi_i(y^{(S)})$ and $x + w \cdot h$, we conclude that $Q_1 = 1 + o(1)$.

Also, by expanding W^2, we have

$$\begin{aligned}
W^2 =&[\mathbb{E}(\prod_{i=1}^{k-1} \prod_{w\in\{0,1\}^{k-1},w_i=0} \nu(\phi_i(y^{(S)}))|y_1, \cdots, y_{k-2} \in \mathbb{Z}_N)]^2\\
=&\mathbb{E}([\prod_{i=1}^{k-1} \prod_{w\in\{0,1\}^{k-1},w_i=0} \nu(\phi_i(y^{(S)}))][\prod_{i=1}^{k-1} \prod_{w\in\{0,1\}^{k-1},w_i=0} \nu(\phi_i(\tilde{y}^{(S)}))]|y_1, \cdots, y_{k-2}, \tilde{y}_1, \cdots, \tilde{y}_{k-2} \in \mathbb{Z}_N)
\end{aligned}$$

Thus

$$\begin{aligned}
Q_2 =&\mathbb{E}(W^2[\prod_{w\in\{0,1\}^{k-1}} \nu(x + w \cdot h)]|x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1})\\
=&\mathbb{E}([\prod_{i=1}^{k-1} \prod_{w\in\{0,1\}^{k-1},w_i=0} \nu(\phi_i(y^{(S)}))][\prod_{i=1}^{k-1} \prod_{w\in\{0,1\}^{k-1},w_i=0} \nu(\phi_i(\tilde{y}^{(S)}))][\prod_{w\in\{0,1\}^{k-1}} \nu(x + w \cdot h)]\\
&|y_1, \cdots, y_{k-2}, \tilde{y}_1, \cdots, \tilde{y}_{k-2} \in \mathbb{Z}_N, x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1})
\end{aligned}$$

Invoking the $(2(k-1)2^{k-2} + 2^{k-1}, 2(k-2) + 1 + (k-1), k)$-linear forms condition on ν with variables $y_1, \cdots, y_{k-2}, \tilde{y}_1, \cdots, \tilde{y}_{k-2}, x, h_1, \cdots, h_{k-1}$ and forms $\phi_i(y^{(S)})$, $\phi_i(\tilde{y}^{(S)})$ and $x + w \cdot h$, we conclude that $Q_2 = 1 + o(1)$.

Now let $V := \prod_{w\in\{0,1\}^{k-1}} \nu(x+w\cdot h)$. Then (8), which is what we want to prove, can be written as

$$\mathbb{E}(|W-1|V|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1}) = o(1). \tag{9}$$

By Cauchy Schwarz, we have

$$\begin{aligned}
&\mathbb{E}(|W-1|V|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1})\\
=&\mathbb{E}(|W-1|V^{\frac{1}{2}}\cdot V^{\frac{1}{2}}|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1})\\
\leq&[\mathbb{E}((|W-1|V^{\frac{1}{2}})^2|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1})\cdot\mathbb{E}((V^{\frac{1}{2}})^2|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1})]^{\frac{1}{2}}\\
=&[\mathbb{E}(|W-1|^2V|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1})\cdot\mathbb{E}(V|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1})]^{\frac{1}{2}}\\
=&[\mathbb{E}((W^2V-2WV+V)|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1})\cdot\mathbb{E}(V|x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^{k-1})]^{\frac{1}{2}}\\
=&[(Q_2-2Q_1+Q_0)(Q_0)]^{\frac{1}{2}}\\
=&[(o(1))(1+o(1))]^{\frac{1}{2}}\\
=&o(1)
\end{aligned}$$

Therefore (9) and hence (8) is proven. This completes the proof of Theorem 5.4. □

Corollary 5.6. *Suppose ν is a k-pseudorandom measure, and $f_0, f_1, \cdots, f_{k-1}\in L^1(\mathbb{Z}_N)$ satisfy*

$$|f_j(x)|\leq\nu(x)+1\forall x\in\mathbb{Z}_N, 0\leq j\leq k-1.$$

Then

$$\mathbb{E}(\prod_{j=0}^{k-1} f_j(x+jr)|x,r\in\mathbb{Z}_N) = O(\inf_{0\leq j\leq k-1}||f_j||_{U^{k-1}}) + o(1).$$

Proof. Note that the only difference from Theorem 5.4 is that here the condition is $|f_j(x)|\leq\nu(x)+1$ instead of $|f_j(x)|\leq\nu(x)$. But if $|f_j(x)|\leq\nu(x)+1$, then $|f_j(x)|/2\leq(\nu(x)+1)/2\equiv\nu'(x)$ where ν' is also a k-pseudorandom measure by Lemma 4.4. Hence by Theorem 5.4, we will have

$$\mathbb{E}(\prod_{j=0}^{k-1} f_j(x+jr)/2|x,r\in\mathbb{Z}_N) = O(\inf_{0\leq j\leq k-1}||f_j/2||_{U^{k-1}}) + o(1).$$

It then follows that

$$\mathbb{E}(\prod_{j=0}^{k-1} f_j(x+jr)|x,r\in\mathbb{Z}_N) = O(\inf_{0\leq j\leq k-1}||f_j||_{U^{k-1}}) + o(1).$$

□

Corollary 5.6 is the form of the Generalized von Neumann Theorem that we are going to use. As noted in section 4, eventually we will apply the Generalized von Neumann Theorem to a function bounded by $\nu+1$ instead of ν.

6 Gowers Anti-uniformity Norms

After studying the Gowers uniformity norm U^d on $L^\infty(\mathbb{Z}_N)$, we now consider its dual norm $(U^d)^*$. Recall that in general, the dual space X^* of a Banach space X with norm $||\cdot||$ is the space of linear functionals from X to $\mathbb{R}$ with the following norm $||\cdot||_*$: for $L \in X^*$,

$$||L||_* := \sup\{|L(x)| : x \in X, ||x|| \leq 1\}.$$

It follows easily that $|L(x)| \leq ||L||_*||x||$ for all $x \in X$ and $L \in X^*$, since $|L(x/||x||)| \leq \sup\{|L(x')| : x' \in X, ||x'|| \leq 1\} = ||L||_*$.

For $X = L^\infty(\mathbb{Z}_N)$ with the U^d norm ($d \geq 2$), we can define the dual norm $(U^d)^*$ as: for $g \in L^\infty(\mathbb{Z}_N)$,

$$||g||_{(U^d)^*} := \sup\{|\langle f, g\rangle| : f \in L^\infty(\mathbb{Z}_N), ||f||_{U^d} \leq 1\}.$$

We have $|\langle f, g\rangle| \leq ||f||_{U^d}||g||_{(U^d)^*}$ for all $f, g \in L^\infty(\mathbb{Z}_N)$. We also say that g is *Gowers anti-uniform* if $||g||_{(U^d)^*} = O_d(1)$ and $||g||_{L^\infty} = O_d(1)$.

Now consider the case $d = k - 1$. For $F \in L^1(\mathbb{Z}_N)$, we define the dual function DF of F as

$$DF(x) := \mathbb{E}(\prod_{w \in \{0,1\}^{k-1}, w \neq \{0\}^{k-1}} F(x + w \cdot h) | h \in \mathbb{Z}_N^{k-1}).$$

Note that DF is almost the same as $\langle (f)_{w \in \{0,1\}} \rangle_{U^{k-1}}$ except that the term inside the expectation misses $F(x)$. The dual function $DF(x)$ has the following properties.

Proposition 6.1. *Suppose $F \in L^1(\mathbb{Z})$ and ν is a k-pseudorandom measure. Then*

$$\langle F, DF\rangle = ||F||_{U^{k-1}}^{2^{k-1}} \tag{10}$$

and

$$||DF||_{(U^{k-1})^*} = ||F||_{U^{k-1}}^{2^{k-1}-1}. \tag{11}$$

If we further have

$$|F(x)| \leq \nu(x) + 1 \ \forall x \in \mathbb{Z}_N, \tag{12}$$

Then

$$||DF||_{L^\infty} \leq 2^{2^{k-1}-1} + o(1). \tag{13}$$

Proof. The proof is relatively straight forward. (10) can be easily verified by expanding the left-hand side (recall $\langle f, g\rangle := \mathbb{E}(fg)$) and using the definition of $||F||_{U^{k-1}}^{2^{k-1}}$. To see (11), note that for

any arbitrary function f, we can consider $(f_w)_{w\in\{0.1\}^{k-1}}$ where $f_w := f$ if $w = \{0\}^{k-1}$ and $f_w := F$ otherwise. Then by the Gowers Cauchy Schwarz inequality,

$$|\langle f, DF\rangle| = |\langle (f_w)_{w\in\{0,1\}^{k-1}}\rangle_{U^{k-1}}| \leq \prod_{w\in\{0,1\}^{k-1}} ||f_w||_{U^{k-1}} = ||f||_{U^{k-1}}||F||_{U^{k-1}}^{2^{k-1}-1}.$$

Note that if $0 < ||f||_{U^{k-1}} \leq 1$, then

$$|\langle f, DF\rangle| \leq \frac{|\langle f, DF\rangle|}{||f||_{U^{k-1}}} \leq ||F||_{U^{k-1}}^{2^{k-1}-1}.$$

Hence by the definition of the $(U^{k-1})^*$ norm,

$$||DF||_{(U^{k-1})^*} = \sup\{|\langle f, DF\rangle| : ||f||_{U^{k-1}} \leq 1\} \leq ||F||_{U^{k-1}}^{2^{k-1}-1}.$$

On the other hand, this supremum is actually achieved by taking $f := F/||F||_{U^{k-1}}$,

$$|\langle F/||F||_{U^{k-1}}, DF\rangle| = |\langle F, DF\rangle|/||F||_{U^{k-1}} = ||F||_{U^{k-1}}^{2^{k-1}}/||F||_{U^{k-1}} = ||F||_{U^{k-1}}^{2^{k-1}-1}.$$

Hence $||DF||_{(U^{k-1})^*} = ||F||_{U^{k-1}}^{2^{k-1}-1}$.

To prove (13), note that by (12), we have $|F(x)| \leq 2\nu_{1/2}(x)$ for all $x \in \mathbb{Z}_N$ where $\nu_{1/2}$ is also a k-pseudorandom measure, so

$$\begin{aligned}|DF(x)| \leq& \mathbb{E}(\prod_{w\in\{0,1\}^{k-1}, w\neq\{0\}^{k-1}} 2\nu(x + w\cdot h)|h \in \mathbb{Z}_N^{k-1})\\ =& 2^{2^{k-1}-1}\mathbb{E}(\prod_{w\in\{0,1\}^{k-1}, w\neq\{0\}^{k-1}} \nu(x + w\cdot h)|h \in \mathbb{Z}_N^{k-1})\end{aligned}$$

Apply the $(2^{k-1}-1, k-1, 1)$-linear forms condition on $\nu_{1/2}$, and noting that the $o(1)$ term is uniform in x, we conclude that

$$\mathbb{E}(\prod_{w\in\{0,1\}^{k-1}, w\neq\{0\}^{k-1}} \nu(x + w\cdot h)|h \in \mathbb{Z}_N^{k-1}) = 1 + o(1)$$

for all $x \in \mathbb{Z}_N$. Hence $||DF||_{L^\infty} \leq 2^{2^{k-1}-1} + o(1)$ as desired. □

The results of Proposition 6.1 are crucial to the proof of the Generalized Szemerédi's Theorem. It implies that if a function F bounded by a pseudorandom measure is not Gowers uniform (i.e. F has large U^{k-1} norm), then by (10) F correlates with a Gowers anti-uniform function DF. The key is that DF is bounded (by inequality (13)). This is a somewhat remarkable result since we never require F to be bounded by a constant (F is only required to be bounded by ν). Recall the main technique to prove the generalized Szemerédi's Theorem is to decompose a function f into a Gowers uniform component f_U and a Gowers anti-uniform component $f_{U^\perp}$. The way to obtain such

a decomposition is through conditional expectation. The basic idea is that we set $f_U = f - \mathbb{E}(f|\mathcal{B})$ and $f_{U^\perp} = \mathbb{E}(f|\mathcal{B})$, where $\mathcal{B}$ is some σ-algebra over the base field $\mathbb{Z}_N$. Initially, $\mathcal{B}$ will be the trivial σ-algebra, so that f_U captures most of the variation of f and $f_{U^\perp}$ is just a constant. If f_U is not Gowers uniform, then it correlates with some Gowers anti-uniform function; with boundedness, we will then be able to construct a finer σ-algebra such that the new f_U will have smaller Gowers uniformity norm. As the procedure is repeated, $f_{U^\perp}$ captures more and more variation of f while f_U captures less and less. The procedure will be repeated until f_U has small enough Gowers uniformity norm. At that point we can apply the Generalized von Neumann Theorem on f_U and the traditional Szemerédi's Theorem on $f_{U^\perp}$ to conclude the proof of the Generalized Szemerédi's Theorem. The argument will be made more explicit in later sections.

A function of the form DF for some F bounded by $\nu+1$ are said to be a *basic Gowers anti-uniform function.* Since the idea is to consider σ-algebras over $\mathbb{Z}_N$, it is natural to consider the algebra generated by basic Gowers anti-uniform functions (as we shall see later, a function is associated with a σ-algebra). The next proposition is essentially saying that a pseudorandom measure is close to the constant measure, if it is restricted to the algebra generated by a collection of basic Gowers anti-uniform functions.

Proposition 6.2. *Let ν be a k-pseudorandom measure, $K \geq 1$ be a fixed integer, and I be the inteval $[-2^{2^{k-1}}, 2^{2^{k-1}}]$. Let $\Phi : I^K \to \mathbb{R}$ be an arbitrary fixed continuous function, and $DF_1, \cdots, DF_K$ be basic Gowers anti-uniform functions. Define $\psi : \mathbb{Z}_N \to \mathbb{R}$ by*

$$\psi(x) := \Phi(DF_1(x), \cdots, DF_K(x)).$$

Then we have

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

Furthermore, if Φ ranges over some compact subspace $E \subset C^0(I^K)$, where $C^0(I^K)$ denotes the space of continuous functions from I^K to $\mathbb{R}$ with the topology of uniform convergence, then we have

$$\langle \nu - 1, \psi \rangle = o_{K,E}(1),$$

i.e., the $o(1)$ term is uniform in Φ.

Proof. We will first prove the result for the case where Φ is a polynomial, then since the claim is an estimation statement, we will then use the Weierstrass Approximation Theorem to conclude the same result when Φ is a general continuous function. By assumption, we have $|F_j(x)| \leq \nu(x) + 1 = 2\nu_{1/2}(x)$ for all $j \leq K$, and it is clear that the constant factor will not matter, by an argument similar to the proof of Corollary 5.6. Thus it suffices to prove the claim for the case in which $|F_j(x)| \leq \nu(x)$.

Let us further restrict ourselves to consider the case in which Φ is the K-th elementary symmetric polynomial in K variables: $\Phi(x_1, \cdots, x_K) = x_1 \cdots x_K$ (recall K is fixed). We have the following lemma.

Lemma 6.3. $||\prod_{j=1}^{K} DF_j||_{(U^{k-1})^*} = O_K(1)$.

Remark. Most of the work is done for the case where Φ is of the form $\prod_{j=1}^{K} DF_j$. It should not be surprising that $||\prod_{j=1}^{K} DF_j||_{(U^{k-1})^*}$ is bounded since each DF_j is.

Proof. The basic technique in the proof is to expand the left hand side as a double expectation, then change the order of expectation to express it as an expectation of a Gowers inner product, which, by the Gowers Cauchy Schwarz inequality, can be bounded by the product of individual Gowers uniformity norms. Then by a version of Hölder's inequality, we can further bound this by the product of expectations of individual Gowers uniformity norms. Finally, the expectation of each Gowers uniformity norm will be bounded due to the correlation condition on ν, and noting that the number of such expectations is fixed (dependent on k only), the lemma follows.

By definition of the $(U^{k-1})^*$ norm, we need to show that

$$\langle f, \prod_{j=1}^{K} DF_j \rangle = O_K(1)$$

for all $||f||_{U^{k-1}} \leq 1$. Expanding the left hand side, we get

$$\begin{aligned}\langle f, \prod_{j=1}^{K} DF_j \rangle =& \mathbb{E}(f(x)[\prod_{j=1}^{K} \mathbb{E}(\prod_{w\in\{0,1\}^{k-1}, w\neq\{0\}^{k-1}} F_j(x + w \cdot h^{(j)})|h^{(j)} \in \mathbb{Z}_N^{k-1})]|x \in \mathbb{Z}_N) \\ =& \mathbb{E}(f(x)[\prod_{j=1}^{K} \prod_{w\in\{0,1\}^{k-1}, w\neq\{0\}^{k-1}} F_j(x + w \cdot h^{(j)})]|h^{(j)} \in \mathbb{Z}_N^{k-1}, x \in \mathbb{Z}_N)\end{aligned}$$

Write $h^{(j)} = h + H^{(j)}$ for $h, H^{(j)} \in \mathbb{Z}_N^{k-1}$ and average over h and $H^{(j)}$ (it should be noted that we have used this change-of-variable trick a few times by now), we have

$$\begin{aligned}&\langle f, \prod_{j=1}^{K} DF_j \rangle \\ =&\mathbb{E}(f(x)[\prod_{j=1}^{K} \prod_{w\in\{0,1\}^{k-1}, w\neq\{0\}^{k-1}} F_j(x + w \cdot h + w \cdot H^{(j)})]|h \in \mathbb{Z}_N^{k-1}, H^{(j)} \in \mathbb{Z}_N^{k-1}, x \in \mathbb{Z}_N) \\ =&\mathbb{E}(\mathbb{E}(f(x)[\prod_{j=1}^{K} \prod_{w\in\{0,1\}^{k-1}, w\neq\{0\}^{k-1}} F_j(x + w \cdot h + w \cdot H^{(j)})]|x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1})|H \in (\mathbb{Z}_N^{k-1})^K) \\ =&\mathbb{E}(\mathbb{E}(f(x)[\prod_{w\in\{0,1\}^{k-1}, w\neq\{0\}^{k-1}} \prod_{j=1}^{K} F_j(x + w \cdot h + w \cdot H^{(j)})]|x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1})|H \in (\mathbb{Z}_N^{k-1})^K).\end{aligned}$$

If we define $(f_w^H)_{w\in\{0,1\}^{k-1}}$ where $f_w^H(x) := f(x)$ if $w = \{0\}^{k-1}$ and $f_w^H(x) := \prod_{j=1}^K F_j(x + w\cdot H^{(j)})$ if $w \neq \{0,1\}^{k-1}$, then by the definition of the Gowers inner product, we can write

$$\langle f, \prod_{j=1}^K DF_j\rangle = \mathbb{E}(\langle (f_w^H)_{w\in\{0,1\}^{k-1}}\rangle_{U^{k-1}} | H \in (\mathbb{Z}_N^{k-1})^K).$$

By the Gowers Cauchy Schwarz inequality, we have

$$|\langle f, \prod_{j=1}^K DF_j\rangle| \leq \mathbb{E}(\prod_{w\in\{0,1\}^{k-1}} ||f_w^H||_{U^{k-1}} | H \in (\mathbb{Z}_N^{k-1})^K). \tag{14}$$

Claim 6.4 (Hölder's Inequality). *For all $n \geq 1$ and $f : X \to \mathbb{R}^+$, we have*

$$\mathbb{E}(\prod_{i=1}^n f_i(x) | x \in X) \leq \prod_{i=1}^n [\mathbb{E}(f_i(x)^n | x \in X)]^{\frac{1}{n}}$$

if the quantities exist.

Proof. We will prove by indunction on n. The base case is clear. By Hölder's inequality, we have

$$\begin{aligned}
\mathbb{E}(\prod_{i=1}^n f_i(x) | x \in X) =& \mathbb{E}([\prod_{i=1}^{n-1} f_i(x)] f_n(x) | x \in X) \\
\leq& [\mathbb{E}((\prod_{i=1}^{n-1} f_i(x))^{\frac{n}{n-1}} | x \in X)]^{\frac{n-1}{n}} [\mathbb{E}(f_n(x)^n | x \in X)]^{\frac{1}{n}} \\
=& [\mathbb{E}(\prod_{i=1}^{n-1} f_i(x)^{\frac{n}{n-1}} | x \in X)]^{\frac{n-1}{n}} [\mathbb{E}(f_n(x)^n | x \in X)]^{\frac{1}{n}} \\
\leq& [\prod_{i=1}^{n-1} \mathbb{E}(f_i(x)^{\frac{n}{n-1}(n-1)} | x \in X)]^{\frac{1}{n-1}\frac{n-1}{n}} [\mathbb{E}(f_n(x)^n | x \in X)]^{\frac{1}{n}} \\
=& [\prod_{i=1}^{n-1} \mathbb{E}(f_i(x)^n | x \in X)]^{\frac{1}{n}} [\mathbb{E}(f_n(x)^n | x \in X)]^{\frac{1}{n}} \\
=& \prod_{i=1}^n [\mathbb{E}(f_i(x)^n | x \in X)]^{\frac{1}{n}}.
\end{aligned}$$

□

Thus by (14) and Claim 6.4, we have

$$|\langle f, \prod_{j=1}^K DF_j\rangle| \leq [\prod_{w\in\{0,1\}^{k-1}} \mathbb{E}(||f_w^H||_{U^{k-1}}^{2^{k-1}} | H \in (\mathbb{Z}_N^{k-1})^K)]^{1/2^{k-1}}. \tag{15}$$

Now for a fixed $w \neq \{0\}^{k-1}$, the map $H \mapsto (w \cdot H^{(1)}, \cdots, w \cdot H^{(K)}) \equiv (u^{(1)}, \cdots, u^{(K)}) \equiv u \in \mathbb{Z}_N^K$ is a surjective and all fibers have the same cardinality, thus

$$\begin{aligned}
&\mathbb{E}(||f_w^H||_{U^{k-1}}^{2^{k-1}}|H \in (\mathbb{Z}_N^{k-1})^K)\\
=&\mathbb{E}(\mathbb{E}(\prod_{w' \in \{0,1\}^{k-1}} f_w^H(x + w' \cdot h')|x \in \mathbb{Z}_N, h' \in \mathbb{Z}_N^{k-1})|H \in (\mathbb{Z}_N^{k-1})^K)\\
=&\mathbb{E}(\mathbb{E}(\prod_{w' \in \{0,1\}^{k-1}} \prod_{j=1}^{K} F_j(x + w \cdot H^{(j)} + w' \cdot h')|x \in \mathbb{Z}_N, h' \in \mathbb{Z}_N^{k-1})|H \in (\mathbb{Z}_N^{k-1})^K)\\
=&\mathbb{E}(\mathbb{E}(\prod_{w' \in \{0,1\}^{k-1}} \prod_{j=1}^{K} F_j(x + u^{(j)} + w' \cdot h')|x \in \mathbb{Z}_N, h' \in \mathbb{Z}_N^{k-1})|u^{(j)} \in \mathbb{Z}_N)\\
=&\mathbb{E}(\prod_{j=1}^{K} \mathbb{E}(\prod_{w' \in \{0,1\}^{k-1}} F_j(x + u^{(j)} + w' \cdot h')|u^{(j)} \in \mathbb{Z}_N)|x \in \mathbb{Z}_N, h' \in \mathbb{Z}_N^{k-1})\\
\leq&\mathbb{E}([\mathbb{E}(\prod_{w' \in \{0,1\}^{k-1}} \nu(x + u' + w' \cdot h')|u' \in \mathbb{Z}_N)]^K|x \in \mathbb{Z}_N, h' \in \mathbb{Z}_N^{k-1})\\
=&\mathbb{E}([\mathbb{E}(\prod_{w' \in \{0,1\}^{k-1}} \nu(x' + w' \cdot h')|x' \in \mathbb{Z}_N)]^K|h' \in \mathbb{Z}_N^{k-1}).
\end{aligned}$$

With this and (15), it suffices to prove

$$\mathbb{E}([\mathbb{E}(\prod_{w' \in \{0,1\}^{k-1}} \nu(x' + w' \cdot h')|x' \in \mathbb{Z}_N)]^K|h' \in \mathbb{Z}_N^{k-1}) = O_K(1). \tag{16}$$

Now invoke the 2^{k-1}-correlation condition on ν, then we have

$$\mathbb{E}(\prod_{w' \in \{0,1\}^{k-1}} \nu(x' + w' \cdot h')|x' \in \mathbb{Z}_N) \leq \sum_{w', w'' \in \{0,1\}^{k-1}, w' \neq w''} \tau((w' - w'') \cdot h')$$

where $\tau : \mathbb{Z}_N \to \mathbb{R}^+$ satisfies $\mathbb{E}(\tau^q) = O_q(1)$ for all $q \geq 1$. Thus

$$\begin{aligned}
&\mathbb{E}([\mathbb{E}(\prod_{w' \in \{0,1\}^{k-1}} \nu(x' + w' \cdot h')|x' \in \mathbb{Z}_N)]^K|h' \in \mathbb{Z}_N^{k-1})\\
\leq&\mathbb{E}([\sum_{w', w'' \in \{0,1\}^{k-1}, w' \neq w''} \tau((w' - w'') \cdot h')]^K|h' \in \mathbb{Z}_N^{k-1})\\
\leq&\mathbb{E}(2^{k-1}[\max\{\tau((w' - w'') \cdot h') : w', w'' \in \{0,1\}^{k-1}, w' \neq w''\}]^K|h' \in \mathbb{Z}_N^{k-1})
\end{aligned}$$

But note that for each distinct $w', w'' \in \{0,1\}^{k-1}$, the map $h' \mapsto (w' - w'') \cdot h'$ is surjective and all fibers have the same cardinality, so $\mathbb{E}(\tau((w' - w'') \cdot h')^K|h' \in \mathbb{Z}_N^{k-1}) = \mathbb{E}(\tau^K) = O_K(1)$. Hence

$$\mathbb{E}([\mathbb{E}(\prod_{w' \in \{0,1\}^{k-1}} \nu(x' + w' \cdot h')|x' \in \mathbb{Z}_N)]^K|h' \in \mathbb{Z}_N^{k-1}) \leq 2^{k-1}\mathbb{E}(\tau^K) = O_K(1),$$

completing the proof of Lemma 6.3. □

With Lemma 6.3, it is easy to see that the $(U^{k-1})^*$ norm of any monomial $DF_1^{d_1} \cdots DF_K^{d_K}$ is $O_{d,K}(1)$ where $d = \sum_{j=1}^K d_j$ is the degree of the monomial, because we can regard the monomial $DF_1^{d_1} \cdots DF_K^{d_K}$ as the d-th elementary symmetric polynomial in the d variables $DF_1, \cdots, DF_1, \cdots, DF_K, \cdots, DF_K$ (since we never require the DF_j be distinct). By the triangle inequality of the $(U^{k-1})^*$ norm, we conclude that $||P(DF_1, \cdots, DF_K)||_{(U^{k-1})^*} = O_{K,d,P}(1)$ for any polynomial P where d is the degree of P.

We are now ready to consider the general case in which Φ is any continuous fucntion. From (13), we know that for large enough N, DF_j takes on values in $[-2^{2^{k-1}}, 2^{2^{k-1}}]$. Thus by Weierstrass, for any $\epsilon > 0$, there exists a polynomial P depending on K, ϵ, and Φ such that

$$||\Phi(DF_1, \cdots, DF_K) - P(DF_1, \cdots, DF_K)||_{L^\infty} \leq \epsilon.$$

Thus

$$|\langle \nu - 1, \Phi - P \rangle| = |\mathbb{E}((\nu - 1)(\Phi - P))| \leq \mathbb{E}(|\nu - 1||\Phi - P|) \leq \epsilon \mathbb{E}(|\nu - 1|) \leq \epsilon(2 + o(1)).$$

On the other hand, by Lemma 5.3, we have $||\nu - 1||_{U^{k-1}} = o(1)$, thus

$$|\langle \nu - 1, P \rangle| = |||\nu - 1||_{U^{k-1}} \langle \frac{\nu - 1}{||\nu - 1||_{U^{k-1}}}, P \rangle| \leq ||\nu - 1||_{U^{k-1}} ||P||_{(U^{k-1})^*} = o_{K,\epsilon,P}(1) = o_{K,\epsilon,\Phi}(1)$$

since P depends on K, ϵ, and Φ. Thus by the triangle inequality, keeping ϵ fixed, we have

$$|\langle \nu - 1, \Phi \rangle| \leq |\langle \nu - 1, \Phi - P \rangle| + |\langle \nu - 1, P \rangle| \leq 2\epsilon + o_{K,\epsilon,\Phi}(1)$$

Now let $\epsilon \to 0$, it follows that $\langle \nu - 1, \Phi \rangle = o_{K,\Phi}(1)$. To obtain the uniform bound $o_{K,E}(1)$ where E is a compact subspace of $C^0(I^K)$, we can consider a covering of E by a finite collection of open balls each with radius ϵ centered at some Φ_α, and take the maximum over these $o_{K,\Phi_\alpha}(1)$ bounds to get the uniform bound $o_{K,E}(1)$ when Φ ranges over E. □

Note that if DF_j were not bounded, we would not be able to apply the Weierstrass Approximation Theorem to obtain the estimate for a general continuous function Φ. Thus once again we see that boundedness is important.

The way we are going to apply Proposition 6.2 is to use it to generate σ-algebras over $\mathbb{Z}_N$. To do that, we need to study σ-algebras over $\mathbb{Z}_N$.

7 σ-algebras Associated with Gowers Anti-uniform Functions

Let us first define the concept of σ-algebras over $\mathbb{Z}_N$ and the notion of measurability.

Definition 7.1 (σ-algebras over $\mathbb{Z}_N$). *A collection $\mathcal{B}$ of subsets of $\mathbb{Z}_N$ is a σ-algebra over $\mathbb{Z}_N$ if it contains the empty set $\emptyset$ and the whole set $\mathbb{Z}_N$, and is closed under complementation and countable unions. An element A in $\mathcal{B}$ is an atom if it is a minimal element in $\mathcal{B}$ (with respect to set inclusion), i.e. there does not exist any element in $\mathcal{B}$ that is a proper subset of A. If $\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_K$ are σ-algebras over $\mathbb{Z}_N$, then the σ-algebra generated by them, denoted by $\vee_{j=1}^K \mathcal{B}_j$, is the smallest σ-algebra over $\mathbb{Z}_N$ containg all of $\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_K$.*

It is clear that the atoms in a σ-algebra $\mathcal{B}$ form a partition of $\mathcal{B}$, and $\mathcal{B}$ consists of arbitrary unions of its atoms. We say that $\mathcal{B}$ is generated by its atoms. It is also clear that the atoms in $\vee_{j=1}^K \mathcal{B}_j$ are the intersections of the atoms in $\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_K$.

Definition 7.2 (Measurability). *The elements in the σ-algebra $\mathcal{B}$ are measurable sets. A function $f : \mathbb{Z}_N \to \mathbb{R}$ is measurable with respect to $\mathcal{B}$, or $\mathcal{B}$-measurable, if for each $x \in \mathbb{R}$, $f^{-1}(x) \in \mathcal{B}$, or equivalently, f is constant on each atom of $\mathcal{B}$.*

For any $f : \mathbb{Z}_N \to \mathbb{R}$, we define $\mathbb{E}(f|\mathcal{B}) : \mathbb{Z}_N \to \mathbb{R}$ by

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

where $\mathcal{B}(x)$ is the unique atom in $\mathcal{B}$ containing x. Then it is clear that $\mathbb{E}(f|\mathcal{B})$ is $\mathcal{B}$-measurable. $\mathbb{E}(f|\mathcal{B})$ can also be interpreted as the projection of $f \in L^q(\mathbb{Z}_N)$ to the subspace $L^q(\mathcal{B})$.

It is clear that the Law of Iterated Expectation holds: if $\mathcal{B}'$ is a sub σ-algebra of $\mathcal{B}$, then

$$\mathbb{E}(\mathbb{E}(f|\mathcal{B})|\mathcal{B}') = \mathbb{E}(f|\mathcal{B}').$$

It is also clear that if $\mathcal{B}'$ is a sub σ-algebra of $\mathcal{B}$ and f is $\mathcal{B}'$-measurable, then f is also $\mathcal{B}$-measurable.

8 Generalized Koopman-von Neumann Structure Theorem and Proof of Generalized Szemerédi's Theorem

The key result we need to prove the Generalized Szemerédi's Theorem is the Generalized Koopman-von Neumann Structure Theorem, which allows us to decompose a function into a Gowers uniform component and a Gowers anti-uniform component. As we have mentioned, this is the main idea in the proof of the Generalized Szemerédi's Theorem. With the terminology in the previous section, we are now ready to state the theorem.

Theorem 8.1 (Generalized Koopman-von Neumann Structure Theorem). *Suppose ν is a k-pseudorandom measure, and $f : \mathbb{Z}_N \to \mathbb{R}$ satisfies $0 \le f(x) \le \nu(x)$ for all $x \in \mathbb{Z}_N$. Let $\epsilon \in (0, 1)$ be a small positive real number, and $N > N_0(\epsilon)$ be sufficiently large. Then there exists a σ-algebra $\mathcal{B}$ over $\mathbb{Z}_N$ and an exceptional set $\Omega \in \mathcal{B}$ such that*

$$\mathbb{E}(\nu \mathbf{1}_\Omega) = o_\epsilon(1) \tag{17}$$

$$||(1-\mathbf{1}_\Omega)\mathbb{E}(\nu-1|\mathcal{B})||_{L^\infty} = o_\epsilon(1) \tag{18}$$

$$||(1-\mathbf{1}_\Omega)(f-\mathbb{E}(f|\mathcal{B}))||_{U^{k-1}} \leq \epsilon^{1/2^k}. \tag{19}$$

Let us now prove the Generalized Szemerédi's Theorem assuming the Generalized Koopman-von Neumann Structure Theorem.

Proof. The main idea in the proof is to decompose a function into $f_U := (1-\mathbf{1}_\Omega)(f-\mathbb{E}(f|\mathcal{B}))$ and $f_{U^\perp} := (1-\mathbf{1}_\Omega)\mathbb{E}(f|\mathcal{B})$. We then expand the left hand side of the statement of Theorem 3.2, and collect the terms involving at least one f_U. Then by the Generalized von Neumann Theorem, each term is at most $O(||f_U||_{U^{k-1}})$, which is at most $O(\epsilon^{1/2^k})$ by (19) and thus can be discarded as $\epsilon \to 0$. The term involving only $f_{U^\perp}$ will have an expectation that is bounded below by Szemerédi's theorem, since $f_{U^\perp}$ can be verified to satisfy the condition for the theorem, by (17) and (18). Combining the 2 results and taking $\epsilon \to 0$, the Generalized Szemerédi's Theorem follows.

Let f, δ be as in the statement of Theorem 3.2. Let $\epsilon \in (0, \delta]$ be an arbitrary fixed real number. Pick $\mathcal{B}$ and Ω according to Theorem 8.1. Define $f_U := (1-\mathbf{1}_\Omega)(f-\mathbb{E}(f|\mathcal{B}))$ and $f_{U^\perp} := (1-\mathbf{1}_\Omega)\mathbb{E}(f|\mathcal{B})$. Then since $\Omega \in \mathcal{B}$, we have $\mathbf{1}_\Omega = \mathbb{E}(\mathbf{1}_\Omega|\mathcal{B})$ which is $\mathcal{B}$-measurable. Hence by (17),

$$\begin{aligned}\mathbb{E}(f_{U^\perp}) &= \mathbb{E}((1-\mathbf{1}_\Omega)\mathbb{E}(f|\mathcal{B})) = \mathbb{E}(f) - \mathbb{E}(\mathbf{1}_\Omega\mathbb{E}(f|\mathcal{B})) = \mathbb{E}(f) - \mathbb{E}(\mathbb{E}(\mathbf{1}_\Omega f|\mathcal{B})) = \mathbb{E}(f) - \mathbb{E}(\mathbf{1}_\Omega f)\\ &\geq \mathbb{E}(f) - \mathbb{E}(\nu\mathbf{1}_\Omega) \geq \delta - o_\epsilon(1).\end{aligned}$$

On the other hand, by (18),

$$\begin{aligned}||f_{U^\perp}||_{L^\infty} &= ||(1-\mathbf{1}_\Omega)\mathbb{E}(f|\mathcal{B})||_{L^\infty}\\ &= ||(1-\mathbf{1}_\Omega) + (1-\mathbf{1}_\Omega)\mathbb{E}(f-1|\mathcal{B})||_{L^\infty}\\ &\leq ||(1-\mathbf{1}_\Omega)||_{L^\infty} + ||(1-\mathbf{1}_\Omega)\mathbb{E}(f-1|\mathcal{B})||_{L^\infty}\\ &\leq ||(1-\mathbf{1}_\Omega)||_{L^\infty} + ||(1-\mathbf{1}_\Omega)\mathbb{E}(\nu-1|\mathcal{B})||_{L^\infty}\\ &= 1 + o_\epsilon(1).\end{aligned}$$

It is also clear that $f_{U^\perp} \geq 0$ since $f \geq 0$. Now let $g = \frac{f_{U^\perp}}{1+o_\epsilon(1)}$. (This is a slight abuse of notation; what it means is that suppose $f_{U^\perp}(x) \leq 1 + h(\epsilon)$ where $h(\epsilon) \to 0$ as $N \to \infty$ for each ϵ, then we let $g = \frac{f_{U^\perp}}{h}$.) Then we have $0 \leq g(x) \leq 1$, and

$$\mathbb{E}(g) \geq \frac{\delta - o'_\epsilon(1)}{1+o_\epsilon(1)} = \delta - o_\epsilon(1).$$

Thus invoking Szemerédi's Theorem on g, we have

$$\mathbb{E}(g(x)g(x+r)\cdots g(x+(k-1)r)|x, r \in \mathbb{Z}_N) \geq c(k,\delta) - o_\epsilon(1) - o_{k,\delta}(1).$$

Thus

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

It is clear that

$$\mathbb{E}(\frac{f_{U^\perp}}{2}(x)\frac{f_{U^\perp}}{2}(x+r)\cdots\frac{f_{U^\perp}}{2}(x+(k-1)r)|x,r\in\mathbb{Z}_N)\geq c'(k,\delta)-o_\epsilon(1)-o_{k,\delta}(1)$$

for some other positive constant $c'(k,\delta)$ (which could be set to $\frac{c(k,\delta)}{2^k}$).

On the other hand, let us consider the Gowers uniform component f_U. By deifinition, we have $f_U=(1-\mathbf{1}_\Omega)f-f_{U^\perp}$, so

$$|f_U|=|(1-\mathbf{1}_\Omega)f-f_{U^\perp}|\leq|(1-\mathbf{1}_\Omega)f|+|f_{U^\perp}|\leq\nu+1+o_\epsilon(1)\leq 2(\nu+1).$$

Thus $|\frac{f_U}{2}|\leq\nu+1$. By (19), $||\frac{f_U}{2}||_{U^{k-1}}=\frac{1}{2}||f_U||_{U^{k-1}}\leq\frac{1}{2}\epsilon^{1/2^k}$. Thus by the Generalized von Neumann Theorem,

$$\mathbb{E}(f_0(x)f_1(x+r)\cdots f_{k-1}(x+(k-1)r)|x,r\in\mathbb{Z}_N)=O(\frac{1}{2}\epsilon^{1/2^k})+o_\epsilon(1)$$

where each f_j is either $\frac{f_U}{2}$ or $\frac{f_{U^\perp}}{2}$, with at least one f_j equal to $\frac{f_U}{2}$.

Now define $\tilde{f}:=\frac{f_U}{2}+\frac{f_{U^\perp}}{2}$. Then by the binomial formula,

$$\begin{aligned}&\mathbb{E}(\tilde{f}(x)\tilde{f}(x+r)\cdots\tilde{f}(x+(k-1))|x,r\in\mathbb{Z}_N)\\ =&\mathbb{E}(\prod_{j=0}^{k-1}(\frac{f_U}{2}(x+jr)+\frac{f_{U^\perp}}{2}(x+jr))|x,r\in\mathbb{Z}_N)\\ =&\mathbb{E}(\prod_{j=0}^{k-1}\frac{f_{U^\perp}}{2}(x+jr)|x,r\in\mathbb{Z}_N)+\sum_{f_j\in\{\frac{f_U}{2},\frac{f_{U^\perp}}{2}\},(f_0,\cdots,f_{k-1})\neq(\frac{f_{U^\perp}}{2},\cdots,\frac{f_{U^\perp}}{2})}\mathbb{E}(\prod_{j=0}^{k-1}f_j(x+jr)|x,r\in\mathbb{Z}_N)\\ \geq&c'(k,\delta)-o_\epsilon(1)-o_{k,\delta}(1)-(2^k-1)(O(\frac{1}{2}\epsilon^{1/2^k})+o_\epsilon(1))\\ \geq&c'(k,\delta)-O(\epsilon^{1/2^k})-o_\epsilon(1)-o_{k,\delta}(1).\end{aligned}$$

Since $0\leq\tilde{f}=\frac{f_U+f_{U^\perp}}{2}=\frac{(1-\mathbf{1}_\Omega)f}{2}\leq f$, we have

$$\mathbb{E}(f(x)f(x+r)\cdots f(x+(k-1)r)|x,r\in\mathbb{Z}_N)\geq c'(k,\delta)-O(\epsilon^{1/2^k})-o_\epsilon(1)-o_{k,\delta}(1).$$

Finally, take $\epsilon=\epsilon(k,\delta)$ to be small enough, we conclude that

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

for some positive constant $c''(k,\delta)$. The Generalized Szemerédi's Theorem thus follows. □

9 Proof of the Generalized Koopman-von Neumann Structure Theorem

To complete the proof of the Generalized Szemerédi's Theorem, it remains to prove the Generalized Koopman-von Neumann Structure Theorem. The method of the proof is closely related to the argument presented in Furstenberg's ergodic theory proof of Szemerédi's theorem ([3], [4]). Since a large portion of the arguments involves straightforward but somewhat lengthy calculations, instead of giving a step by step proof, we are going to present a skeleton of the proof, with some of the computational details omitted. Before we start the proof, let us study σ-algebras over $\mathbb{Z}_N$ associated with a general bounded function.

Proposition 9.1. *Suppose ν is a k-pseudorandom measure, and $0 < \epsilon < 1$, $0 < \eta < 1/2$. For any $G : \mathbb{Z}_N \to I$ where $I := [-2^{2^{k-1}}, 2^{2^{k-1}}]$, there exists a σ-algebra $\mathcal{B}_{\epsilon,\eta}(G)$ that satisfies:*

- *For any σ-algebra $\mathcal{B}$ over $\mathbb{Z}_N$, $||G - \mathbb{E}(G|\mathcal{B} \vee \mathcal{B}_{\epsilon,\eta}(G))||_{L^\infty} \leq \epsilon$.*
- *G is generated by at most $O(1/\epsilon)$ atoms.*
- *For any atom A in $\mathcal{B}_{\epsilon,\eta}(G)$, there exists a continuous function $\Psi_A : I \to [0,1]$ such that*

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

Furthermore, Ψ_A can be chosen from a set $E_{\epsilon,\eta}$, where $E_{\epsilon,\eta}$ is a compact subset of $C^0(I)$ which depends only on ϵ and η, and is independent of G, ν, N, or A.

We say $\mathcal{B}_{\epsilon,\eta}(G)$ is the σ-algebra generated by G.

Proof. Changing the order of integration, summation, and expectation, we observe that

$$\begin{aligned}
&\int_0^1 \sum_{n\in\mathbb{Z}} \mathbb{E}(\mathbf{1}[G(x) \in [\epsilon(n-\eta+\alpha), \epsilon(n+\eta+\alpha)]](\nu(x)+1)|x \in \mathbb{Z}_N)d\alpha \\
=&\mathbb{E}(\sum_{n\in\mathbb{Z}} \int_0^1 \mathbf{1}[G(x) \in [\epsilon(n-\eta+\alpha), \epsilon(n+\eta+\alpha)]](\nu(x)+1)d\alpha|x \in \mathbb{Z}_N) \\
=&\mathbb{E}((\nu(x)+1)2\eta|x \in \mathbb{Z}_N) \\
=&O(\eta).
\end{aligned}$$

Hence there exists some $\alpha \in [0,1]$ such that

$$\sum_{n\in\mathbb{Z}} \mathbb{E}(\mathbf{1}[G(x) \in [\epsilon(n-\eta+\alpha), \epsilon(n+\eta+\alpha)]](\nu(x)+1)|x \in \mathbb{Z}_N) = O(\eta). \tag{20}$$

Now let $\mathcal{B}_{\epsilon,\eta}(G)$ be the σ-algebra generated by the following collection of atoms:

$$\mathcal{A} := \{G^{-1}([\epsilon(n+\alpha), \epsilon(n+1+\alpha))) : n \in \mathbb{Z}\}.$$

Note that $\mathcal{A}$ is a collection of atoms because $\{[\epsilon(n+\alpha), \epsilon(n+1+\alpha)) : n \in \mathbb{Z}\}$ is a partition of $\mathbb{R}$, and thus $\mathcal{A}$ is a partition of $\mathbb{Z}_N$. By construction, the range of G on any atom of $\mathcal{B}_{\epsilon,\eta}(G)$ has length at most ϵ. For any σ-algebra $\mathcal{B}$, since each atom of $\mathcal{B} \vee \mathcal{B}_{\epsilon,\eta}(G)$ is contained in some atom of $\mathcal{B}_{\epsilon,\eta}(G)$, the first property follows. Furthermore, note that since G only takes values on I, we observe that n only needs to range from $-2^{2^{k-1}}/\epsilon$ to $2^{2^{k-1}}/\epsilon$, thus only $O(1/\epsilon)$ atoms are needed to generate $\mathcal{B}_{\epsilon,\eta}(G)$, verifying the second property. To verify the third property, we let $\psi_\eta : \mathbb{R} \to [0,1]$ be a continuous piecewise linear function with support on $[-\eta, 1+\eta]$ such that it is 1 on $[\eta, 1-\eta]$ and is linear on $[-\eta, \eta]$ and on $[1-\eta, 1+\eta]$. For each atom $A \in \mathcal{B}_{\epsilon,\eta}(G)$ of the form $G^{-1}([\epsilon(n+\alpha), \epsilon(n+1+\alpha)))$, we define $\Psi_A(x) := \psi_\eta(x/\epsilon - n - \alpha)$. Then $\mathbf{1}_A - \Psi_A(G)$ pointwise bounded by $\mathbf{1}[G(x) \in [\epsilon(n-\eta+\alpha), \epsilon(n+\eta+\alpha)]] + \mathbf{1}[G(x) \in [\epsilon(n+1-\eta+\alpha), \epsilon(n+1+\eta+\alpha)]]$. Thus by (20), $||(\mathbf{1}_A - \Psi_A(G))(\nu+1)||_{L^1} = O(\eta)$. It is clear that Ψ_A can be chosen from a compact subset since n and α are bounded. Thus the third property is also verified, completing the proof. □

Having studied the σ-algebras generated by a single bounded function, we shall now study σ-algebras generated by multiple bounded functions. We will focus on the case involving basic Gowers anti-uniform functions. The next proposition is a direct generalization of Proposition 9.1. It allows us to translate the results we obtain for basic Gowers anti-uniform functions in section 6 into the language of σ-algebras.

Proposition 9.2. *Let ν be a k-pseudorandom measure, K be a fixed integer, $0 < \epsilon < 1, 0 < \eta < 1/2$ be real numbers. Let $DF_1, \cdots, DF_K$ be basic Gowers anti-uniform functions. Let $\mathcal{B}_{\epsilon,\eta}(DF_j)$ be the σ-algebra generated by DF_j. Let $\mathcal{B}_{\epsilon,\eta} := \mathcal{B}_{\epsilon,\eta}(DF_1) \vee \cdots \vee \mathcal{B}_{\epsilon,\eta}(DF_K)$ be the σ-algebra generated by $DF_1, \cdots, DF_K$. Then for η sufficiently small (depending on ϵ, K) and N sufficient large (depending on ϵ, K, η), we have*

$$||DF_j - \mathbb{E}(DF_j|\mathcal{B}_{\epsilon,\eta})||_{L^\infty} \leq \epsilon \text{ for all } j \tag{21}$$

and there exists an exceptional set $\Omega \in \mathcal{B}_{\epsilon,\eta}$ such that

$$\mathbb{E}((\nu+1)\mathbf{1}_\Omega) = O_{K,\epsilon}(\eta^{1/2}) \tag{22}$$

and

$$||(1-\mathbf{1}_\Omega)\mathbb{E}(\nu - 1|\mathcal{B}_{\epsilon,\eta})||_{L^\infty} = O_{K,\epsilon}(\eta^{1/2}). \tag{23}$$

Remark. Ω is *exceptional* because in practice, we will take $\eta \ll \epsilon$ to be small enough such that Ω is essentially empty.

Proof. (21) follows immediately from the first consequence of Proposition 9.1. For the remaining part, let Ω be the set of atoms A in $\mathcal{B}_{\epsilon,\eta}$ where $\mathbb{E}((\nu+1)\mathbf{1}_A) \leq \eta^{1/2}$. By construction, it is clear

that (22) holds. To show (23), first note that if $x \in \Omega$ then it is a trivial result. So suppose $x \notin \Omega$. We want to show $\mathbb{E}(\nu - 1|\mathcal{B}_{\epsilon,\eta})(x) = O_{K,\epsilon}(\eta^{1/2})$. Let A be the atom containing x. Then $\mathbb{E}((\nu+1)\mathbf{1}_A) > \eta^{1/2}$, and hence $\mathbb{E}(\mathbf{1}_A) \geq (\eta^{1/2} - \mathbb{E}((\nu-1)\mathbf{1}_A))/2$. From the third property of Proposition 9.1, followed by a simple induction on K using Hölder's inequality and the triangle inequality, we can find a continuous function $\Psi_A : I^K \to [0,1]$ such that

$$||(\nu+1)(\mathbf{1}_A - \Psi_A(DF_1, \cdots, DF_K))||_{L^1} = O_K(\eta). \tag{24}$$

On the other hand, by Proposition 6.2, we have

$$\mathbb{E}((\nu-1)(\Psi_A(DF_1, \cdots, DF_K))) = o_{K,\epsilon,\eta}(1) \tag{25}$$

since Ψ_A can be chosen from a compact subset of $C^0(I^K)$ depending only on K, ϵ, η. From (24) and (25), we have

$$\mathbb{E}((\nu-1)\mathbf{1}_A) = o_{K,\epsilon,\eta}(1) + O_{K,\epsilon}(\eta).$$

Hence

$$\mathbb{E}(\nu - 1|A) = \frac{\mathbb{E}((\nu-1)\mathbf{1}_A)}{\mathbb{E}(\mathbf{1}_A)} \leq \frac{\mathbb{E}((\nu-1)\mathbf{1}_A)}{(\eta^{1/2} - \mathbb{E}((\nu-1)\mathbf{1}_A))/2} = o_{K,\epsilon,\eta}(1) + O_{K,\epsilon}(\eta^{1/2}),$$

from which (23) is verified. □

Proposition 9.2 gives us the machinery to obtain the following *Furstenberg tower* in an iterative manner, which is the key step (or better described as an algorithm) in proving Theorem 8.1.

Proposition/Algorithm 9.3 (Furstenberg tower). *Let ν be a k-pseudorandom measure, and $f : \mathbb{Z}_N \to \mathbb{R}^+$ be pointwise bounded by ν. Let $0 < \epsilon < 1$ be a parameter chosen up front (which can be made arbitrarily small), let K be a nonnegative integer (which we need to iterate over in the algorithm), and let $0 < \eta \ll \epsilon$ (i.e. η should be thought of as being small compared to ϵ). Suppose $F_1, \cdots, F_K : \mathbb{Z}_N \to \mathbb{R}$ satisfy*

$$|F_j(x)| \leq (1 + O_{K,\epsilon}(\eta^{1/2}))(\nu(x) + 1) \tag{26}$$

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

$$\mathcal{B}_K := \mathcal{B}_{\epsilon,\eta}(DF_1) \vee \cdots \vee \mathcal{B}_{\epsilon,\eta}(DF_K).$$

Suppose there is an exceptional set $\Omega_K \subset \mathbb{Z}_N$ such that

$$\mathbb{E}((\nu+1)\mathbf{1}_{\Omega_K}) = O_{K,\epsilon}(\eta^{1/2}) \tag{27}$$

and

$$||(1 - \mathbf{1}_{\Omega_K})\mathbb{E}(\nu - 1|\mathcal{B}_K)||_{L^\infty} = O_{K,\epsilon}(\eta^{1/2}). \tag{28}$$

If $F_{K+1} := (1 - \mathbf{1}_{\Omega_K})(f - \mathbb{E}(f|\mathcal{B}_K))$ *satisfies*

$$||F_{K+1}||_{U^{k-1}} > \epsilon^{1/2^k}, \tag{29}$$

i.e. if F_{K+1} *is not Gowers uniform enough for the smallness level* ϵ *that we choose up front, then we have (and none of the following would apply if* F_{K+1} *is sufficiently Gowers uniform):*

$$||(1 - \mathbf{1}_{\Omega_K})\mathbb{E}(f|\mathcal{B}_K)||_{L^\infty} \leq 1 + O_{K,\epsilon}(\eta^{1/2}) \tag{30}$$

and

$$|F_{K+1}(x)| \leq (1 + O_{K,\epsilon}(\eta^{1/2}))(\nu(x) + 1). \tag{31}$$

Furthermore, let $\mathcal{B}_{K+1} := \mathcal{B}_K \vee \mathcal{B}_{\epsilon,\eta}(DF_{K+1})$*, then there is an exceptional set* $\Omega_{K+1} \supset \Omega_K$ *such that*

$$\mathbb{E}((\nu + 1)\mathbf{1}_{\Omega_{K+1}}) = O_{K,\epsilon}(\eta^{1/2}) \tag{32}$$

$$||(1 - \mathbf{1}_{\Omega_{K+1}})\mathbb{E}(\nu - 1|\mathcal{B}_{K+1})||_{L^\infty} = O_{K,\epsilon}(\eta^{1/2}) \tag{33}$$

and

$$||(1 - \mathbf{1}_{\Omega_{K+1}})\mathbb{E}(f|\mathcal{B}_{K+1})||_{L^2}^2 \geq ||(1 - \mathbf{1}_{\Omega_K})\mathbb{E}(f|\mathcal{B}_K)||_{L^2}^2 + 2^{-2^k+1}\epsilon. \tag{34}$$

Remark. This proposition might seem overly technical and daunting at first glance, but it really is a formalism of the idea briefly introduced in section 6, after the proof of Proposition 6.1. To re-emphasize, the strategy is the following. We want to decompose an arbitrary function $f \in L^1(\mathbb{Z}_N)$ into a Gowers uniform part and a Gowers anti-uniform part. The idea is that we project f into a larger and larger subspace until the residual is small enough (in terms of Gowers uniformity norm). This proposition provides us with a concrete algorithm of achieving this (through the machinery of conditional expection restricted to some σ-algebra). The mechanism that guarantees this alrorithm will stop in a finite number of steps is: the L^2 norm of $\mathbb{E}(f)$ (also defined as the *energy* in ergodic theory) will increase by a constant amount during each iteration, and by construction this quantity is bounded above, thus by the time the quantity exceeds the bound, the algorithm must have stopped.

To understand the Proposition more quantitatively, one shall think of F_K as being Gowers anti-uniform (in fact at the termination point of the algorithm, this will be the $f_{U^\perp}$ in the proof of Theorem 8.1), and $O_{K,\epsilon}(\eta^{1/2})$ as being essentially $o_\epsilon(1)$. (29) is the condition for which the algorithm will be repeated (or the converse of (29) gives the condition for which the algorithm will stop). (31), (32), (33) is saying that the properties (26), (27), (28) are invariant throughout the iteration (which allows us to run the algorithm). (30) is saying that the L^2 norm of $\mathbb{E}(f)$ is bounded above, while (34) is saying it will increase by a fixed amount during each iteration. Thus these 2

together guarantees the algorithm will stop. The role that Proposition 9.2 plays is that it provides us with a concrete method of coming up with the set Ω_{K+1}, which ensures us that the algorithm can be continued while maintaining the properties (26), (27), (28).

Before proving Proposition 9.3, let us quickly prove Theorem 8.1 assuming we have Proposition 9.3.

Proof. We will run the algorithm as presented in Proposition 9.3 until it terminates, at which point we conclude all the properties required by Theorem 8.1 are satisfied. To start, we pick an arbitrary tolerance level ϵ, and let $K_0 := \lceil 2^{2^k}/\epsilon + 1 \rceil$. The parameter η will be chosen later, which would be small compared to ϵ. We run the following:

- Initialization step. Initialize $K := 0$, $\Omega_0 := \emptyset$.
- Iterative step. At the K-th step, we have $\mathcal{B}_K$, Ω_K, and $DF_1, \cdots, DF_K, DF_{K+1}$ as in Proposition 9.3. There are two possible cases:
  - If $||F_{K+1}||_{U^{k-1}} \leq \epsilon^{1/2^k}$ (i.e. the condition for continuation fails): then we terminate the algorithm by setting $\Omega := \Omega_K$, $\mathcal{B} := \mathcal{B}_K$.
  - Else if $||F_{K+1}||_{U^{k-1}} > \epsilon^{1/2^k}$ and $K \leq K_0$ (so we will continue): then we define $\mathcal{B}_{K+1}$ as in Proposition 9.3 (note that we will already have DF_{K+1} from the K-th (previous) iteration), and choose Ω_{K+1} as in Proposition 9.2. By Proposition 9.3, the invariance of the properties are verified. Then we increment K to $K+1$ and repeat the Iterative step.
- Force-quit step. If we ever reach the situation in which $K > K_0$: then we force the algorithm to terminate and set $\Omega := \Omega_{K_0}$, $\mathcal{B} := \mathcal{B}_{K_0}$. (As we will prove below, this shall never be reached by any correct implementation of the algorithm, and the algorithm will terminate successfully at some step K for which $K \leq K_0$.)

It is clear that if the algorithm terminates successfully at some step $K \leq K_0$, then by (27), (28), and (the converse of) (29), we know that (17), (18), and (19) from Theorem 8.1 will be satisfied, except that we will have $O_{K,\epsilon}(\eta^{1/2})$ instead of $o_\epsilon(1)$ in (17) and (18). By letting η go to 0 for fixed ϵ, we can indeed replace $O_{K,\epsilon}(\eta^{1/2})$ by $o_\epsilon(1)$. Thus it remains to verify that the Force-quit step will not be reached. As mentioned before, this will be guaranteed by (30) and (34). Specifically, suppose for a contradiction that K_0 is reached. Then by (34), we must have

$$||(1 - \mathbf{1}_{\Omega_{K_0}})\mathbb{E}(f|\mathcal{B}_{K_0})||_{L^2}^2 \geq K_0 2^{-2^k+1}\epsilon \geq 2.$$

But on the other hand, by (30), we must have

$$||(1 - \mathbf{1}_{\Omega_{K_0}})\mathbb{E}(f|\mathcal{B}_{K_0})||_{L^2}^2 \leq ||(1 - \mathbf{1}_{\Omega_{K_0}})\mathbb{E}(f|\mathcal{B}_{K_0})||_{L^\infty}^2 \leq 1 + O_{K,\epsilon}(\eta^{1/2}).$$

By letting $\eta \to 0$ for fixed ϵ, we get a contradiction. Theomem 8.1 thus follows. □

To complete the proof of Theorem 8.1, it remains to prove Proposition 9.3. As we will see shortly, the invariance of (26), (27), and (28) is fairly easy to obtain. The major task is to ensure the fixed increment of the L^2 norm of $\mathbb{E}(f|\mathcal{B}_K)$ as we increment K.

Proof. From (28), by writng $\nu = (\nu - 1) + 1$ and applying the triangle inequality, we have

$$||(1 - \mathbf{1}_{\Omega_K})\mathbb{E}(\nu|\mathcal{B}_K)||_{L^\infty} = 1 + O_{K,\epsilon}(\eta^{1/2}).$$

Since f is pointwise bounded by ν, (30) follows. From (30) and the definition of F_{K+1}, again applying the triangle inequality and noting that f is bounded by ν, we get

$$|F_{K+1}(x)| \leq (1 + O_{K,\epsilon}(\eta^{1/2})) + \nu(x),$$

and thus in particular (31) follows. From (26), each DF_j is a basic Gowers anti-uniform function (up to a multiplicative factor of $1 + O_{K,\epsilon}(\eta^{1/2})$ which is negligible), thus by Proposition 9.2, there exists an exceptional set $\Omega'_{K+1} \subset \mathcal{B}_{K+1}$ such that (22) and (23) holds. Set $\Omega_{K+1} := \Omega_K \cup \Omega'_{K+1}$. Then (33) can be directly verified from (23); and by noting $\mathbf{1}_{\Omega_{K+1}} \leq \mathbf{1}_{\Omega_K} + \mathbf{1}_{\Omega'_{K+1}}$, (32) can be verified from (22) and (27). So it remains to verify (34), i.e. the increment of the L^2 norm of $\mathbb{E}(f|\mathcal{B})$. To do that, we will need to use condition (29), estimate certain inner products that are relatively easy to compute, then use the triangle inequality to derive estimates of the desired quantity. The computation is fairly involved and routine, with the main trick being to identify orthogonality relationship between certain functions and take advantage of that. The details are presented as follows.

By the definition of F_{K+1} and (10), since $||F_{K+1}||_{U^{k-1}} > \epsilon^{1/2^k}$, we have

$$|\langle (1 - \mathbf{1}_{\Omega_K})(f - \mathbb{E}(f|\mathcal{B}_K)), DF_{K+1}\rangle| = |\langle F_{K+1}, DF_{K+1}\rangle| = ||F_{K+1}||_{U^k-1}^{2^{k-1}} > \epsilon^{1/2}. \tag{35}$$

On the other hand, from (26), we have

$$\begin{aligned}|\langle (\mathbf{1}_{\Omega_{K+1}} - \mathbf{1}_{\Omega_K})(f - \mathbb{E}(f|\mathcal{B}_K)), DF_{K+1}\rangle| \leq & ||DF_{K+1}||_{L^\infty}\mathbb{E}(|(\mathbf{1}_{\Omega_{K+1}} - \mathbf{1}_{\Omega_K})(f - \mathbb{E}(f|\mathcal{B}_K))|) \\ \leq & ||DF_{K+1}||_{L^\infty}\mathbb{E}((\mathbf{1}_{\Omega_{K+1}} - \mathbf{1}_{\Omega_K})(1 + O_{K,\epsilon}(\eta^{1/2}))(\nu + 1)))\end{aligned}$$

where the last inequality follows by noting that if $x \in \Omega_K$, then it contributes nothing to the right hand side, and thus we only need to consider the case $x \notin \Omega_K$ so that $f - \mathbb{E}(f|\mathcal{B}_K) = F_{K+1}$. Since $||DF_{K+1}||_{L^\infty} = O_{K,\epsilon}(1)$, from the above, (27), and (32), we obtain

$$|\langle (\mathbf{1}_{\Omega_{K+1}} - \mathbf{1}_{\Omega_K})(f - \mathbb{E}(f|\mathcal{B}_K)), DF_{K+1}\rangle| \leq O_{K,\epsilon}(\eta^{1/2}). \tag{36}$$

Now note that

$$\begin{aligned}&|\langle (1 - \mathbf{1}_{\Omega_{K+1}})(f - \mathbb{E}(f|\mathcal{B}_K)), DF_{K+1} - \mathbb{E}(DF_{K+1}|\mathcal{B}_{K+1})\rangle| \\ \leq & ||DF_{K+1} - \mathbb{E}(DF_{K+1}|\mathcal{B}_{K+1})||_{L^\infty}\mathbb{E}(|(1 - \mathbf{1}_{\Omega_{K+1}})(f - \mathbb{E}(f|\mathcal{B}_K))|).\end{aligned}$$

By (21) and (30), we thus have

$$|\langle(1-\mathbf{1}_{\Omega_{K+1}})(f-\mathbb{E}(f|\mathcal{B}_K)), DF_{K+1}-\mathbb{E}(DF_{K+1}|\mathcal{B}_{K+1})\rangle| \leq O(\epsilon). \tag{37}$$

Differencing (35) and (36) and then applying the triangle inequality with (37), we obtain

$$|\langle(1-\mathbf{1}_{\Omega_{K+1}})(f-\mathbb{E}(f|\mathcal{B}_K)), \mathbb{E}(DF_{K+1}|\mathcal{B}_{K+1})\rangle| \geq \epsilon^{1/2} - O_{K,\epsilon}(\eta^{1/2}) - O(\epsilon). \tag{38}$$

Noting that the inner product is an expectation $\mathbb{E}(\cdot|\mathcal{B})$ where $\mathcal{B} := 2^{\mathbb{Z}_N}$ is the power set of $\mathbb{Z}_N$, by introducing the sub-algebra $\mathcal{B}_{K+1}$ and invoking the Law of Iterated Expectation, (38) can be rewritten as

$$|\langle(1-\mathbf{1}_{\Omega_{K+1}})(\mathbb{E}(f|\mathcal{B}_{K+1})-\mathbb{E}(f|\mathcal{B}_K)), \mathbb{E}(DF_{K+1}|\mathcal{B}_{K+1})\rangle| \geq \epsilon^{1/2} - O_{K,\epsilon}(\eta^{1/2}) - O(\epsilon).$$

Now by the Cauchy Schwarz inequality with respect to this inner product (which induces the L^2 norm), and noting that $||\mathbb{E}(DF_{K+1}|\mathcal{B}_{K+1})||_{L^2} \leq ||\mathbb{E}(DF_{K+1}|\mathcal{B}_{K+1})||_{L^\infty} \leq ||DF_{K+1}||_{L^\infty} = 2^{2^{k-1}-1} + O_{K,\epsilon}(\eta^{1/2})$ (since it is basic Gowers anti-uniform), we have

$$||(1-\mathbf{1}_{\Omega_{K+1}})(\mathbb{E}(f|\mathcal{B}_{K+1})-\mathbb{E}(f|\mathcal{B}_K))||_{L^2} \geq 2^{-2^{k-1}+1}\epsilon^{1/2} - O_{K,\epsilon}(\eta^{1/2}) - O(\epsilon). \tag{39}$$

Note that if Ω_K stays empty throughout the algorithm, then (39) would imply (34) since $\mathbb{E}(f|\mathcal{B}_{K+1})$ $-\mathbb{E}(f|\mathcal{B}_K)$ is orthogonal to $\mathbb{E}(f|\mathcal{B}_K)$. But the presence of Ω_K requires us to do further estimations. Again, our tool is the triangle inequality.

Comparing (39) and the left hand side of (34), we would like to write $\mathbb{E}(f|\mathcal{B}_{K+1})$ as $[\mathbb{E}(f|\mathcal{B}_K)] + [\mathbb{E}(f|\mathcal{B}_{K+1}) - \mathbb{E}(f|\mathcal{B}_K)]$, and expand the left hand side of (34) as an inner product to obtain:

$$\begin{aligned}||(1-\mathbf{1}_{\Omega_{K+1}})\mathbb{E}(f|\mathcal{B}_{K+1})||_{L^2}^2 =& ||(1-\mathbf{1}_{\Omega_{K+1}})\mathbb{E}(f|\mathcal{B}_K)||_{L^2}^2 + ||(1-\mathbf{1}_{\Omega_{K+1}})(\mathbb{E}(f|\mathcal{B}_{K+1})-\mathbb{E}(f|\mathcal{B}_K))||_{L^2}^2 \\ &+ 2\langle(1-\mathbf{1}_{\Omega_{K+1}})\mathbb{E}(f|\mathcal{B}_K), (1-\mathbf{1}_{\Omega_{K+1}})(\mathbb{E}(f|\mathcal{B}_{K+1})-\mathbb{E}(f|\mathcal{B}_K))\rangle. \ (*)\end{aligned}$$

We have the estimate for the second term from (39). To estimate the first term, note that the difference between it and the term on the right hand side of (34) is the set Ω_K, thus it is natural to consider the difference $\mathbf{1}_{\Omega_{K+1}} - \mathbf{1}_{\Omega_K}$. Note that for $x \notin \Omega_K$, we have $|\mathbb{E}(f|\mathcal{B})| \leq 1 + O_{K,\epsilon}(\eta^{1/2}) \leq 2$ for η small enough by (30). Thus by (32),

$$||(\mathbf{1}_{\Omega_{K+1}} - \mathbf{1}_{\Omega_K})\mathbb{E}(f|\mathcal{B}_K)||_{L^2}^2 \leq 2||(\mathbf{1}_{\Omega_{K+1}} - \mathbf{1}_{\Omega_K})||_{L^2}^2 \leq 2\mathbb{E}(\mathbf{1}_{\Omega_{K+1}}) \leq O_{K,\epsilon}(\eta^{1/2}). \tag{40}$$

Writing $1-\mathbf{1}_{\Omega_K}$ as $(\mathbf{1}_{\Omega_{K+1}} - \mathbf{1}_{\Omega_K}) + (1-\mathbf{1}_{\Omega_{K+1}})$, by the triangle inequality we have

$$||(1-\mathbf{1}_{\Omega_K})\mathbb{E}(f|\mathcal{B}_K)||_{L^2}^2 \leq (||(\mathbf{1}_{\Omega_{K+1}} - \mathbf{1}_{\Omega_K})\mathbb{E}(f|\mathcal{B}_K)||_{L^2} + ||(1-\mathbf{1}_{\Omega_{K+1}})\mathbb{E}(f|\mathcal{B}_K)||_{L^2})^2.$$

Expanding the square and using (30), combined with (40), we obtain

$$||(1-\mathbf{1}_{\Omega_K})\mathbb{E}(f|\mathcal{B}_K)||_{L^2}^2 \leq ||(1-\mathbf{1}_{\Omega_{K+1}})\mathbb{E}(f|\mathcal{B}_K)||_{L^2}^2 + O_{K,\epsilon}(\eta^{1/4}). \tag{41}$$

As we have seen several times, the $O_{K,\epsilon}(\eta^{1/4})$ will not matter since η will be made small for fixed ϵ. Thus we have successfully estimated the first term. Now it remains to estimate the cross term. Noting that $(1-\mathbf{1}_{\Omega_{K+1}})^2 = (1-\mathbf{1}_{\Omega_{K+1}})$, the cross term (ignoring the factor of 2) can be written as

$$\langle(\mathbf{1}_{\Omega_K} - \mathbf{1}_{\Omega_{K+1}})\mathbb{E}(f|\mathcal{B}_K), \mathbb{E}(f|\mathcal{B}_{K+1}) - \mathbb{E}(f|\mathcal{B}_K)\rangle \tag{42}$$

since the difference between the two expressions will be 0 by observing that $\mathbb{E}(f|\mathcal{B}_{K+1}) - \mathbb{E}(f|\mathcal{B}_K)$ is orthogonal to $L^2(\mathcal{B}_K)$. Furthermore, (42) equals

$$\langle(\mathbf{1}_{\Omega_K} - \mathbf{1}_{\Omega_{K+1}})\mathbb{E}(f|\mathcal{B}_K), f - \mathbb{E}(f|\mathcal{B}_K)\rangle$$

by interpreting the inner product as an expectation and applying the Law of Iterated Expectation with the sub-algebra $\mathcal{B}_{K+1}$. Again, note that for $x \notin \Omega_K$ (which is the relevant case, since the expression will be 0 otherwise), $||\mathbb{E}(f|\mathcal{B}_K)||_{L^\infty} \leq 1 + O_{K,\epsilon}(\eta^{1/2}) \leq 2$ for η small enough by (30), and $f - \mathbb{E}(f|\mathcal{B}_K)$ is bounded by $(1 + O_{K,\epsilon}(\eta^{1/2}))(\nu + 1)$, thus by (27) and (32) we have

$$\langle(\mathbf{1}_{\Omega_K} - \mathbf{1}_{\Omega_{K+1}})\mathbb{E}(f|\mathcal{B}_K), f - \mathbb{E}(f|\mathcal{B}_K)\rangle = O_{K,\epsilon}(\eta^{1/2}). \tag{43}$$

Thus the cross term is $O_{K,\epsilon}(\eta^{1/2})$. Combined with (*), (39), (41), and letting $\eta \to 0$ for fiexed ϵ, we have established (34). This completes the proof of Proposition 9.3, and thus Theorem 8.1. □

10 Proof of the Green-Tao Theorem

Assuming Proposition 3.3 (Goldston-Yıldırım), combined with the Generalized Szemerédi's Theorem that we proved, we are now ready to prove the Green-Tao Theorem:

Theorem 10.1 (Green-Tao Theorem). *The prime numbers contain infinitely many arithmetic progressions of length k for each positive integer k.*

Proof. As mentioned in section 3, we first need to verify $\mathbb{E}(f)$ is bounded below for our choice of f. A strong version of the Dirichlet Theorem (which is the Siegel-Walfisz theorem) allows us to conclude this. Then by the Generalized Szemerédi's Theorem, the product of f evaluated at an arithmetic progression of length k will have an expectation that is bounded below. Since the support of f is contained in the primes, it follows that there is an arithmetic progression of length k in the primes each time the product is nonzero. To conclude the Green-Tao theorem, we will need to verify that each arithmetic progression counted this way is a genuine arithmetic progression. But this follows from a simple argument.

Let f and ν be as in Proposition 3.3. To use the Generalized Szemerédi's Theorem on f, it remains to verify that $\mathbb{E}(f) \geq \delta$ for some positive constant δ which does not depend on N. Recall that we use $\mathbb{P}$ to denote the set of primes. Then

$$\mathbb{E}(f) = \frac{k^{-1}2^{-k-5}}{N} \sum_{\epsilon_k N \leq n \leq 2\epsilon_k N} \frac{\phi(W)}{W} \log(Wn+1)\mathbf{1}[Wn+1 \in \mathbb{P}]$$

By the Siegel-Walfisz theorem[3], if $q = O((\log x)^m)$ for some real number m, then

$$\psi_q(x) := \sum_{n \leq x, n \equiv 1 \mod q} \Lambda(n) = \frac{1}{\phi(q)}(1 + o(1))x.$$

Since $w(N) \log w(N) = O(\log \log N)$, we have $w(N) \log w(N) \leq m \log \log N$ for some m for all sufficiently large N. Thus

$$W = e^{\sum_{p \leq w(N)} \log p} \leq e^{w(N) \log w(N)} \leq e^{m \log \log N} = (\log N)^m$$

and thus $W = O((\log(W\epsilon_k N + 1))^m)$ and $W = O((\log(2W\epsilon_k N + 1))^m)$. Therefore

$$\psi_W(W\epsilon_k N + 1) = \frac{1}{\phi(W)}(1 + o(1))(W\epsilon_k N + 1)$$

and

$$\psi_W(2W\epsilon_k N + 1) = \frac{1}{\phi(W)}(1 + o(1))(2W\epsilon_k N + 1)$$

Observe that if n' is a prime in $[W\epsilon_k N + 1, 2W\epsilon_k N + 1]$, and $(n')^m$ is a prime power in $[W\epsilon_k N + 1, 2W\epsilon_k N + 1]$, then $m \leq \log_{n'}(2W\epsilon_k N + 1) \leq \log_{(W\epsilon_k N+1)}(2W\epsilon_k N + 1)$. Thus for sufficiently large N, we have $m < 2$. This implies that higher powers of primes can be discarded.

Hence for sufficently large N,

$$\begin{aligned}
\mathbb{E}(f) &= \frac{k^{-1}2^{-k-5}}{N} \frac{\phi(W)}{W} \sum_{\epsilon_k N \leq n \leq 2\epsilon_k N} \log(Wn + 1)\mathbf{1}[Wn + 1 \in \mathbb{P}] \\
&= \frac{k^{-1}2^{-k-5}}{N} \frac{\phi(W)}{W} \sum_{W\epsilon_k N + 1 \leq Wn + 1 \leq 2W\epsilon_k N + 1} \log(Wn + 1)\mathbf{1}[Wn + 1 \in \mathbb{P}] \\
&= \frac{k^{-1}2^{-k-5}}{N} \frac{\phi(W)}{W} \sum_{W\epsilon_k N + 1 \leq n' \leq 2W\epsilon_k N + 1} \Lambda(n') \\
&= \frac{k^{-1}2^{-k-5}}{N} \frac{\phi(W)}{W} (\psi_W(2W\epsilon_k N + 1) - \psi_W(W\epsilon_k N + 1)) \\
&= \frac{k^{-1}2^{-k-5}}{N} \frac{\phi(W)}{W} (\frac{1}{\phi(W)}(1 + o(1))W\epsilon_k N) \\
&= k^{-1}2^{-k-5}\epsilon_k(1 + o(1)).
\end{aligned}$$

[3]If $w(N)$ is chosen to be a constant (which is valid as noted in Proposition 3.3), then Dirichlet's Theorem can be applied instead to obtain the same result. But in greater generality, $w(N)$ could be a slowly growing function of N rather than a constant, thus Dirichlet's Theorem does not quite suffice, and we would need Siegel-Walfisz theorem instead. The author thanks N. Elkies for pointing this out.

Thus invoking the Generalized Szemerédi's Theorem on f, we obtain

$$\mathbb{E}(f(x)f(x+r)\cdots f(x+(k-1)r)|x,r\in\mathbb{Z}_N)\geq c(k,k^{-1}2^{-k-5}\epsilon_k)-o_k(1). \tag{44}$$

Note that for each $x,r\in\mathbb{Z}_N$, $f(x)f(x+r)\cdots f(x+(k-1)r)=O_k(\log^k N)$. Thus by (44), the number of arithmetic progressions in $\mathbb{Z}_N$ is $\Omega_k(N^2/\log^k N)$ (we say $f=\Omega(g)$ if $g=O(f)$). Also note that the number of arithmetic progressions corresponding to $r=0$ is at most $O(N)$, thus the number of nontrivial arithmetic progressions in $\mathbb{Z}_N$ is still $\Omega_k(N^2/\log^k N)$.

Finally, to obtain the Green-Tao Theorem, it suffices to verify that each arithmetic progression counted by this expression is a genuine arithmetic progression, not just an arithmetic progression in $\mathbb{Z}_N$.

Proposition 10.2. *If $f(x)f(x+r)\cdots f(x+(k-1)r)\neq 0$, then it counts a genuine arithmetic progression.*

Proof. If $f(x)f(x+r)\cdots f(x+(k-1)r)\neq 0$, then for each $i=0,1,\cdots,k-1$, we have $\epsilon_k N+m_iN\leq x+ir\leq 2\epsilon_k N+m_iN$ for some nonnegative integer m_i. Define

$$j:=\max_{0\leq i\leq k-1}\{i:\epsilon_k N\leq x+ir\leq 2\epsilon_k N\}.$$

Since $\epsilon_k N\leq x\leq 2\epsilon_k N$, we have $m_0=0$, and j is well defined. It is clear that if $j=k-1$, then $x,x+r,\cdots,x+(k-1)r$ is a genuine arithmetic progression. Suppose $1\leq j<k-1$. Then we must have $x+jr+r>N$. Since $2\epsilon_k N\geq x+jr$, this implies $2\epsilon_k N+r>N$, or $r>(1-2\epsilon_k)N$. On the other hand, $r\leq\frac{2\epsilon_k N-x}{j}\leq 2\epsilon_k N$, thus combining these 2 inequalities, we obtain $(1-2\epsilon_k)N<2\epsilon_k N$. But this implies $\epsilon_k>\frac{1}{4}$, which is a contradiction.

Now suppose $j=0$. We must then have $x+r\geq\epsilon_k N+N$. Since $x\leq 2\epsilon_k N$, we have $r\geq\epsilon_k N+N-x\geq(1-\epsilon_k)N$. In particular, $r>\epsilon_k N$, thus we must have $m_i=m_{i-1}+1$ for all $i=1,2,\cdots,k-1$. Hence $m_i=i$ for all $i=0,1,\cdots,k-1$. It then follows that $x,x+(r-N),\cdots,x+(k-1)(r-N)$ is a genuine arithmetic progression (in decreasing order). □

By Proposition 10.2, each *increasing* arithmetic progression will be counted twice by the expression $f(x)f(x+r)\cdots f(x+(k-1)r)$ (if it is ever counted). Thus the number of increasing arithmetic progressions with elements in the set $S_N:=\{p\in\mathbb{P}:p\leq N,p\equiv 1 \mod W\}$ is at least $\frac{1}{2}\Omega_k(N^2/\log^k N)=\Omega_k(N^2/\log^k N)$. Since $S_N\subset\mathbb{P}$ for each N, taking $N\to\infty$, the Green-Tao Theorem follows. □

11 Linear Forms and Correlation Estimates on ν (Goldston-Yıldırım)

To complete the proof of the Green-Tao theorem, we would need to prove Proposition 3.3. The method of the proof is very closely related to the arguments presented in Goldston and Yıldırım's

proof of small gaps between primes ([5]). The actual proof is fairly technical in nature and we are not going to present the complete proof here. Instead, we are going to quote the following 2 propositions, from which the linear forms condition and correlation condition shall be verified for ν.

Proposition 11.1 (Linear Forms Condition) (Goldston-Yıldırım). *Let m, t be positive integer parameters. Let $L := (L_{ij})_{1\leq i\leq m, 1\leq j\leq t}$ be a m-by-t matrix with integer coefficients L_{ij} such that no two rows of L are rational multiples of each other, and no row consists entirely of 0. Further, suppose $|L_{ij}| \leq \sqrt{w(N)}/2$ for all i, j. Let b be an arbitrary fixed element of $\mathbb{Z}_N^t$. Define $\psi : \mathbb{Z}_N^t \to \mathbb{Z}_N^m$ as $\psi(x) = Lx + b$. Denote $\theta_i := W\psi_i + 1$. Let $B := \prod_{i=1}^t I_i \subset \mathbb{R}$ be a product of t intervals I_i, each of which has length at least R^{10m}, where R is as defined in Proposition 3.3. Then for a sufficiently slowly growing function $w(N)$ (which we assume), we have*

$$\mathbb{E}(\Lambda_R(\theta_1(x))^2 \cdots \Lambda_R(\theta_m(x))^2 | x \in B) = (1 + o_{m,t})(\frac{W \log R}{\phi(W)})^m,$$

where $\Lambda_R(\cdot)$ is as defined in Proposition 3.3.

Proposition 11.2 (Correlation Condition) (Goldston-Yıldırım). *Let m be a positive integer parameter, and let B be an interval of length at least R^{10m} where R is as defined in Proposition 3.3. Suppose $h_1, \cdots, h_m$ are distinct integers bounded by N^2 in absolute value. Let $\Delta := \prod_{1\leq i<j\leq m} |h_i - h_j|$. Then for sufficiently large N (depending on m) and a sufficiently slowly growing function $w(N)$, we have*

$$\mathbb{E}(\Lambda_R(W(x + h_1) + 1)^2 \cdots \Lambda_R(W(x + h_m) + 1)^2 | x \in B) \leq (1 + o_m(1))(\frac{W \log R}{\phi(W)})^m \prod_{p|\Delta, p\in\mathbb{P}} (1 + O_m(p^{-\frac{1}{2}})),$$

where $\Lambda_R(\cdot)$ is as defined in Proposition 3.3.

One should note that Proposition 11.1 and Proposition 11.2 are almost identical to the linear forms condition and correlation condition, respectively. However, there are some technical difficulties that prevent us from directly arriving at the desired condition from the 2 propositions. In order to verify the linear forms condition, because of the piecewise definition of ν in Proposition 3.3, we will need to estimate the left hand side of Definition 4.1 by dividing $\mathbb{Z}_N^t$ into several parts and estimate each part separately. And in order to verify the correlation condition, we will need to construct a weight function τ with the required properties. One choice of τ motivated by Proposition 11.2 would be

$$\tau(n) := \tau_m(n) := O_m(1) \prod_{p|n, p\in\mathbb{P}} (1 + p^{-\frac{1}{2}})^{O_m(1)}$$

for $n \neq 0$, and

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

for some large constant C.

We can then verify τ have the desired properties and thus the correlation condition.

The proof of Proposition 11.1 and Proposition 11.2 is largely due to Goldston and Yıldırım, and it involves a substantial amount of complex analysis. The idea is to express the expectation in terms of certain contour integrals, and use complex analytic mathods to estimate those integrals. The details are fairly involved and are omitted from the paper. Interested readers can find related arguments in [5] and a complete proof of Proposition 11.1 and Proposition 11.2 in [7].

12 A Note on Szemerédi's Theorem

In this section, we are going to briefly discuss the significance of Szemerédi's theorem and the main ideas in the proof. There are at least three types of proofs: Szemerédi's orginal combinatorial proof using the Szemerédi regularity lemma ([9]); Furstenberg's ergodic theoretic proof using a Furstenberg tower ([3],[4]); and Gowers' harmonic analytic proof using the theory of additive combinatorics ([6]). In fact, each proof has far reaching implications than the establishment of Szemerédi's theorem itself. Szemerédi's proof extends the famous van der Waerden theorem[4] (1927) and the earlier Szmeredi regularity lemma, which has a huge application in graph theory and theoretical computer science. Furstenberg's proof makes a beautiful connection between additive combinatorics and ergodic theory, and actually gives rise to a new brach of mathematics known as *Ergodic Ramsey theory*[5]. Gowers' proof gives an explicit and relatively strong bound on how large N needs to be, which, if further optimized, could potentially lead to the Green-Tao Theorem by more elementary arguments.

Although the arguments of these proofs are very different in nature, they do share a common technique, which is to break the object under consideration into a *structured component* and an *error component*. The idea is then to establish some structure theorem which would be applied to the structured component and lead to the result, and to establish some error-controling theorem which would allow us to discard the negligible error. By observing such common theme in these existing proofs, Tao is able to give yet another proof of the Szemerédi's theorem using ergodic theory ([10]), which is more elementary compared to Furstenberg's proof in the sense that it remains in the finitary setting $\mathbb{Z}_N$ and does not involve the axiom of choice. In fact the proof of the Generalized Szemerédi's theorem in this paper greatly reflects this theme. The object under consideration here is the product $\mathbb{E}(f(x)f(x+r)\cdots f(x+(k-1)r)|x,r\in\mathbb{Z}_N)$; the structured component is $f_{U^\perp}$; the error component is f_U; the structure theorem is the Generalized Koopman-von Neumann Structure Theorem (Theorem 8.1); the error-controling theorem is the Generliazed von Neumann Theorem (Theorem 5.4). Such identification in Tao's proof of the Szemerédi's theorem is almost identical.

[4]The van der Waerden theorem states the following. For any positive integers k, r, there exists some number N such that for every coloring $c:\{1,2,\cdots,N\}\to\{1,2,\cdots,m\}$, $\{1,2,\cdots,N\}$ contains an arithmetic progression of length k on which c is constant.

[5]For readers unfamiliar with the field, it is a branch of mathematics where problems motivated by additive combinatorics are solved by ergodic theory methods.

Since Tao's proof of the Szemerédi's theorem is greatly parallel to the arguments in this paper, we choose to examine it in more detail (as opposed to other proofs). Instead of presenting Tao's proof of the Szemerédi's theorem in its entirety, we are going to present the main ideas only. The following 3 theorems are the main ingredients in the proof.

Theorem 12.1 (Generalized von Neumann Theorem). *Let $k \geq 2$ be a parameter, and $\lambda_0, \cdots, \lambda_{k-1}$ be distinct elements of $\mathbb{Z}_N$. Then for any bounded functions $f_0, f_1, \cdots, f_{k-1} : \mathbb{Z}_N \to \mathbb{C}$, we have*

$$|\mathbb{E}(f_0(x - \lambda_0 r) \cdots f_{k-1}(x - \lambda_{k-1} r)|x, r \in \mathbb{Z}_N)| \leq \min_{1 \leq j \leq k} ||f_j||_{U^{k-1}}.$$

Theorem 12.2 (Recurrence). *Let $d \geq 0$, $k \geq 1$ be integer parameters and $\delta > 0$. Let $f_{U^\perp}$, f_{UAP} be nonnegative bounded functions such that* [6]

$$||f_{U^\perp} - f_{UAP}||_{L^2} \leq \frac{\delta^2}{1024k} \tag{45}$$

$$\mathbb{E}(f_{U^\perp}) \geq \delta \tag{46}$$

$$||f_{UAP}||_{UAP^d} < M \tag{47}$$

Then for all $\mu \in \mathbb{Z}_N$ and $N_1 \geq 1$, we have

$$\mathbb{E}(f_{U^\perp}(x) f_{U^\perp}(x - \mu r) \cdots f_{U^\perp}(x - (k-1)\mu r)|x \in \mathbb{Z}_N, 0 \leq r \leq N_1) \gg_{d,k,\delta,M} 1. \tag{48}$$

Theorem 12.3 (Structure Theorem). *Let $k \geq 3$ be a parameter, and let $f : \mathbb{Z}_N \to \mathbb{R}^+$ be bounded and satisfy $\mathbb{E}(f) \geq \delta > 0$ for some constant δ independent of N. Let $F : \mathbb{R}^+ \to \mathbb{R}^+$ be an arbitrary function which may depend on k or δ but not on N. Then there is a positive number $M = O_{k,\delta,F}(1)$, a bounded function f_U, and nonnegative bounded functions $f_{U^\perp}$, f_{UAP} such that*

- $f = f_U + f_{U^\perp}$
- *(45), (46), (47) hold for $d = k - 2$*
- $||f_U||_{U^{k-1}} \leq \frac{1}{F(M)}$.

[6] The subscript *UAP* stands for *uniformly almost periodic*. A uniformly almost periodic function F of order $k-2$ typically takes the form $F(x) = e^{2\pi i P(x)/N}$ where $P : \mathbb{Z}_N \to \mathbb{Z}_N$ is a polynomial of order at most $k-2$. However, these are not the only examples, and the UAP^{k-2} norm is designed to capture such notion of almost periodicity by quantifying how close an arbitrary function is to such form. The exact definition is a bit technical and can be found in [10]. One should think of the UAP^{k-2} norm as being similar to but larger than the $(U^{k-1})^*$ norm.

One should note that there is a direct analogy between Theorem 12.1 and Theorem 5.4, and between Theorem 12.3 and Theorem 8.1. In fact, the proofs are also very similar. Theorem 12.1 can be proved by Cauchy-Schwarz inequality, and Theorem 12.3 can be proved by using a Furstenberg tower type argument.

Theorem 12.2 is the hardest to prove among the three theorems. It plays the role of the Szemerédi's theorem in our proof of the Generalized Szemerédi's theorem, which is crucial to the entire proof of the Szemerédi's theorem, just as Szemerédi's theorem is crucial in our proof of the Generalized Szemerédi's theorem. The proof of 12.2 also involves a Furstenberg type argument, plus the application of the van der Waerden theorem. The details in the proof are fairly technical (for instance, it involves numerical estimates) and are omitted here. Interested readers can find a complete proof in [10].

With these three theorems, the Szemerédi's theorem follows from a simple argument similar to the one in section 8.

13 Concluding Remarks

The choice of f in Proposition 3.3 might seem a bit fancy and somewhat arbitrary. In fact, a careful reader should note that, in order to use the Generalized Szemerédi's Theorem to conclude the Green-Tao Theorem, it suffices to pick a function with positive expectation that is bounded by a pseudorandom measure and has support contained in the primes. The first step is thus to consider the indicator function of primes: $f(n) := \mathbf{1}[n \in \mathbb{P}]$. However, by the Prime Number Theorem, $\mathbb{E}(f) = \pi(N)/N \sim 1/\log N$, which is not bounded below and thus the Generalized Szemerédi's Theorem does not apply.

This suggests us consider a function that evaluates to something greater than 1 to ensure the expectation is bounded away from 0, but with support still contained in the primes. This motivates the choice of the von Mangoldt function, defined as

$$\Lambda(n) := \left\{ \begin{array}{cc} \log p & \text{if } n = p^m \text{ for some prime } p \\ 0 & \text{otherwise .} \end{array} \right\}$$

By the Prime Number Theorem, we have $\mathbb{E}(\Lambda) = 1 + o(1)$, which is bounded below as desired. However, Λ cannot be bounded by any pseudorandom measure. This is because for any positive integer q, as N gets large, the support of Λ will concentrate on the residue classes $a \mod q$ where $(a, q) = 1$. In other words, the density of the residue classes $a \mod q$ for which $(a, q) = 1$ in the support of Λ will be asymptotically close to 1. But on the other hand, a pseudorandom measure must be uniform across all residue classes $\mod q$, as required by the linear forms condition. This prevents us from bounding Λ by any pseudorandom measure. To make the argument more explicit, consider $A_N := Supp(\Lambda) \cap \mathbb{Z}_N$, $S_N := \{p^m : p \in \mathbb{P}, p|q, m \in \mathbb{Z}^+\} \cap \mathbb{Z}_N$. Then note that if $n \in A_N$ and $(n, q) \neq 1$, we must have $n \in S_N$. Clearly $|S_N| = \sum_{p|q, p\in\mathbb{P}} \lfloor \log_p N \rfloor \leq \sum_{p|q, p\in\mathbb{P}} \log_2 N = \omega(q) \log_2 N$,

where $\omega(q)$ is the number of distinct prime factors of q. But on the other hand, $|A_N| \geq \pi(N)$, thus $|S_N|/|A_N| \to 0$ as $N \to \infty$, which means the support of Λ will concentrate on the residue classes $a \mod q$ where $(a, q) = 1$. In fact, more is true: $\sum_{n \in S_N} \Lambda(n) \leq \log q |S_N|$, thus combining with the Prime Number Theorem $\sum_{n \in A_N} \Lambda(n) = N(1 + o(1))$, we have $\sum_{n \in A_N \setminus S_N} \Lambda(n) = N(1 + o(1))$. On the other hand, if ν is a pseudorandom random measure, then for each $a \mod q$, by the linear forms condition with $m = 1, t = 1, L_0 = q$, and $\psi(x) = qx + a$, we would obtain $\sum_{n \leq N} \nu(\psi(n)) = N(1 + o(1)) = q \sum_{n \leq N, n \equiv a \mod q} \nu(n)$, and thus $\sum_{n \leq N, n \equiv a \mod q, (a,q)=1} \nu(n) = \frac{\phi(q)}{q} N(1 + o(1))$. Since $\phi(q)/q$ can be made arbitrarily small, we cannot bound Λ by any constant multiple of a pseudorandom measure.

Finally, in order to make the argument work, we consider a modified von Mangoldt function as defined in Proposition 3.3. Green and Tao call this modification the W-trick. With this modification, it turns out that the obstructions to pseudorandomness can be eliminated, and we can find a pseudorandom measure that majorizes f, thanks to Proposition 11.1 and Proposition 11.2. This allows us to deduce the Green-Tao Theorem from the Generalized Szemerédi's Theorem.

From the proof of the Green-Tao Theorem, we see that the theorem implies something stronger than there being infinitely many arithmetic progressions of length k in the primes. It asserts that there are at least $\Omega_k(N^2/\log^k N)$ such arithmetic progressions in which each element is at most N. The following simple heuristic might help understand this result. For a large N, we can have a model in which each number from 1 to N is prime with probability π, independently. Then by the Prime Number Theorem, we shall have $\pi = 1/\log N$. Then the probability that a sequence of length k consists entirely of primes will be $1/\log^k N$. Since there are approximately N^2 arithmetic progressions in $\mathbb{Z}_N$ (up to some multiplicative factor depending on k), it follows that there shall be $N^2/\log^k N$ arithmetic progressions of primes (up to a factor), which is consistent with the claim of the Green-Tao Theorem. This is certainly not a proof of the Green-Tao theorem, since the primes are by no means randomly and independently distributed. However, the consistency between the heuristic and the theorem suggests that one might approach problems involving the primes from a probablistic point of view. In fact, the proof of the Green-Tao Theorem has the flavor of this idea by formalizing the notion of pseudorandomness and making connection with ergodic theory such as the Generalized Koopman-von Neumann Structure Theorem.

14 Computing the Green-Tao Numbers

The Green-Tao Theorem is merely an existence statement; in particular, it does not give an explicit algorithm of finding an arithmetic progression of primes for an arbitrary length k. Neither does it provide an upper bound on N below which such progression is guaranteed to exist. Thus naturally, the next question to ask is: for an arbitrary fixed positive integer k, what is the smallest number $N(k)$ such that there exists an arithmetic progression of primes of length k where all the elements are no larger than $N(k)$. This is the problem that we are going to study in this section.

Following the approach in Kullmann's paper ([12]), we have the following definition:

Definition 14.1 (Green-Tao Number). *Let $k_1, \cdots, k_m$ be positive integers. The Green-Tao number $grt_m(k_1, \cdots, k_m)$ is defined as the smallest positive integer N_0 such that for every $N \geq N_0$ and every $f : \{p_1, p_2, \cdots, p_N\} \to \{1, 2, \cdots, m\}$ where p_j is the j-th prime number, there exists some $i \in \{1, 2, \cdots, m\}$ such that $f^{-1}(i)$ contains an arithmetic progression of length k_i.*

There are a few quick remarks about this definition. If $k_1 = k_2 \cdots = k_m = 2$, then by the Pigeonhole Principle it is clear that $grt_m(k_1, \cdots, k_m) = m + 1$. In this case $(k_1, \cdots, k_m)$ is called a *trivial parameter tuple*. If $m = 1$, then $grt_1(k_1)$ is the smallest number of primes we need before an arithmetic progression of primes of length k_1 is obtained. In this case (k_1) is called a *simple parameter tuple*. The Green-Tao Theorem guarantees that $grt_1(k_1)$ is finite for any k_1. Let us also mention that $(k_1, \cdots, k_m)$ is called a *core parameter tuple* if it is non-simple and each $k_i \geq 3$.

In general, $k_1, \cdots, k_m$ may not be the same, in which case $(k_1, \cdots, k_m)$ is called a *mixed parameter tuple*. If $k_1 = k_2 \cdots = k_m$, then $(k_1, \cdots, k_m)$ is called a *diagonal parameter tuple*. It is clear that if $k \geq \max\{k_1, \cdots, k_m\}$, then $grt_m(k, \cdots, k) \geq grt_m(k_1, \cdots, k_m)$.

In this section, we focus on computing the Green-Tao numbers of the form $grt_m(k, \cdots, k)$. To do that, we need to introduce the notion of a hypergraph and translate the problem into a generalized satisfiability problem.

Definition 14.2 (Hypergraph). *A hypergraph G is a pair $G := G(V, E)$ where V is a finite set of vertices and E is collection of subsets of V.*

Note that a hypergraph generalizes the notion of a graph. In a graph, $E \subset V \times V$; but in a hypergraph, we have $E \subset \mathcal{P}(V)$. A hypergraph coloring problem is defined as follows.

Definition 14.3 (Hypergraph Coloring). *An m-coloring of a hypergraph $G(V, E)$ is a function $f : V \to \{1, 2, \cdots, m\}$ such that for all hyperedges $e \in E$ with $|e| \geq 2$, there exist $v_i, v_j \in e$ such that $f(v_i) \neq f(v_j)$, i.e., e is not monochromatic. If G has an m-coloring, then we say G is m-colorable.*

With this, we have the following correspondence.

Lemma 14.4. *Let k, m be fixed positive integer parameters, and n be an arbitrary positive integer. Let $G_n := G(V_n, E_n)$ where $V_n := \{p_1, p_2, \cdots, p_n\}$ and E_n is the collection of arithmetic progressions of length k consisting of elements in V_n. Then $grt_m(k, \cdots, k) > n$ if and only if G_n is m-colorable.*

This is clear by the above definitions. Thus to decide whether $grt_m(k, \cdots, k) > n$, it is equivalent to decide whether G_n is m-colorable. We can further reduce this problem to the following generalized SAT problem.

Definition 14.5 (Generalized Clause). *A generalized clause c is of the form $(y_1, y_2, \cdots, y_l)$ where each y_i is a boolean variable for $i = 1, 2, \cdots, l$. A generalized clause $c := (y_1, y_2, \cdots, y_l)$*

is evaluated to true if and only if not all of its literals are evaluated to true. In other words, $(y_1, y_2, \cdots, y_l) = \vee_i^l(\neg y_i)$.

With this notion of a generalized clause, we have the following correspondence.

Lemma 14.6. *Let* k, m *be fixed positive integer parameters, and* n *be an arbitrary positive integer as before. Let* G_n *be defined as before. For* $0 \leq i \leq n$, $0 \leq j \leq m$, *let* $x_{p_i,j}$ *be boolean variables. Then there is a mapping* $f : V_n \to \{1, 2, \cdots, m\}$ *if and only if there is a truth assignment to the* $x_{p_i,j}$ *such that for every* $i = 1, 2, \cdots, n$, *exactly one of* $x_{p_i,1}, x_{p_i,2}, \cdots, x_{p_i,m}$ *is evaluated to true. (one should think of* $x_{p_i,j} :=$ *true if and only if* $f(p_i) = j$.*) Let* $F(x)$ *be a formula that encodes this. Furthermore, if for every* $e = (e_1, e_2, \cdots, e_{|e|}) \in E_n$, *we define the generalized clause* $c(e, j) := (x_{e_1,j}, x_{e_2,j}, \cdots, x_{e_{|e|},j})$, *then* G_n *is* m*-colorable if and only if* $F(x) \wedge (\wedge_{e \in E_n, 1 \leq j \leq m} c(e, j))$ *is satisfiable.*

Proof. It is clear that $F(x)$ is satisfiable if and only if there is a mapping $f : V_n \to \{1, 2, \cdots, m\}$. In fact, each satisfied truth assignment to $F(x)$ corresponds to exactly one mapping from V_n to $\{1, 2, \cdots, m\}$. Given this, in order for a mapping f to be an m-coloring, we require that no edge is monochromatic, but this is exactly captured by $\wedge_{e \in E_n, 1 \leq j \leq m} c(e, j)$. Therefore, deciding whether $F(x) \wedge (\wedge_{e \in E_n, 1 \leq j \leq m} c(e, j))$ is satisfiable is equivalent to deciding whether G_n is m-colorable. □

Combining Lemma 14.4 and Lemma 14.6, we conclude that deciding whether $grt_m(k, \cdots, k) > n$ is equivalent to deciding whether $F(x) \wedge (\wedge_{e \in E_n, 1 \leq j \leq m} c(e, j))$ is satisfiable.

After transforming the problem of computing the Green-Tao numbers into SAT, we can use SAT solvers to compute the Green-Tao numbers. For instance, the following results are obtained.

Table 1: The first few Green-Tao numbers $grt_1(k)$

k	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$grt_1(k)$	2	4	9	10	37	155	263	289	316	21966	23060	58464	2253121	9686320

Table 2: The first few Green-Tao numbers $grt_2(k, k)$

k	3	4	5
$grt_2(k, k)$	23	512	≥ 34309

Table 3: The first few Green-Tao numbers $grt_3(k, k, k)$

k	3	4
$grt_3(k, k, k)$	137	> 1662

More results can be found in [12].

15 References

[1] PrimeGrid, Primes in Arithmetic Progression Records; http://users.cybercity.dk/˜dsl522332/math/aprecords.htm.

[2] N. Elkies, The asymptotic formula for primes in arithmetic progressions; the Extended Riemann Hypothesis, concerning the zeros of $L(s, \chi)$, lecture notes for *Math 229: Introduction to Analytic Number Theory* (2010); http://www.math.harvard.edu/˜elkies/M229.09/pnt_q.pdf.

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

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

[5] B. Goldston and C. Yıldırım, Small gaps between primes I; http://arxiv.org/pdf/math.NT/0504336.pdf.

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

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

[8] B. Kra, The Green-Tao theorem on arithmetic progressions in the primes: an ergodic point of view, *Bull. Amer. Math. Soc.* **43** (2006), 3-23.

[9] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199-245.

[10] T. Tao, A quantitative ergodic theory proof of Szemerédi's theorem, *Electronic J. Combinatorics* **13** (2006), 49pp.

[11] A. Walfisz, Zur additiven Zahlentheorie. II, *Mathematische Zeitschrift* **40** (1936), 592-607.

[12] O. Kullmann, Exact Ramsey Theory: Green-Tao numbers and SAT, *Theory and Applications of Satisfiability Testing - SAT 2010*, Technical Report arXiv:1004.0653v2 [cs.DM], arXiv, April 2010